New perspectives in IQA: from biochemical applications to controversial chemical bonds

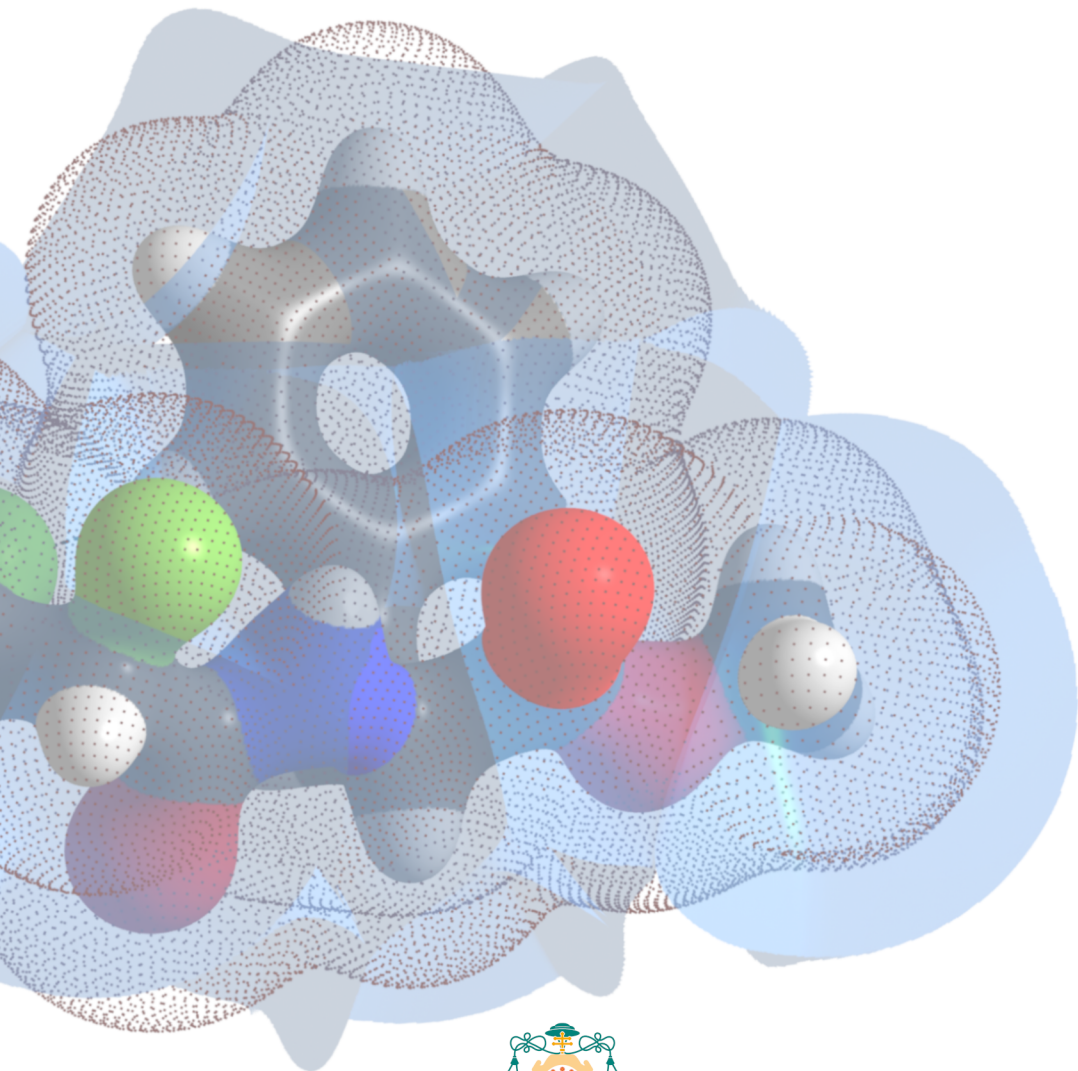Fernando Jiménez Grávalos

UNIVERSIDAD DE OVIEDO

PROGRAMA DE DOCTORADO EN QUÍMICA TEÓRICA Y MODELIZACIÓN COMPUTACIONAL

New perspectives in IQA: from biochemical applications to controversial chemical bonds

Author:
Fernando Jiménez Grávalos

Directors:
Ángel Martín Pendás,
Dimas Suárez Rodríguez

Universidad de Oviedo

Grupo de Química Teórica y Computacional
Departamento de Química Física y Analítica

2021

A mis abuelos

Acknowledgements

En el proceso de desarrollo y elaboración de la presente tesis doctoral han interferido constructivamente multitud de actores, cuya contribución individual puede ser variada, pero únicamente con la suma ponderada de todas ellas ha sido posible alcanzar los objetivos previstos y redactar el presente documento.

Como decimos, el trabajo aquí recogido resulta de una combinación lineal de muy diversas contribuciones. Por una parte, la familia. Apoyo incondicional, moral; profunda creencia en la aptitud del doctorando (algo que roza lo místico, pues no puede ser demostrado) y fuente constante de ánimos y motivación. Padres y hermano, gracias. (Hago una excepción en mi gélido y nórdico carácter para expresarles, sin que sirva de precedente, mi más sincero sentimiento de agradecimiento).

Dicen que los amigos son la familia elegida. Un dicho muy manido, pero al que no le falta razón. Gracias, Kancún, por el también constante apoyo, en especial en los baches más profundos en los que he caído (sesión de terapia en San Francisco). Kancún, y Kancún extendido (antiguos miembros). Kancún extendido, y Cervera en toda su esencia. Porque qué habría sido de la vida de este (todavía, mientras escribe estas líneas) doctorando sin el hogar común. En los últimos años, lugar de descanso y diversión, pero, siempre, desde un principio, fuente de inspiración (terrenal y divina, aunque esto último parece haber llegado a su fin), origen de las más destacadas experiencias vitales y forja de la personalidad de quien escribe. A todos los cerveranos que me alegráis el dia, la tarde y la noche, gracias.

Hasta ahora no ha sido posible personificar los agradecimientos, pues llenaríamos líneas de nombres de personas con las que coincido, con mayor o menor cercanía, en Cervera. Pero, aunque origen y foco principal, no se encuentran las amistades únicamente allí. Dejando atrás la adolescencia, fue en Logroño, en la UR, donde fue posible establecer los primeros vínculos sociales de manera autónoma. Varela, Álvaro, Enrique, Sergio, Vanessa, Miguel y Rebeca, formáis la familia logroñesa de este aspirante a doctor. Gracias por la amistad surgida en la carrera y mantenida ocho años después de terminarla. En especial, gracias a los Dres. Fernández Ochoa y Lara Garnica por sus consejos y ánimos desde la propia experiencia. Han sido verdaderamente útiles. Tampoco podemos olvidar a los directores del primer trabajo de investigación de un servidor, el TFG; gracias, Pedro y Pilar, por vuestra ayuda, compañía e inspiración en aquellos momentos. Aunque la línea escogida después no haya sido la de la espectroscopía, quedarán en la memoria aquellos (intensos) meses de primera aproximación a la investigación.

A pesar de que el doctorado se haya desarrollado en Oviedo, el verdadero primer contacto con el mundo de la química teórica se produjo durante el máster. Primer contacto con esta rama, ¡y primera salida del hogar familiar!, que hasta se mudó en el tránsito a la educación superior por comodidad y para minimizar preocupaciones. Dos años muy intensos en los que no faltó estrés por compartir entre Almudena, Anna, Sonia, Noe y el resto de la clase de Barcelona y de sus amigos, que nos veían día a día sufrir. Más entretenidos fueron los intensivos y la *winter*, donde, ahí sí, pudimos juntarnos toda la clase que se encontraba dispersa por España (Carlos, Julen, del que creía que me había librado, pero no) y el extranjero para vernos las caras, compartir pareceres sobre los *deadlines* vencidos y por vencer, y, también, y muy necesariamente, socializar. Gracias por aquellos momentos. Gracias también a quien dirigió el TFM de este alumno tan perdido, Jordi, y a Juan (Novoa) por su cercanía y amabilidad constantes.

Introduciendo aquí un salto temporal, tampoco podemos olvidarnos de quienes han brindado su apoyo y comprensión en las últimas etapas de redacción de esta

memoria. Gracias, clase del máster de profesorado de la UCM, por haber comprendido las temporales ausencias que no buscaban escaquearse del millar de trabajos grupales que se iban acumulando, sino, simplemente, sobrevivir al solapamiento temporal de escritura de tesis y desarrollo de máster.

Arribamos, ahora sí, a la vetusta ciudad: Oviedo. Dicen que la vida está llena de baches, que todo ser humano pasa por la crisis de los veintitantos y que no hay doctorando que no sufra el llamado síndrome del impostor. Todo ello se dio cita en esta encantadora ciudad y, de no ser por el apoyo cercano de todos los seres de los que me rodeé, esta tesis no podría casi ni haberse empezado. Gracias, Dani, por la ayuda a este recién llegado; Hussien, por la cercana compañía; José Luis, por el soporte técnico constante y paciente; José Manuel, por el acompañamiento, los ánimos y el ejemplo de trabajo; Sara, por la reivindicativa personalidad que más quisiera yo que se me hubiese pegado; Alberto, fuente de amabilidad y disposición, por el tiempo intensamente compartido; Eduardo, por la primera pandemia vivida, cuarentena conjuntamente sufrida; María, nuestra más joven doctoranda, por la compañía y también la referencia en aquello que nos asemeja; Miguel... *coach* de este doctorando, soporte emocional, ejemplo y modelo y, en las últimas etapas, mi más fiel compañero. Junto con el Julen. Hete soportando mis desvaríos, aguantando mis líos de nuevo; consciencia en su ausencia. Gracias a todos. No me puedo olvidar tampoco de mis compañeros durante mi estancia en la RWTH, en Aquisgrán. Jil, Leo, Michael, Emile, Jasper, gracias por vuestra compañía. Y gracias también a Arne, el profesor Lüchow, por su amabilidad durante toda mi estancia allí.

Y amigos de fuera del laboratorio del QCTOVI. Yeferson Javier. Es difícil plasmar por escrito todo lo que siento que agradezco. Pandemia de por medio en el mejor momento, amistad intensa y placentera, despertar de una nueva era. Imborrables son los recuerdos que me llevo. Jaime y Pablo, gracias también por vuestra compañía, cenas y galas eurovisivas. Y, cómo no, Go fit cross. Sin haber aprendido aún la difrencia entre *thruster* y *cluster*, me llevo un muy buen recuerdo de este grupo. Cuatro años de sufrimiento físico compartido, una Farinato superada y cenas en Bueño bien merecidas. Gracias por todos esos momentos.

Como colofón a esta sección abordaremos la manifestación explícita del gran setimiento de gratitud hacia los profesores del QTCOVI, y del departamento, con quienes tantas veces he coincidido en pasillos, charlas, asignaturas de colaboración docente, etc. Ramón e Isabel, ha sido un placer conoceros. Me perdí en el máster la asignatura que entonces daba Isabel, pero pude al menos conocer después un poco más a Ramón en los laboratorios de cuarto. Buenos profesores y mejores personas. Me ha parecido también siempre muy agradable compartir charlas de cafés y comidas con Michi, al único al que sí conocía de las asignaturas de nuestro máster. Aaay, Evelio, qué paciencia has tenido que tener conmigo. Viéndome comer mis (para mí) suculentos platos que serían la envidia de cualquier chef con estrella Michelin, pero, sobre todo, paciencia con la investigación y con la docencia. Quedémonos con que de la experiencia he aprendido mucho [insertar carita sonriente]. Como en toda familia, y el QTCOVI no es una mera unión de individuos con un fin investigador, sino que comparte experiencias, vivencias, reflexiones, (también) discusiones, cenas, recenas, diversiones... hay una figura maternal. Gracias, Aurora, por todo ese apoyo moral, esa escucha activa y paciente, y esas siempre buenas referencias de la mejor gastronomía ovetense. Natalia, gracias también por tu compañía y tu ejemplo. Dimas... mucha paciencia has tenido que tener con este doctorando, no está pagado. Pero al final dicha paciencia ha dado resultado y la tesis se ha acabado (menos mal). Gracias por todo lo que he aprendido, que no es poco. Y es que los comienzos estuvieron marcados por el escaso dominio de la disciplina, problemas que se arrastraron

durante la mayor parte del desarrollo de estos estudios por una no siempre óptima organización. Sin embargo, reflexionando a pocos metros del destino de este largo viaje no puede uno obviar el gran bagaje con el que se despide, bagaje adquirido de manera gradual e imperceptible y que no se limita a lo estrictamente académico, sino que conlleva un gran aspecto personal. Ello no podría haberse logrado de ninguna manera sin la contribución del profesor Pendás. Muchísimas gracias, Ángel, por tus enseñanzas, lecciones, discusiones y aclaraciones, y, sobre todo, por tu altruismo y esa dedicación incesante que es capaz de lograr que doctorandos tan perdidos y desorientados como yo acaben satisfactoria[aceptable]mente sus estudios doctorales.

Gracias a todos. Esta tesis es vuestra.

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: Nuevas perspectivas en IQA: de las aplicaciones bioquímicas a enlaces químicos controvertidos	Inglés: New perspectives in IQA: from biochemical applications to controversial chemical bonds

2.- Autor	
Nombre: Fernando Jiménez Grávalos	DNI/Pasaporte/NIE:
Programa de Doctorado: Química Teórica y Modelización Computacional	
Órgano responsable: Facultad de Química-Centro Internacional de Posgrado de la Universidad de Oviedo	

RESUMEN (en Inglés)

Since its birth around fifteen years ago, the interacting quantum atoms (IQA) energy decomposition has proven to be useful, particularly when the nature of a given chemical bond has to be ascertained. However appealing this reference-free, orbital-invariant technique might be, it has suffered from the strong limitations its high computational demands impose, as well as the unavailability of the second-order reduced density matrix in some popular electronic structure methods. Nonetheless, these problems have recently started to be overcome and new fields of application of this genuine method await to be explored.

In this line, the present doctoral thesis aims at paving the way for a future systematic and consistent application of IQA in computational biochemistry using Hartree-Fock (HF) and Density Functional Theory (DFT) methods. In this regard, we first study the performance of the IQA decomposition of DFT energies as implemented through the scaling technique developed at the University of Oviedo. In addition, the combined IQA-D3 methodology (within an HF-D3 framework that enables a separate treatment of dispersion) is tested on a number of compounds and with different applications. Moreover, its interacting quantum fragments (IQF) variant is shown as a chemically appealing alternative to the complicated atomic analysis one faces when a large number of atoms comes into play. Such a grouped analysis, that yields the energy contributions emanating from functional groups or more general molecular fragments, is tested in the determination of the effects behind the conformational preferences that different fluorine-containing systems present, for which the so-called fluorine *gauche* effect is specifically analysed. But the conformation and activity of biomolecules is very influenced by the surrounding environment, which may contribute decisively in the preferred spatial display of a large macromolecule and determine its activity. For this reason, and as a first step towards more sophisticated descriptions, the IQA atomic and group energies are enhanced with the solvent contribution from an electrostatic continuum solvation model: COSMO. IQA-COSMO energies serve to dissect the electrostatic contribution to the solvation energy into atomic and fragment components, for which a mapping between the atoms/functional groups selected and the value of their respective solvation energies is observed.

Although IQA is generally applied to three-dimensional partitions of space according to the quantum theory of atoms in molecules (QTAIM), other possibilities exist. Moreover, even an orbital decomposition of the wave function of a system can be applied so as to provide a real space view of some orbital-based approximations. It is the case of the charge penetration energy, that emanates from the consideration that two molecules maintain their independent entity upon mutual interaction, and so their molecular densities, that remain unaltered. Accordingly, when two molecules approach each other, their densities just superimpose and interpenetrate one another. In this regard, the IQF

analysis allows for an assessment of the penetration energy – the energy difference between the total electrostatic interaction involving the two interacting densities and that obtained by the multipole approximation– in terms of inter- and intramolecular contributions. Besides, an IQA/IQF intermolecular electrostatic framework is constructed on the basis of its electrostatic interaction pair term that provides a faithful description, as observed after comparison with popular force field electrostatic models.

A second grand line of this thesis is devoted to the elucidation of controversial chemical bonds, such as beryllium, halogen and charge-shift bonding. These applications serve to reinforce the claim on the usefulness IQA and other complementary real space, orbital-invariant quantum chemical topology (QCT) techniques offer. With this, the common electrostatic interpretation of both beryllium and halogen bonding is evaluated and compared with the covalent contributions to these interactions. Special attention is paid to the σ-hole concept, closely related to such an electrostatic interpretation. Charge-shift bonding, in turn, is intended to be considered as a third grand category of bonding as it is argued to be different in nature than the ionic and covalent families. The different QCT tools permit the classification of this bond type according to the information the different descriptors provide in comparison with those reported for typical ionic and covalently-bonded species.

All in all, the results of this thesis point out the numerous advantages of adopting a QCT-IQA description of chemical systems as a robust and physically-rooted alternative to orbital-based methodologies, and whose scope of application is becoming wider and wider.

RESUMEN (en español)

Desde su surgimiento unos quince años atrás, el método de descomposición de la energía de átomos cuánticos interaccionantes (IQA) ha demostrado ser un útil descriptor del enlace químico. Sin embargo, la aplicación de dicho método libre de referencias e invariante orbital se ha visto fuertemente impedida por su alto coste computacional, así como la ausencia de la matriz densidad reducida de segundo orden en varios métodos de cálculo de la estructura electrónica muy populares. Estos problemas, no obstante, han comenzado a ser recientemente solucionados, abriendo nuevas puertas para la aplicación de este genuino método a campos que hasta la fecha esperaban a ser explorados.

Siguiendo esta línea, la presente tesis doctoral pretende sentar las bases para una futura aplicación sistemática, a la par que consistente, de IQA en bioquímica computacional utilizando los métodos tanto de Hartree-Fock (HF) como la teoría del funcional de la densidad (DFT). En este sentido, se presenta en primer lugar un estudio sobre la efectividad de la descomposición IQA de las energías DFT siguiendo la implementación basada en la técnica de escalado desarrollada en la Universidad de Oviedo. Por otra parte, se comprueba el desempeño de la metodología combinada IQA-D3 (dentro de un marco HF-D3 que permite el tratamiento separado de la dispersión) en la descripción de diversos compuestos y con diferentes aplicaciones. Además, su variante de fragmentos cuánticos interaccionantes (IQF) se muestra como una alternativa atractiva desde el punto de vista químico al complicado análisis atómico al que uno se enfrenta cuando el número de átomos comienza a ser sustancial. Este análisis grupal, que proporciona contribuciones energéticas asociadas a grupos funcionales o fragmentos moleculares más generales, se pone a prueba en la determinación de los efectos que subyacen bajo las preferencias conformacionales de sistemas que contienen flúor, prestando especial atención al llamado efecto *gauche* del flúor. Pero la conformación y la actividad de las biomoléculas está muy influida por el entorno que las rodea, pudiendo contribuir decisivamente a la preferente disposición espacial de una gran macromolécula y determinar su actividad. Por este motivo, y como un primer paso hacia descripciones más sofisticadas, se presenta una primera incorporación de las contribuciones del disolvente a las energías atómicas y grupales IQA a partir de un modelo electrostático de

disolvente continuo: COSMO. Las energías IQA-COSMO permiten diseccionar la contribución electrostática a la energía de solvatación en componentes atómicas y de fragmento, observándose una adecuada correspondencia entre los átomos/grupos funcionales seleccionados y el valor de sus respectivas energías de solvatación.

Aunque IQA se aplica en general siguiendo una partición del espacio tridimensional conforme a la teoría cuántica de átomos en moléculas (QTAIM), existen otras posibilidades. No solo eso, sino que es posible incluso aplicar una descomposición orbital de la función de onda de un sistema para proporcionar una visión del espacio real de determinadas aproximaciones basadas en orbitales. Este es el caso de la energía de penetración de carga, que surge de la consideración de que dos moléculas mantienen su entidad independiente al interactuar mutuamente y, por tanto, sus densidades moleculares permanecen inalteradas. En consecuencia, cuando dos moléculas se aproximan, sus densidades se superponen y se interpenetran mutuamente. A este respecto, IQF permite evaluar la energía de penetración —la diferencia energética entre la interacción electrostática total entre dos densidades interaccionantes y la obtenida por medio de la aproximación multipolar— en términos de contribuciones inter e intramoleculares. Asimismo, se construye también un marco de la electrostática intermolecular IQA/IQF sobre la base del término electrostático de pares, que proporciona una fiel descripción, tal y como se observa en su comparación con modelos electrostáticos de populares campos de fuerzas.

Una segunda gran línea de esta tesis está dedicada a la elucidación de enlaces químicos controvertidos, como los de berilio, de halógeno o por desplazamiento de carga. Estas aplicaciones sirven para reforzar la afirmación sobre la utilidad que ofrece tanto IQA como otras técnicas complementarias de la topología químico cuántica (QCT), pertenecientes al espacio real e invariantes orbitales. Con todo ello, se evalúa la interpretación electrostática que comúnmente se realiza de los enlaces de berilio y de halógeno, y se compara con las contribuciones covalentes a dichas interacciones. Se presta una especial atención al concepto de agujero σ, íntimamente ligado a tal interpretación electrostática. El enlace por desplazamiento de carga, por su parte, se ha pretendido que sea considerado como una tercera gran categoría de enlace, argumentando para ello que su naturaleza difiere sustancialmente de la de las familias iónica y covalente. Las diferentes herramientas de la QCT permiten clasificar este tipo de enlace en función de la información que aportan diferentes descriptores en comparación con los reportados para especies con enlaces iónicos y covalentes típicos.

En definitiva, los resultados de esta tesis señalan las numerosas ventajas de adoptar una descripción QCT-IQA de los sistemas químicos como una alternativa robusta y con sólidas bases físicas a las metodologías orbitales, y cuyo ámbito de aplicación está resultando cada vez más amplio.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN QUÍMICA TEÓRICA Y MODELIZACIÓN COMPUTACIONAL

Contents

List of Abbreviations

AIM	Atoms-in-molecules
AMOEBA	Atomic multipole optimised energetics for biomolecular simulation
ASC	Apparent surface charges
BCP	Bond critical point
BJ	Becke-Johnson damping
BLW	Block-localised wave function
BO	Born-Oppenheimer
CASSCF	Complete active space self-consistent field
CC	Coupled cluster
CD	Cumulant density
CCD	Coupled cluster doubles
CCP	Cage critical point
CCSD	Coupled cluster singles and doubles
CI	Configuration interaction
CISD	Configuration interaction singles and doubles
COSMO	Conductor-like screening model
CP	Charge penetration
CSB	Charge-shift bonding
CSF	Configuration state function
CT	Charge transfer
D3	Grimme's third generation dispersion correction
DFT	Density functional theory
DI	Delocalisation index
DMA	Distributed multipole analysis
DMRG	Density matrix renormalisation group
EDF	Electron distribution function
ELF	Electron localisation function
ESP	Electrostatic potential
FCI	Full configuration interaction
GAFF	General Amber force field
GGA	Generalised gradient approximation
GS	Ground state
GTO	Gaussian-type orbital
HB	Hydrogen bond
HF	Hartree-Fock
HK	Hohenberg-Kohn
IAS	Interatomic surface
IMPT	Intermolecular perturbation theory
IQA	Interacting quantum atoms

IQF	Interacting quantum fragments
KS	Kohn-Sham
LB	Lewis base
LCAO	Linear combination of atomic orbitals
LDA	Local density approximation
LI	Localisation index
LPBWE	Lone pair bond weakening effect
LSDA	Local spin density approximation
MCSCF	Multiconfiguration self-consistent field
mGGA	Meta-generalised gradient approximation
MM	Molecular mechanics
MNSol	Minnesota solvation database
MO	Molecular orbital
MPn	Møller-Plesset nth-order
MPPT	Møller-Plesset perturbation theory
MRCI	Multireference configuration interaction
n-RDM	nth-order reduced density matrix
NA	Nuclear attractor
NAdO	Natural adaptive orbital
NBO	Natural bond orbital
NCI	Non-covalent interaction
NNA	Non-nuclear attractor
NOVB	Non-orthogonal valence bond
NPA	Natural population analysis
PES	Potential energy surface
PT	Perturbation theory
QCT	Quantum chemical topology
QM	Quantum mechanics
QMC	Quantum Monte Carlo
QTAIM	Quantum theory of atoms in molecules
RASSCF	Restricted active space self-consistent field
RCP	Ring critical point
RDM	Reduced density matrix
RE	Resonance energy
RESP	Restrained electrostatic potential
RHF	Restricted Hartree-Fock
RMS	Root-mean-square error
RSRS	Real space resonance structure
SAPT	Symmetry-adapted perturbation theory
SAS	Solvent accesible surface
SCF	Self-consistent field
SD	Slater determinant
STO	Slater-type orbital
TF	Thomas-Fermi
VB	Valence bond
VDD	Voronoi deformation densities

VSCC	Valence shell charge concentration
VSEPR	Valence shell electron pair repulsion
x	Exchange
XB	Halogen bond
xc	Exchange-correlation
XCR	Exchange-correlation-repulsion

List of Tables

List of Figures

Chapter 1

Introduction

The complexity of the natural world is something that burdens any person who pays enough attention to what is waiting out there. The apparent simplicity of still water in a puddle or the plain, thin, symmetric leaf of a pine tree conceal a complex net of interacting constituents that make the whole incomprehensible without all its parts. The, at first sight, simple leaf is composed of tens of thousands of cells that are alive as independent units, but which cannot survive without the support of their partners. These provide the structure for the transport of mineral salts, protection against external threats or simply make possible gas exchange. Each of these cells produces its own organic matter, transforms nutrients into more complex structures, undergoes mitosis, etc. They are in turn made of differentiated structures such as membrane, nucleus, Golgi apparatus, mitochondria and other organelles that are nothing but macro-chemical networks of a repeated number of smaller molecules or sets of atoms interacting in such a way that all the functions of the organelles and, ultimately, cell life are preserved. Hence, the macroscopic apparent simplicity turns into an overwhelming complexity when going a bit deeper upon the first perception we get from our senses. We, humans, are in fact a collection of organs and structures composed of cells in an intricate display. We are all aware of the complexity of the human being, not only at the biological level, but also —and more easily averted— sociologically. We believe other human beings are complex in the way they behave, they act, the decisions they make... and, ultimately, they are merely the result of a set of simpler constituents interacting congruently; their feelings, the result of chemical reactions triggered by cellular response. But, although we may agree in the realisation of the high complexity of the world, we have been capable of surpassing our biological limits and thus developing cures for diseases caused by the tiniest virus, making the sempiternal human dream of flying come true, communicating with electromagnetic waves from thousands kilometers away or even conquering the outer space by travelling beyond the edges of our own planet. It has been so thanks to our capacity to fragment complexity, disentangle mixed facets of nature to focus on one or a few and build a model. We create models in our minds everyday, even in an unconscious way. We act upon stimuli, and to cope with them our brains create a model, a simplified representation of the real event we have to manage [1, 2], so we can react to it and anticipate the consequences of one or another reaction. From the spatial representation of a room to process the distribution of objects in space to the use of our memory to avoid controversial topics with some relative on Christmas Eve, we use models in our daily life continuously. Models are inherent to our way of dealing with the real world, and models are the milestone of the framework we use to comprehend it: science [3]. Science emerges as the best manner of understanding nature. It is primarily based on observation, later formulation of a hypothesis and finally evaluation of the proposed explanation (the real process may be not so simple [4], but this one serves as an idealised example). It is roughly the formalisation of

the natural process we daily do in our minds, with the imposition of a testing phase
to check the reasoning formulated. This is how a mere conjecture loses its subjective
character due to the particular view of the individual who proposed it and acquires
explanatory power. There is another difference: science seeks generality, as it tries to
capture as many empirical aspects as possible within the developing models. How-
ever, and although an apparently solid theory serves to explain many facets of life
and nature, scientists eventually come across new scenarios that do not fit into the
already known models. And that does not pose a bump in the road at all, but entails
an exciting new path to be discovered towards a more complete understanding of
nature. It was exactly what happened in the late 19th century and principles of the
20th, when the failure of the very rooted theory of classical mechanics gave rise to
the birth of a new one: quantum mechanics.

1.1   The birth of a new theory

By the 19th century, science was ready to experience a development, supported by
the Industrial Revolution, as it had never seen before. The so-called scientific revo-
lution of the 16th and, especially, 17th centuries brought the breeding ground for a
nascent modern science based upon observation and rationalisation, leaving aside
the dominant Aristotelian tradition [5, 6]. In this context, the Earth lost its central-
ity in the universe, the planets no longer revolved about our planet, nor the stars
were fixed in the sky; the invention of the microscope permitted the discovery of
the cell in many forms such as plant, blood or sperm cells, bacteria and other mi-
croorganisms, what ultimately made dismiss the idea of spontaneous generation;
the mysterious ability of the compass needle to point toward the north was not due
to Polaris, the pole star, nor to a magnetic island somewhere in the north, but to the
fact that the Earth was itself magnetic... Such discoveries and breakthrough in the
way of interpreting the world were reinforced in the century of the Enlightenment
with other remarkable advances such as the law of conservation of mass, that posed
the beginning of modern chemistry, the dismissal of the theory of phlogiston with
the discovery of oxygen and its role played in combustion, the Linnaean classifica-
tion of species —the starting point of modern taxonomy—, the first contemporary
theory of evolution that confronted with the biblical idea of immutable species, or
the developments of the steam engine and the invention of the battery, among many
others [7]. This age was the time of diffusion and popularisation of science and
prepared the western society for the prolific 19th century, the also called Age of
Science [8]. Dalton's atomic theory, Darwin's On the origin of species, Lyell's Princi-
ples of Geology, Maxwell's equations of electromagnetism, Mendel's laws of hered-
ity, Mendeleev's periodic table, the rise of thermodynamics as a science, Pasteur's
discovery of microorganisms as cause of fermentation and disease or Faraday's in-
vention of the dynamo and electric motor are fruits of this splendorous era in the
different, many newly-created branches of science [5, 7]. Such was the confidence in
the development and scientific maturity achieved that scientists began to think that
the major scientific contributions, the guidelines of each discipline —especially in
physics—, had already been set. The dominant belief by the late 19th century was
that only improvements due to more precise measurements were expected, reveal-
ing scarce new knowledge [8]. They were wrong.

Experimental physics had proven the validity of the (classical) theories at that
time for a large number of phenomena, and the motions of mechanical objects were
correctly described by Newtonian mechanics at both the astronomical and terrestrial

scales [9]. Yet, a subatomic particle, the electron, could be discovered as if it were a classical, Newtonian particle. Nonetheless, some experimental results did not follow the classical rules, though it was chiefly ascribed to the absence of a suitable atomic model. Yet, some of the phenomena observed should have been understood by the theories known, but they were actually not. The specific heats of solids at low temperature, the observed only five degrees of freedom in the motion of a free diatomic molecule or the electromagnetic spectrum of the blackbody at thermal equilibrium are some of the unsolved problems that what we call now *classical* mechanics could not explain. Historically, it is the latter that has been considered as the problem whose solution posed the breakdown with the hitherto completely accepted theory of mechanics and marked the beginning of quantum mechanics [10].

1.1.1   A conundrum with the blackbody radiation

Oscillating charges radiate. This statement is the basis for the understanding of the emission due to a body in thermal equilibrium. Experiments show that every object has a certain radiative emission and that such emission covers a given range of wavelengths, having a peak of intensity about one of them and becoming thus the *observed* wavelength of radiation [11]. Empiricism also reveals that the rise in the temperature of the object entails an increase in the observed frequency, what could be related with the inner vibrations of the atoms and molecules conforming the object that are composed of charged subatomic particles —in addition to the neutron, that was discovered in 1932. With the aim of obtaining a relationship between the observed wavelength of an emitting body and the temperature it presents, physics devised the blackbody, a perfect absorber and emitter.

When an object is irradiated with light, some of the incident radiation is absorbed and some is reflected; a blackbody is therefore an idealised object. Nevertheless, physics employed an equivalent device: an object composed of a cavity with a diminutive hollow. In such object, the oscillating charges in the cavity wall emit electromagnetic waves constantly in every direction, but, as they are emitted, they are absorbed by another part of the cavity. Therefore, in thermodynamic equilibrium, the constantly emitting radiation inside the cavity is compensated by a simultaneous absorption, the waves that escape from the cavity being negligible due to the insignificant size of the hole —and so those that enter through it. Experiments with this device in the 19th century showed an energy density $\rho(\nu)$ (or intensity, since they are proportional) distribution conforming a curve, rising from lower frequencies ν to higher in a rate proportional to ν^2, reaching a maximum and then decaying exponentially. It was shown that the curve moves towards higher frequencies when increasing temperature (Wien's displacement law) and becomes more peaked, but the pattern remains. There were two main attempts to attain a relation between the energy density and the frequencies they correspond to basing on either thermodynamic arguments or statistical mechanics [10, 12]. On the one hand, by using thermodynamic arguments and fitting to the empirical curves, Wien obtained the relation

$$\rho(\nu) = c_1 \nu^3 e^{c_2 \nu / T}, \tag{1.1}$$

where c_1 and c_2 are constants and T represents the temperature of the system. Wien's formula agreed with measurements at high frequencies, but not at low. A different approach was followed by Rayleigh and Jeans who, by employing in this case classical electrodynamics and statistical mechanics were led to

$$\rho(\nu) = c\kappa T \nu^2, \tag{1.2}$$

κ and c being the Boltzmann's and another constant, respectively, and becoming only valid for low frequencies of the emission spectrum. In fact, the Rayleigh-Jeans law predicts a continuous increment of energies with ν (the *ultraviolet catastrophe*), what leads to a divergent integrated total energy and the conclusion that the blackbody contains an infinite amount of energy. Therefore, neither the thermodynamic derivation nor the electrodynamical one could satisfactorily explain the experimental emission curve. Rayleigh and Jeans had assumed that the oscillators that compose the cavity walls have a continuous range of energies available. In an attempt to circumvent the problems raised with the classical electrodynamics formulation, Max Planck introduced a revolutionary hypothesis: the energy of a given oscillator of frequency ν must be an integral multiple of a basic quantity or *quantum $h\nu$*, where h is a very small constant. Thus, he found the relation

$$\rho(\nu) = ch\nu^3 \frac{e^{-h\nu/\kappa T}}{1 - e^{-h\nu/\kappa T}}. \tag{1.3}$$

This new formula approaches Rayleigh-Jeans' at small frequencies and tends to Wien's law for higher. Yet, it perfectly agrees with intermediate frequencies, being capable of reproducing the whole blackbody spectrum at different temperatures. Planck's hypothesis sounded strange, but was able to explain the experimental measurements. This finding constituted a breakthrough and was supported by successive new explanations of phenomena such as the photoelectric effect —for which Einstein found the same h constant from Planck's theory and showed the universality of Planck hypothesis that radiation can be only exchanged in units of $h\nu$—, the Compton effect, the energy levels of material systems or the laws for the absorption and emission of radiation. Planck marked the beginning of what we nowadays call quantum mechanics, the theory of mechanics that describes phenomena in the microscopic world; phenomena that appear when looking deep enough but that are otherwise imperceptible from our macroscopic view. Such macroscopic imperceptibility impedes us from noticing the quantisation of matter, as the basic units of energy $h\nu$ are too small for our direct perception. An apparent continuum is nevertheless what is revealed and the physical laws based on it are satisfying in the macroscopic world. Thus, a model that reports good agreement and predictability to a given date is not an absolute truth, but a construction on the basis of rationalisation from empirical experience, and eventually the new advances in the different fields may unveil previously unknown phenomena that escape from the models thus far developed. This was what happened with the failure of classical mechanics, paving the way for a more complete theory with the inclusion of the quantum mechanics, that has in Bohr's correspondence principle the linkage between the two. Two theories developed for the explanation of two apparently different physical worlds, but connected in a complete and elegant fashion.

1.2 A wave function to rule them all

The construction process of the new theory was done by analogies with the already known classical mechanics [9, 10, 13, 14]. According to Newton's laws, known the current state of a (macroscopic) particle, its future can be predicted. In classical mechanics it is sufficient to know the positions and velocities, along with the forces acting on the target particles, to determine the future state of such objects. The quantum world is nevertheless non-deterministic. Heisenberg's uncertainty principle prevents the exact and simultaneous determination of conjugate properties such

as position and momentum. Therefore, it is not possible to know the state (in classical mechanics parlance) of a microscopic particle at any time. To operate with such systems and taking into account the wave-particle duality of the quantum world, a wave or state function Ψ is postulated to act as a system's state descriptor. It is not a mathematical object to unequivocally determine the properties of the system regarded, but rather a function that describes its behaviour in a probabilistic fashion.

The wave function of a system composed of N particles with masses $\{m_i\}_{i=1,N}$ depends on the $3N$ spatial coordinates of all its constituents $\{\mathbf{r}_i\}_{i=1,N}$ plus time t and satisfies the Schrödinger wave equation (in reality a diffusion equation since a wave equation presents a second derivative with respect to time and Schrödinger's has one [14]), that in its non-relativistic formulation reads

$$i\hbar\frac{\partial\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N,t)}{\partial t} = \hat{H}(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N,t)\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N,t). \tag{1.4}$$

This equation relates the temporal evolution of the wave function with the Hamiltonian operator $\hat{H}$ as applied to Ψ. $\hat{H}$ is the operator associated to the total energy of the system and can thus be split into a kinetic energy operator $\hat{T} = \sum_{i=1}^{N} -\frac{\hbar^2}{2m_i}\nabla_i^2$ and a potential energy one $\hat{V}(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N,t)$. When the system is subject to a time independent potential, a solution to the partial differential equation can be of the form

$$\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N,t) = \psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)\theta(t), \tag{1.5}$$

where ψ does not evolve with time and is thus called a stationary state. For such cases, the (time-dependent) Schrödinger equation becomes

$$i\hbar\psi\frac{d\theta}{dt} = \sum_{i=1}^{N} -\frac{\hbar^2}{2m_i}\theta\nabla_i^2\psi + V(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)\psi\theta. \tag{1.6}$$

If we divide both sides by $\psi\theta$, we obtain

$$i\hbar\frac{1}{\theta}\frac{d\theta}{dt} = \sum_{i=1}^{N} -\frac{\hbar^2}{2m_i}\frac{1}{\psi}\nabla_i^2\psi + V(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N), \tag{1.7}$$

where only the left part shows dependency on time and the right on space. Therefore, as time changes, only the left-hand side of the equation changes (and so does the right-hand side when the spatial coordinates vary). Since both sides are equal by the previous relation, both the time-dependent term and the spatial-dependent part must equal a constant, which corresponds to the total energy E. The Schrödinger equation can thus be separated into two differential equations

$$\sum_{i=1}^{N} -\frac{\hbar^2}{2m_i}\nabla_i^2\psi + V(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)\psi = E\psi \tag{1.8}$$

$$i\hbar\frac{d\theta}{dt} = E\theta. \tag{1.9}$$

The solution to the latter is

$$\theta \propto e^{-iEt/\hbar}, \tag{1.10}$$

what, when applied to the spatial part ψ to obtain the complete wave function

$$\Psi = \psi e^{-iEt/\hbar}, \tag{1.11}$$

only provokes a change in its phase. (Note that the time-dependent function θ is proportional to the exponential. The constant of proportionality can be disregarded if one seeks a normalised Ψ.) The energy is therefore not altered and dependent upon the spatial, time-independent function ψ only —hence, its name stationary state. Such function is a solution to the so-called time-independent Schrödinger equation (succinctly referred to as Schrödinger equation in the following)

$$\hat{H}\psi = E\psi. \tag{1.12}$$

Equation 1.12 is in fact an eigenvalue equation, where the wave function is an eigenstate of the Hamiltonian and the energy, its corresponding eigenvalue.

The wave function dictates the behaviour of a given system, however, the lack of a physical analogue impedes a one-to-one correspondence with any measurable feature or object. In other words, Ψ is not an observable. The wave function itself does not reside in the real world, but can be interpreted in statistical terms. According to the Copenhagen or Born's interpretation, Ψ is a probability amplitude and its square modulus $|\Psi|^2$ corresponds to a probability density. Thus, the probability of a single particle to be found in the volume element $d\mathbf{r}$ around $\mathbf{r}$ (that is, between $\mathbf{r}$ and $\mathbf{r} + d\mathbf{r}$) at time t is $|\Psi(\mathbf{r}, t)|^2 d\mathbf{r}$ (whenever Ψ be normalised, otherwise the probability is proportional). Note that if the particle is in a stationary state, the probability density is determined by $\psi(\mathbf{r})$ alone. In fact, in Chemistry it is customary to speak about electron density that is nothing but such a probability density applied to N-electron systems. As will be explained in more detail in Section 2.1, a molecular wave function can be approximated as a stationary state of the electrons moving in the field created by the fixed nuclei, and such a wave function does not only depend on the positions of each electron, but on their spin coordinates $s(\sigma)$ too, giving rise to the spatial-spin coordinates $\mathbf{x}$. Hence, the electron density $\rho(\mathbf{r})$ is defined as

$$\rho(\mathbf{r}) = N \int d\sigma_1 \int d\mathbf{x}_2 \cdots \int d\mathbf{x}_N \psi^*(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) \psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N), \tag{1.13}$$

where the factor N accounts for the indistinguishability of electrons. Such electron density therefore corresponds to averaging $|\psi|^2$ over all but one spatial coordinate and all the spin ones.

The probabilistic interpretation of the wave function implies that if we were to find the particle described by a 1-particle Ψ, an infinite set of experiments would locate it between $\mathbf{r}_1$ and $\mathbf{r}_1 + d\mathbf{r}_1$ with $|\Psi(\mathbf{r}_1, t)|^2 d\mathbf{r}_1$ frequency and between $\mathbf{r}_2$ and $\mathbf{r}_2 + d\mathbf{r}_2$ with $|\Psi(\mathbf{r}_2, t)|^2 d\mathbf{r}_2$ regularity. Concerning a system composed of N equivalent particles, such as electrons, $\rho(\mathbf{r})$ accounts for the probability of finding those electrons per volume element around $\mathbf{r}$, and can be determined experimentally by, for example, X-ray diffraction, representing the observed electron distribution in space.

As regards to the calculation of properties, the distribution given by $|\Psi|^2$ is utilised to estimate the mean or expected value of a given observable. This means that if we were to perform a measurement of a certain property, only if its associated quantum mechanical operator $\hat{\mathcal{O}}$ has Ψ as an eigenfunction, the property will be then well-defined. Otherwise, a set of experiments will give rise to a set of different possible values with a mean value $\langle \mathcal{O} \rangle$ given by

$$\langle \mathcal{O} \rangle = \frac{\int \psi^* \hat{\mathcal{O}} \psi d\mathbf{r}}{\int \psi^* \psi d\mathbf{r}} \equiv \frac{\langle \psi | \mathcal{O} | \psi \rangle}{\langle \psi | \psi \rangle} \tag{1.14}$$

where for the right-side terms the Dirac bra-ket notation has been introduced (hence, representing an integration over the whole domain of ψ). For a normalised wave function, the mean or expectation value is simply given by the numerator of the previous expression, since $\langle \psi | \psi \rangle = 1$. It is also worth mentioning that when the wave function is expanded in terms of other functions $\{\psi_m\}$, a matrix representation of an operator $\hat{O}$ in such basis is possible (whenever the basis is finite), leading to the so-called matrix elements $O_{km} = \langle \psi_k | O | \psi_m \rangle$ of the operator.

The wave function is thus a key concept in the theory of quantum mechanics. Any scientist who seeks for modelling the microscopic world, predict or interpret its behaviour needs first to adequately obtain a proper wave function for the system of concern.

1.3 A physical basis for chemistry

1.3.1 The quantum theory applied to the simplest chemical systems

Once the foundations of quantum mechanics were settled, the study of the simplest systems such as a those composed of a unique particle in different environments (e.g., the free particle, the particle confined in a box or on a ring, the rigid rotor, etc.) could be done. The following step was to address real but also simple systems such as the hydrogen and other hydrogen-like (i.e., one-electron) atoms.

A hydrogen-like atom can be regarded as a system composed of two mutually-interacting particles with a Hamiltonian operator of the kind

$$\hat{H} = -\frac{\hbar^2}{2m_1}\nabla_1^2 - \frac{\hbar^2}{2m_2}\nabla_2^2 + V(\mathbf{r}_1, \mathbf{r}_2). \tag{1.15}$$

Particles 1 and 2 are, indeed, the nucleus, with a positive charge $+Ze$, and a unique electron, with a negative unit of charge $-e$. This well-known two-body problem can be simplified by taking into account that the potential acting on each particle corresponds to the Coulomb potential, with the only dependency on the relative coordinate $r = |\mathbf{r}_1 - \mathbf{r}_2|$ [9, 14]. The Hamiltonian becomes

$$\hat{H} = -\frac{\hbar^2}{2m_n}\nabla_n^2 - \frac{\hbar^2}{2m_e}\nabla_e^2 + \frac{1}{4\pi\epsilon_0}\frac{Ze^2}{r}, \tag{1.16}$$

where n stands for the nucleus, e for the electron and ϵ_0 denotes the vacuum permittivity. This fact can be exploited and the system treated as one composed of a hypothetical free particle of mass $M = m_n + m_e$, corresponding to the centre of mass, and another with reduced mass μ (where $\frac{1}{\mu} = \frac{1}{m_n} + \frac{1}{m_e}$), describing the relative motion of the particles and subject to the Coulomb potential. These are non-interacting, independent particles and, as such, their whole wave function can be split as the product of the individual two. On the one hand, the well-known wave function of the free particle (i.e., the translational motion of the atom as a whole) and, on the other, the one corresponding to the relative motion of the nucleus and electron. The Schrödinger equation for the latter reads

$$-\frac{\hbar^2}{2\mu}\nabla^2\psi + \frac{1}{4\pi\epsilon_0}\frac{Ze^2}{r}\psi = E\psi, \tag{1.17}$$

which is more convenient to be expressed in spherical coordinates, leading to

$$\frac{1}{r}\frac{\partial^2}{\partial r^2}r\psi + \frac{1}{r^2}\Lambda^2\psi + \frac{Ze^2\mu}{2\pi\epsilon_0\hbar^2 r}\psi = -\frac{2\mu E}{\hbar^2}\psi, \tag{1.18}$$

where Λ^2 represents the Legendrian, the angular part of the Laplacian. As can be appreciated, the radial and angular parts are not coupled, and so is the solution to it,

$$\psi_{nlm_l}(r,\theta,\phi) = R_{nl}(r)Y_{lm_l}(\theta,\phi), \tag{1.19}$$

a product of a radial function R and a spherical harmonic Y. The solution to this problem of general form 1.19 is in fact a function of a set of parameters, namely n, l and m_l, the *quantum numbers*, whose values determine the explicit expression of the wave function, in turn, composed by a set of functions and polynomials with distinct forms depending on the value of the parameters. These also present different associated energies and are the different states of the hydrogenic atoms.

The treatment employed for the two-body problem raised by a nucleus and an electron is similar to considering the centre of mass as the pure nucleus and the particle of reduced mass as the electron. Such a similarity stems from the very different mass ratio between the two particles that shall be used in the next sections as the basis of an approximation central to molecular modelling. As such, the functions ψ_{nlm_l} are known as *atomic orbitals* —a term derived from the word *orbit*, in analogy with the classical theory— and determine the relative movement of a unique electron about a nucleus.

1.3.2 The many-body problem

The hydrogen atom (along with the rest of one-electron atoms) is the only chemical system whose associated Schrödinger equation can be solved analytically, since it can be reduced to two pseudo one-body problems. The inclusion of a third particle, be it another nucleus or an electron, impedes the separation of variables in any coordinate system due to the interdependence of each particle on the rest as accounted for by the interaction terms. Thus, apart from the hydrogen-like atoms, the rest of the atomic and the molecular wave functions cannot in principle be obtained exactly due to their many-body nature. However, this fact does not preclude the usage of an approximate treatment. In order to circumvent this issue, two main families of approximations emerged: those based upon the consideration of the interaction term as a small perturbation of the non-interacting system —and thus called *perturbative* approaches— and those seeking to attain a correct wave function by means of some trial, approximate wave function with parameters to be optimised —the *variational* method.

Perturbation theory

The (time-independent and non-degenerate) perturbation theory is based on a simple model system whose eigenstates (that shall be called $\{\psi_m^{(0)}\}_{m=0,1,2,...}$) are known and that is subject to a certain small perturbation. The true system is characterised by a Hamiltonian $\hat{H}$ that differs from that of the model system $\hat{H}^0$ by a small perturbation $\hat{H}'$ [13, 15]. Hence,

$$\hat{H} = \hat{H}^0 + \lambda\hat{H}', \tag{1.20}$$

where $\lambda \in [0,1]$ accounts for the level of perturbation acquired, $\lambda = 0$ representing the model system and $\lambda = 1$ the completely perturbed system, that is, the true system. (It is a mere instrument to keep track of the perturbative order as shall be seen.)

The set of known eigenstates of the model system, in turn, fulfil the property that are mutually orthogonal (the inner product between two different states is zero). They are also normalised, so that they are said to be *orthonormal*. This property can be succinctly expressed as $\langle \psi_n | \psi_m \rangle = \delta_{nm}$, where δ_{nm} is the *Kronecker delta* and takes value 1 for $n = m$ and 0 otherwise.

Since the Hamiltonian depends on the parameter λ, so do its eigenfunctions ψ_m and energies E_m. By expanding them in a Taylor series about $\lambda = 0$,

$$\psi_m = \psi_m|_{\lambda=0} + \left.\frac{\partial \psi_m}{\partial \lambda}\right|_{\lambda=0} \lambda + \left.\frac{\partial^2 \psi_m}{\partial \lambda^2}\right|_{\lambda=0} \frac{\lambda^2}{2!} + \cdots \tag{1.21}$$

$$E_m = E_m|_{\lambda=0} + \left.\frac{dE_m}{d\lambda}\right|_{\lambda=0} \lambda + \left.\frac{d^2 E_m}{d\lambda^2}\right|_{\lambda=0} \frac{\lambda^2}{2!} + \cdots , \tag{1.22}$$

where $\psi_m|_{\lambda=0} = \psi_m^{(0)}$ and $E_m|_{\lambda=0} = E_m^{(0)}$. Equations 1.21 and 1.22 become

$$\psi_m = \psi_m^{(0)} + \lambda \psi_m^{(1)} + \lambda^2 \psi_m^{(2)} + \cdots \tag{1.23}$$

$$E_m = E_m^{(0)} + \lambda E_m^{(1)} + \lambda^2 E_m^{(2)} + \cdots \tag{1.24}$$

with

$$\psi_m^{(k)} = \frac{1}{k!} \left.\frac{\partial^k \psi_m}{\partial \lambda^k}\right|_{\lambda=0} \quad \text{and} \quad E_m^{(k)} = \frac{1}{k!} \left.\frac{d^k E_m}{d\lambda^k}\right|_{\lambda=0}$$

as the different kth-order corrections to either the wave function or the energy.

The Schrödinger equation for the real system then reads

$$(\hat{H}^0 + \lambda \hat{H}')(\psi_m^{(0)} + \lambda \psi_m^{(1)} + \lambda^2 \psi_m^{(2)} + \cdots) =$$
$$(E_m^{(0)} + \lambda E_m^{(1)} + \lambda^2 E_m^{(2)} + \cdots)(\psi_m^{(0)} + \lambda \psi_m^{(1)} + \lambda^2 \psi_m^{(2)} + \cdots) \tag{1.25}$$

and can be reorganised to grouping the different kth-order terms as

$$\hat{H}^0 \psi_m^{(0)} + \lambda(\hat{H}^0 \psi_m^{(1)} + \hat{H}' \psi_m^{(0)}) + \lambda^2(\hat{H}^0 \psi_m^{(2)} + \hat{H}' \psi_m^{(1)}) + \cdots =$$
$$E_m^{(0)} \psi_m^{(0)} + \lambda(E_m^{(0)} \psi_m^{(1)} + E_m^{(1)} \psi_m^{(0)}) + \lambda^2(E_m^{(0)} \psi_m^{(2)} + E_m^{(1)} \psi_m^{(1)} + E_m^{(2)} \psi_m^{(0)}) + \cdots \tag{1.26}$$

Provided the model system has been chosen appropriately and the expansion converges, the first-, second-order and subsequent corrections to both the wave function and the energy will produce smaller additional changes when including higher order corrections. Since λ is arbitrary, all the terms of a given order at both sides of Equation 1.26 must coincide. Thus, the model system is represented by the 0th-order terms

$$\hat{H}^0 \psi_m^{(0)} = E_m^{(0)} \psi_m^{(0)}, \tag{1.27}$$

the first-order correction comprises the solution to

$$(\hat{H}^0 - E_m^{(0)})\psi_m^{(1)} = (E_m^{(1)} - \hat{H}')\psi_m^{(0)}, \tag{1.28}$$

the second one to

$$(\hat{H}^0 - E_m^{(0)})\psi_m^{(2)} = (E_m^{(1)} - \hat{H}')\psi_m^{(1)} + E_m^{(2)}\psi_m^{(0)}, \tag{1.29}$$

and so on. By manipulating such expressions, the different kth-order corrections are obtained.

First-order correction

In perturbation theory, the corrections to the wave function are treated as a mixing of the zeroth-order m state with the rest of the same-order states. The perturbation thus acts as a distortion on the reference system and induces virtual transitions to other states. As such, the first-order correction to the m-state wave function is expressed as

$$\psi_m^{(1)} = \sum_n a_n \psi_n^{(0)}. \tag{1.30}$$

By substituting in the first-order equation and multiplying on the left by $\langle\psi_m^{(0)}|$ (or, equivalently, by multiplying by $\psi_m^{(0),*}$ and integrating),

$$\sum_n a_n(E_n^{(0)} - E_m^{(0)})\langle\psi_m^{(0)}|\psi_n^{(0)}\rangle = E_m^{(1)}\langle\psi_m^{(0)}|\psi_m^{(0)}\rangle - \langle\psi_m^{(0)}|\hat{H}'|\psi_m^{(0)}\rangle. \tag{1.31}$$

Since the states are orthonormal, the only surviving term in the summation over the states is that for which $n = m$. Therefore, the whole left-hand side of the equation vanishes and the first-order correction to the energy becomes just the matrix element of the perturbation $\hat{H}'$ on the reference state,

$$E_m^{(1)} = \langle\psi_m^{(0)}|\hat{H}'|\psi_m^{(0)}\rangle \equiv H'_{mm}. \tag{1.32}$$

This correction corresponds to the average effect of such a perturbation over the state $\psi_m^{(0)}$.

The calculation of the corresponding first order correction to the wave function implies obtaining an explicit form for the expansion coefficients a_n of Eq. 1.30. The procedure for doing so is similar to the one described for $E_m^{(1)}$. By taking the first-order equation and multiplying on the left by a different bra $\langle\psi_k^{(0)}|$ in this case so that the left-hand side part does not vanish (and with it a_n),

$$\sum_n a_n(E_n^{(0)} - E_m^{(0)})\langle\psi_k^{(0)}|\psi_n^{(0)}\rangle = E_m^{(1)}\langle\psi_k^{(0)}|\psi_m^{(0)}\rangle - \langle\psi_k^{(0)}|\hat{H}'|\psi_m^{(0)}\rangle. \tag{1.33}$$

The orthonormality condition leads in this case to $n = k$ as the only surviving term in the summation, a_k thus becoming

$$a_k = \frac{H'_{km}}{E_m^{(0)} - E_k^{(0)}}, \tag{1.34}$$

which is valid for non-degenerate states whenever $k \neq m$.

Second-order correction

Similarly to the previous derivation of the first-order corrections, the starting point for the second-order ones is the corresponding expansion of $\psi_m^{(2)}$ in terms of

the model system's eigenstates

$$\psi_m^{(2)} = \sum_n b_n \psi_n^{(0)}. \tag{1.35}$$

By considering Equation 1.29 now, substituting the first- and second-order corrections to the wave function and multiplying by $\langle \psi_m^{(0)}|$ on the left, one is led to

$$\sum_n b_n (E_n^{(0)} - E_m^{(0)}) \langle \psi_m^{(0)}|\psi_n^{(0)} \rangle = \sum_{n \neq m} a_n \langle \psi_m^{(0)}|E_m^{(1)} - \hat{H}'|\psi_n^{(0)} \rangle + E_m^{(2)} \langle \psi_m^{(0)}|\psi_m^{(0)} \rangle. \tag{1.36}$$

Due to orthonormality, the left-hand side part and the term with $E_m^{(1)}$ are zero, so the second-order correction to the energy becomes

$$E_m^{(2)} = \sum_{n \neq m} \frac{H'_{nm} H'_{mn}}{E_m^{(0)} - E_n^{(0)}}. \tag{1.37}$$

Since $H'_{nm} H'_{nm} = H'^*_{mn} H'_{nm} = |H'_{mn}|^2$, the second-order contribution will always play a stabilising role for ground states ($m = 0$).

Following the same receipe used so far, the second-order correction to the wave function, the third-order ones, etc., could be obtained. However, for later discussion only the energy corrections up to second order are needed. Moreover, the knowledge of the first-order correction to the wave function suffices to derive corrections to the energy up to third order (according to Wigner's $2n + 1$ rule which states that energy corrections up to $2n + 1$ order can be achieved provided the nth-order correction to the wave function [16]).

This is the general procedure due to Rayleigh and Schrödinger. A simple application is the search for the minimum energy state, that is, the ground state (GS) of the helium atom. In such case, the problem can be viewed as two independent one-particle systems subject to a perturbation caused by the mutual interaction between electrons 1 and 2. Hence,

$$\hat{H}^0 = -\frac{\hbar^2}{2m_e} \nabla_1^2 - \frac{\hbar^2}{2m_e} \nabla_2^2 - \frac{1}{4\pi\epsilon_0} \frac{e^2}{r_{1n}} - \frac{1}{4\pi\epsilon_0} \frac{e^2}{r_{2n}} \tag{1.38}$$

and the perturbation,

$$\hat{H}' = \frac{1}{4\pi\epsilon_0} \frac{e^2}{r_{12}}. \tag{1.39}$$

The model system is described by the product of the hydrogenic $1s$ wave function of each electron. It yields an associated ground state energy in a 38% of error with respect to the experimental one. The correction provided by the perturbation $\hat{H}'$ up to third order results in an almost perfect agreement with a small error of $0.02\ eV$.

Variation theory

An alternative scheme is provided by the variational method. It is based on the variational principle, that states that given a wave function ψ_{trial} fulfilling the boundary conditions of the problem at hand, its associated energy (i.e., the expectation value of the Hamiltonian) is an upper bound to the exact ground state energy E_0 given by the Rayleigh ratio

$$E = \frac{\langle \psi_{trial}|\hat{H}|\psi_{trial} \rangle}{\langle \psi_{trial}|\psi_{trial} \rangle} \geq E_0. \tag{1.40}$$

The procedure of this method involves an optimisation of a set of parameters the trial wave function is dependent upon so as to converge towards E_0. When the set of parameters allows an appropriate exploration of the Hilbert space, the converged wave function corresponds to the the exact ground state one ψ_0.

To exemplify the performance of this second method, the He atom can be again evaluated. The aforementioned perturbation treatment acts over the hydrogenic wave functions as model systems. If a similar trial wave function is used, replacing the atomic number Z by a parameter ζ in the $1s$ hydrogen-like wave functions of each electron, its optimisation leads to an energy 1.53 eV in error (a 1.9%) with the exact value. Since any function fulfilling the boundary conditions can be used as a trial function, another attempt can be done by introducing the interelectronic repulsion term (absent in the previous variational wave function) multiplied by an adjustable parameter. This was the spirit followed by Hylleraas and led to an improved variational wave function whose energy was in a reduced error of 0.3 eV (further corrected, in turn, by subsequent authors).

Nevertheless, the general usage of the variational method entails its linear formulation by Rayleigh and Ritz. Having noticed the difficulties to come up with an adequate trial wave function for each problem to be optimised, these authors proposed the usage of linear variations only, making thus use of a trial wave function of the form

$$\psi_{trial} = \sum_i c_i \psi_i. \tag{1.41}$$

In such method, the wave function is expanded as a linear combination of a set of fixed functions (the basis functions ψ_i), whose coefficients c_i are the parameters to be optimised. When introducing such trial function in the Rayleigh ratio, expression 1.40 becomes

$$E = \frac{\sum_{ij} c_i c_j \langle \psi_i | \hat{H} | \psi_j \rangle}{\sum_{ij} c_i c_j \langle \psi_i | \psi_j \rangle} = \frac{\sum_{ij} c_i c_j H_{ij}}{\sum_{ij} c_i c_j S_{ij}}, \tag{1.42}$$

where H_{ij} are the matrix elements of the Hamiltonian and S_{ij} the overlap integral between functions ψ_i and ψ_j.

The optimisation therefore comprises the simultaneous minimisation of the energy with respect to each coefficient. Thus, $\partial E / \partial c_k = 0$ —whenever the coefficients are independent; otherwise, using Lagrange's method of undetermined multipliers instead—, what ultimately leads to a set of secular equations

$$\sum_i c_i (H_{ik} - E S_{ik}) = 0. \tag{1.43}$$

A non-trivial solution to the above equation entails that the secular determinant be zero,

$$det(H_{ik} - E S_{ik}) = 0, \tag{1.44}$$

what leads to a set of values of E as the roots of their corresponding polynomials, the lowest value being the approximate ground state energy.

1.3.3 Computational chemistry

The development of the quantum theory led to a new subfield of chemistry as the branch dealing with the development and application of methods rooted in quantum mechanics to unravel and solve chemical problems. The so-called quantum chemistry also laid the foundations of a more general theoretical chemistry, that

was intended to explain a predict chemical processes, physical properties of compounds or to rationalise the structures they present on the basis of both conceptual and mathematically rooted models. These models comprise the different and successive descriptions of chemical bonding, the explanation of the reaction rates observed (chemical kinetics) or the rules for chemical reactivity derived from wave function interpretation.

In the early ages of quantum mechanics, equations had to be solved by hand. It was not until the invention of computers that more complex problems could be addressed. Computers provided the means for performing individual operations that otherwise would have been unaffordable. Even if an idealised large number of individuals working ceaselessly were employed, the performance of a computer would be much greater due to the rapid access to memory, connections among different processing units, etc. Provided a set of instructions as a sequence of logical operations conforming an algorithm, a computer is capable of distributing the tasks among its processors to speed up the whole process by means of parallelisation. Methods in theoretical chemistry rose basing on different approaches and taking advantage of the computer architectures at hand. As such, among the plethora of methods available, it has been possible to implement really accurate ones to obtain proper wave functions with calculated energies within narrow error ranges for small and medium-sized molecules. Not only is the theoretical study of chemistry based on quantum physics, but also on other subsequent incorporations to the set of methods that go beyond the rigorous quantum mechanical treatment for atomic-scale systems. These range from the application of classical physics to simulate the dynamics of a large protein in solution to the more modern use of data mining to predict chemical reactions mechanisms and products [17]. All these methods, although may escape from the initially conceived quantum chemistry, are part of a broader discipline focused on the application of the different theoretical approaches (with the help of computers, hence its name) to chemical problems: computational chemistry [18, 19].

Computational chemistry covers indeed a wide range of computational tools rooted in very distinct approaches. These can be chiefly classified in quantum mechanical (QM), as those relying on obtaining a wave function from which all measurable properties can be calculated, and molecular mechanics (MM) methods, that make use of simplified models rooted in simpler classical physics laws to simulate (usually huge) systems or even parts of them when the accurate description by QM is not necessary and/or affordable. Indeed, QM methodologies have traditionally been applied to not very large systems due to the rapid growth in computational requirements with the number of atoms (specifically, electrons and basis functions). As such, either hybrid QM/MM methodologies for large systems where a target moiety is treated at the QM level while the rest is modelled classically, or simpler QM techniques when a profound description is needed have been the choice for those situations. It is the case of non-*ab initio* (i.e., those methods that do not obtain a quantum mechanical wave function step by step from scratch) QM methods such as semi-empirical methods, that aim to simplify costly (two-electron) integrals by setting them to specific values (such as those provided by fitted parameters), or density functional theory (DFT), that also makes use of a set of parameters to estimate the energy and obtaining an approximate pseudo-wave function, as shall be seen later. However, other full QM alternatives also exist. To cite a couple, quantum Monte Carlo (QMC) and density matrix renormalisation group (DMRG), that exploit very different approaches than traditional methods and achieve a high level of accuracy while retaining moderate resources demands. Furthermore, the ongoing

development of quantum computers opens new doors for the exact QM treatment of bigger systems; methods whose implementation in classical computers made it thus far unbelievable [20].

Chemistry began its adventure as a modern science, independent from the mysticism that had accompanied alchemy over the years, as an eminent experimental discipline. The rise of quantum mechanics, however, marked a turning point whose effects can still be perceived nowadays. The current computational chemistry has its own portion in scientific research, albeit it complements many experimental investigations and vice-versa. It has become an indispensable tool for experimentalists for rationalising their results and, at the same time, it itself has marked many of the research lines in synthetic chemistry. The proposition of new catalytic routes, novel materials with promising mechanical or electrical properties or the search for more specific and less harmful drugs have encouraged many experimental work, proving the usefulness of theoretical research in promoting overall chemistry.

The usefulness of computational techniques lies on the set of tools to analyse and interpret the results of a simulation. Special attention deserves the analysis of the wave function, an object that does not belong to the real space but to Hilbert's. The reconciliation of chemistry with quantum mechanics has therefore been (and it still is, indeed) a major concern among theoretical chemists. The interpretation of quantum mechanical results in chemist's parlance has provided a plethora of methods of analysis built upon many different theoretical views, what sometimes leads to contradictory explanations depending on the particular model of choice.

Traditionally, and still reminiscent in the chemist's mind, chemical interpretation accompanied the method derived for the wave function calculation. The first quantum mechanical treatment of a (neutral) molecule was due to Heitler and London in 1927 [13, 19]. They developed a theory known as Valence Bond (VB) that relied directly on overlapping atomic orbitals to describe bonding (initially) between pairs of atoms. Their idea was successful from an interpretative point of view, but suffered from the high computational demands that it required. By contrast, the Molecular Orbital (MO) picture emerged as the preferred one for the so-called electronic structure calculations (i.e., the calculation of the electronic wave function under a fixed nuclei approximation) due to its simply and effective implementation, though, in turn, it suffered from an interpretative issue since this model is based upon linear combinations of atomic orbitals that ultimately lead to highly delocalised molecular orbitals throughout the whole molecule. No clear bonds, nor atomic units are found. Later developments led to molecular orbital models that localised the set of orbitals, maintaining the global wave function unaltered, to more clearly visualise bonds, associate electrons to atoms and explain chemical structure, reactions or stability in simple atomic terms. However, as can be averted from the previous words, a wave function can be constructed from many different sets of orbitals (interconverted through *unitary transformations*), all of them giving a proper final description of the system. The most widely-used interpretative schemes have therefore been linked to the particular method utilised for solving the Schrödinger equation, resorting to non-physical objects such as orbitals to obtain chemical arguments.

A radically different perspective is taken under *orbital-invariant* techniques, such as those relying on real space descriptors. In this family of computational methods, orbitals are merely a means for constructing the target mathematical functions such as the electron density or the more general reduced density matrices (RDMs), so that the set of orbitals employed do not affect the final outcome. This is the theoretical background this thesis aims to spread and popularise. In the following, and with the aim of contextualise the work carried out in the present thesis, two controversial

tasks within quantum chemistry are addressed. On the one hand, the way atoms are defined in a more complex entity, such a molecule, and, on the other, how the total energy obtained from a quantum mechanical calculation can be decomposed and related to chemical concepts. Both conundrums are in the heart of the present doctoral thesis.

1.4   The atomic identity in the molecular whole

The quantum mechanical treatment of chemical systems through the solution of the Schrödinger equation results in a more or less accurate system's wave function, depending on the particular choice of the method to approximate it. Such a mathematical object that describes the behaviour of the system and determines its properties is nevertheless irreducible to smaller parts. In the case of a molecule, the solution of the corresponding Schrödinger equation permits the calculation of the molecular wave function. But chemistry is about atoms interacting with each other, forming bonds, a structured display where the positions of atoms that determine properties change upon a chemical reaction. Moreover, the atomic framework is usually decomposed into subunits, the *functional groups*, that are said to be *transferable*, what means that those groups present specific properties that are carried from one molecule to another where a given functional group is inserted.

The molecular structure hypothesis is in fact a pillar of the chemical science. An isolated atom by itself is a diffuse entity as the electron density it comprises extends to infinity. When approaching to another, there is no way of fully disentangle the two interacting atoms whose electrons move throughout the whole system. The two nuclei can be readily determined, but the sea of electrons about the two nuclear islands cannot be measured as to belonging to one or another atom unequivocally. The wave function reflects this experimental and just mental observation. If two subsystems are in interaction, a correlation between them exists and hence a wave function for the whole describes them. According to the probabilistic interpretation, this is similar to asserting that the wave function of a system does not equal a product of the individual subsystems' ones, that would be the non-correlated case. Instead, the entire system appears to be indivisible. Fortunately, several theoretical models have tackled this chemists issue and proposed *atoms-in-molecules* (AIM) decomposition schemes that range from the simplest atomic charge assignment based on atom-centred basis functions to more sophisticated charge density decompositions built upon more physical grounds [21].

The atomic division of a molecule can be performed in either the Hilbert space, where the wave function resides, or in the three dimensional, real one. Among the Hilbert space approaches is the Mulliken population analysis, probably the most famous of this group. When considering only the electron motion about nuclei and the corresponding electronic wave function being expanded as a linear combination of atom-centred basis functions (such as molecular orbitals in terms of atomic ones, as shall be seen in the following Chapter), the Mulliken scheme distributes the electron charge carried by those basis functions as belonging to a particular nucleus or equally split between two according to the overlap between those functions and whether they are centred on the same or different atomic sites. Other techniques that belong to the same group are the Löwdin and the natural population analyses (NPA), the latter achieved from a particular set of atomic orbitals (the so-called *natural* orbitals) [19, 22]. More physical has been proved the second group of techniques. Instead of resorting to orbitals, these decomposition methods partition the charge

density (or simply and specifically the electron density that surrounds nuclei at a given fixed molecular geometry) among atoms by making use of weight functions $w_A(\mathbf{r})$ as follows:

$$\rho(\mathbf{r}) = \sum_A \rho_A(\mathbf{r}) = \sum_A \int w_A(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}, \tag{1.45}$$

from which atomic charges $Q(A)$ can be calculated by subtracting the electron density ascribed to a given atom to its corresponding nuclear charge,

$$Q(A) = Z_A - \int \rho_A(\mathbf{r})d\mathbf{r}. \tag{1.46}$$

The total molecular density $\rho(\mathbf{r})$ is therefore recovered from the sum of the atomic ones $\rho_A(\mathbf{r})$, and the weight functions fulfil

$$\sum_A w_A(\mathbf{r}) = 1 \ \ \forall \mathbf{r} \in \mathbb{R}^3. \tag{1.47}$$

With the previous requirement as a common feature, several definitions of weight functions have appeared in the last decades and, depending on whether they provide well-defined boundaries or not, their resulting decompositions have been in turn classified as *fuzzy* for the latter case and non-fuzzy for the former.

Within the fuzzy atomic decompositions, the Hirshfeld method is particularly popular. This scheme relies on the *promolecular* density $\rho_{pro}^0(\mathbf{r})$ to weigh the different atomic contributions to the molecular $\rho(\mathbf{r})$. A promolecule is a fictitious molecule constructed from the (usually spherically averaged) atomic ground state densities $\rho_A^0(\mathbf{r})$ placed at the final atomic positions within the molecule,

$$\rho_{pro}^0(\mathbf{r}) = \sum_A \rho_A^0(\mathbf{r}). \tag{1.48}$$

The weights are calculated from the quotient between the atomic isolated density and the promolecular one, yielding atomic densities within the molecule of the form

$$\rho_A^H(\mathbf{r}) = w_A^H(\mathbf{r})\rho(\mathbf{r}) = \frac{\rho_A^0(\mathbf{r})}{\rho_{pro}^0(\mathbf{r})}\rho(\mathbf{r}). \tag{1.49}$$

The Hirshfeld scheme provide atoms as close as possible to the isolated ones and can also be derived from information entropy [21]. It was shown that Hirshfeld weights $w_A^H(\mathbf{r})$ correspond to the functions that minimise the information loss I between the promolecule and the true molecular entity

$$I = \sum_A \int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) d\mathbf{r} \tag{1.50}$$

whenever the atoms in the molecule maintain the same number of electrons they possess as isolated species. A generalisation that overcomes some issues of the original formulation, such as the large dependency on the promolecule chosen, is its iterative version by Bultinck et al. [21].

However, when considering a molecule, such as the water molecule, given by a set of points representing the compact atomic nuclei immersed in a continuous electron density, the simplest way of assigning electrons to nuclei and therefore visualising atoms within a molecule is probably through Voronoi polyhedra [23]. A Voronoi diagram is a partition of space filled with distributed objects in regions whose points
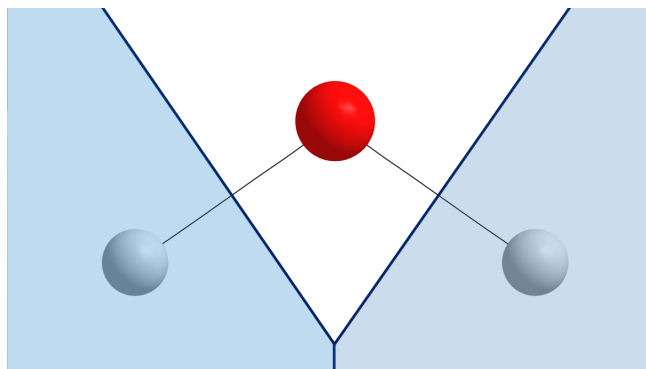
FIGURE 1.1: Voronoi diagram of the water molecule where each atomic region possesses the spatial points closer to its corresponding nucleus and not others. The molecular representation has been obtained from AIMAll [24].

are closer to each object than to others. Therefore, as shown in Figure 1.1, the atomic regions as Voronoi polyhedra for the planar water molecule are delimited by separatrices traced perpendicular to each bond or linking line between atoms.

Based on this simple scheme various AIM methodologies are found. On the one hand, Becke's scheme [25] partitions space in fuzzy Voronoi cells by softening the weight functions. A such, Becke's procedure starts from an ordinary Voronoi partition and then uses polynomials to smooth the weights that are otherwise binary step functions that assign every point of space to one and only one nucleus. Similar procedures were taken by Clark and Davidson [26] and by Mayer and Salvador [27], with the slight modification of including two simultaneous weight functions so as to discriminate between the proper atomic density and that shared between two atoms. By contrast, Fonseca Guerra et al. [22] employed Voronoi deformation densities (VDD) as

$$\rho_A^{VDD} = w_A^V(\mathbf{r})\rho_{def}(\mathbf{r}) = w_A^V(\mathbf{r})\left(\rho(\mathbf{r}) - \rho_{pro}^0(\mathbf{r})\right), \tag{1.51}$$

focusing on the density change from the promolecule to the true molecular distribution and partitioning it by means of Voronoi weights $w_A^V(\mathbf{r})$.

Since the definition of atoms in molecules is not and cannot be unique —they cannot be observed by any experiment—, there are very distinct approaches, not only in Hilbert space relying on one or another set of orbitals, but also in Cartesian space. A very different approach to the above-mentioned ones where this dissimilarity is readily appreciated was accomplished by Li and Parr [28], who relied on atomic densities following a minimum promotion energy, defined as the energy difference between the isolated atomic density and the final atomic portion of the molecular charge distribution.

Among all the AIM schemes, there is one that itself inaugurated a new theory to interpret quantum mechanical results, whose grounds served to accomplish a fruitful connection between chemical concepts such as chemical bonds, bond order, Lewis pairs, acidity or basicity and the information contained in the inscrutable wave function. It is Bader's Quantum Theory of Atoms in Molecules (QTAIM) [29, 30]. Although it constitutes an extensive theory, we are for the moment only concerned with how such a theory recognises atoms as belonging to a more complex entity. Bader's approach is based on the electron density, as the previously-cited AIM schemes, but more specifically, it relies on the topology of such scalar field.
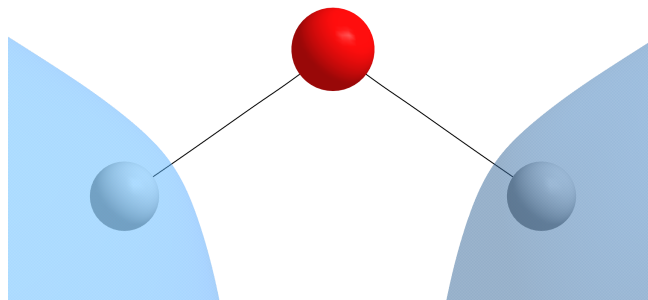
FIGURE 1.2: Bader's QTAIM atomic partitioning of the water molecule. This graphic has been obtained with the AIMAll program [24].

Hence, the name Quantum Chemical Topology (QCT), coined by Paul L.A. Popelier to refer to the set of techniques rooted in the topological study of quantum mechanical functions, such as the electron density, the Laplacian of ρ or the electron localisation function (ELF) [31].

Paramount to QTAIM is the gradient field of ρ: $\nabla\rho(\mathbf{r})$. It constitutes a vector field, a collection of all gradient paths that connect sources to attractors; trajectories that (essentially) originate at infinity where ρ vanishes and (usually) end up at the nuclei. However, not all these trajectories terminate at nuclear positions: some trace paths that end up at points between two nuclei and configure a surface of trajectories that are tangential to those terminating at the nucleus. In other words, these surfaces fulfil

$$\nabla\rho(\mathbf{r}) \cdot \hat{\mathbf{n}} = 0 \ \ \forall \mathbf{r} \in S(\Omega), \tag{1.52}$$

where $\hat{\mathbf{n}}$ is a unitary vector normal to the surface S enclosing the three-dimensional region Ω. These are the *interatomic surfaces* (IAS) in QTAIM and represent the boundaries of the topological atoms. They are also called *zero-flux surfaces*, since no density flux crosses them, what entails interesting consequences that shall be described in more detail in following sections.

Bader's atoms are therefore described by the weight function

$$w_A^B(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \in \Omega_A \\ 0 & \text{elsewhere,} \end{cases} \tag{1.53}$$

which yields an exhaustive (i.e., $\cup_A \Omega_A = \mathbb{R}^3$ and $\cap_A \Omega_A = \emptyset$), non-fuzzy atomic decomposition of the three-dimensional space. The Voronoi partition presents an equivalent weight function, but as a result of the previous definition of the QTAIM boundaries, the size of Bader's atoms depends on how electron density is distributed in space, not only depending on the relative positions of the atomic nuclei. Consequently, the real space picture of topological atoms differs to a large extent from that provided by Voronoi (though it can be attenuated by adapting Voronoi cells by means of chemical size measures such as van der Waals radii [23]). By considering again the example of the water molecule (Fig. 1.2), under QTAIM the electron deficient hydrogen atoms compress a smaller domain, whereas the electron rich and highly electronegative oxygen dominates a larger portion of space that under a Voronoi decompostion is shared with hydrogens.

The so-called topological atoms are therefore parameter-free; they are built upon

the shape, in topological terms, of the electron density, and hold a straightforward chemical picture: a tiny hydrogen atom attached to an electronegative one loses its entity to a large extent, whereas its partner spans so as to acquire a more negative charge (as obtained from the integration of the total charge density within the corresponding boundaries) withdrawn from the hydrogen; two carbon atoms in a hydrocarbon trace a separatrix that divides the space equally between them, as they are equivalent from the chemical and also topological point of view. The QTAIM atoms arise from a theoretical sound approach; however, they do not appear alone in nature, but in a permanent interaction with their surrounding partners with which they exchange electrons continuously (they are *open quantum subsystems*). The atomic identity is therefore diluted and the decision of taking one or another approach to disentangle the continuous electron density so as to achieve a final partition in what chemists call atoms is just a choice. Nonetheless, this apparent ambiguity does not preclude a preferential reasoned used of one method over the others, and this is why only QTAIM atoms have been considered in this thesis, since those are more robust and arguably more physical than others.

1.5 Energy components as a link with chemical ideas

According to the Schrödinger equation, the wave function is the eigenfunction of the Hamiltonian operator, the energy of the system being its eigenvalue. Similarly to the molecular decomposition in terms of atoms, the energy can only be unequivocally split according to the energy terms in the Hamiltonian, but dispersion, induction, charge transfer or dipole-dipole interactions that cope an important part of the chemist's imaginary are not found within it. Here is where energy decomposition analyses (EDAs) [32] come into play. From the perturbative SAPT (symmetry-adapted perturbation theory) to the QCT-rooted IQA (interacting quantum atoms), the set of techniques for analysing the global energy of a system by dissecting it into chemically meaningful contributions have been very fruitful in, for example, ascribing binding of a molecular aggregate to a specific source, measuring the role played by different parts of a molecule in the global stability or in ascertaining the character of a chemical bond or more general interaction.

As described for the atomic decomposition of the molecular system, the corresponding schemes for dissecting the global energy are also based on very distinct approaches. These can be classified in three main groups, namely, perturbative approaches, supermolecule, orbital-based procedures and those rooted in real space techniques. One important feature common to all these methods is the definition of fragments [32, 33]. Those can be complete molecular entities such as two monomers that come into interaction, molecular fragments defined at will or the atoms or groups defined within the different AIM schemes. First of all, perturbation theory, although is a method of approximating both wave function and energy, it also permits the extraction of chemical information from energetics. This can be accomplished by identifying the energy corrections with physical processes. As such, the electrostatic energy between the separate molecular fragments —and, analogously, the different multipolar terms in which it can be decomposed— comes from the first order correction to the energy, the second encompassing induction and dispersion. The problem of its bare application is that it does not fulfil some requirements concerning the entanglement between fragments, that cannot be disregarded unless a long-range regime is considered. For closer distances, however, perturbation theory is, in principle, not applicable. To overcome this issue, several related procedures

have been derived, such as Stone-Hayes' intermolecular perturbation theory (IMPT) [34] or symmetry-adapted perturbation theory (SAPT), developed by Jeziorski, Szalewicz and co-workers [35]. Since they make use of different ansätze, so are their resulting energy terms. Thus, SAPT in its most common implementation is able to distinguish between electrostatic, exchange-repulsion, -induction and -dispersion, and just induction and dispersion energies. A different perspective is taken under the supermolecule approach. In this case, the total interaction energy between two monomers or fragments is intended to be decomposed basing on different intermediate states between the completely isolated species and the final arrangement whose difference determines the total interaction energy. The original procedure is due to Morokuma [36] and several improvements led to the more common versions used nowadays, such as that proposed by Bickelhaupt and Baerends [37]. This EDA provides four main contributions to the interaction energy: the preparation energy arisen from the distortion both fragments suffer when they begin to interact, the electrostatic component between the fragments in their final conformations, Pauli repulsion (similar to the exchange-repulsion term in perturbative approaches) from the antisymmetrisation of the supermolecular wave function and, finally, the orbital interaction energy comprising the final relaxation of the electron clouds in the presence of one another. These are orbital-based EDAs, since they resort to orbitals to calculate the different quantities. This fact can be viewed as chemically intuitive or, on the contrary, artificial. On the one hand, such an interpretation assigns electrons to interacting fragments, even when those are placed close together. Since electrons are indistinguishable, no assignment done in isolation is maintained in interaction with other species. Similar considerations deserve the orbitals. In the supermolecule, those are constructed from the contributions of the different atoms and a satisfactory assignment is not always possible, especially when the MOs are delocalised and one pretends to associate them to only one fragment, even though they have similar contributions from the other. Therefore, the definition of fragments on which these EDAs are based may be controversial [32]. The last energy partition comes from a real space analysis: the interacting quantum atoms (IQA) energy decomposition method [38]. In contrast to the previous schemes, this is an orbital-invariant, reference-free technique to decompose the total energy into atomic or group contributions. Not only does it perform an atomic partitioning of the energy (according to the desired AIM scheme), but it also provides a separation of the different sources of the energy in a physically sound way. IQA is based on the real space partitioning of the so-called *reduced density matrices* (RDMs), mathematical objects through which all the energy terms contained in the Hamiltonian can be calculated. Moreover, they allow for the further decomposition of the interelectronic potential energy into the classical, uncorrelated Coulomb product of independent densities and that due to their interdependence according to their quantum nature, what constitutes a direct link to either the ionicity or covalency of a given bond or non-bonding interaction.

1.6   New horizons in the interacting quantum atoms energy partition

The previous two sections have shown that, although there is no unique way of decomposing both the global, molecular wave function (or related functions such as the electron density) into atomic subunits and the energy of the system into chemically meaningful terms, especially those ascribed to specific parts of the molecule,

the QCT provides a tempting and coherent choice for addressing these problems. As commented before, quantum chemical topology techniques are orbital-invariant, since in contrast to orbital-based techniques, the QCT only needs a proper wave function through which to calculate its particular descriptors, regardless of the set of orbitals such a wave function is built upon.

Within this conceptual framework, the previously-mentioned interacting quantum atoms energy decomposition has been shown as a valuable tool for shedding light in many different situations. To cite a few from a recent review by Guevara-Vela et al. [38], IQA has been capable of unveiling the role played by electrostatics and covalency in different bond situations, from the most typical hydrogen bonds (HB) to more controversial H-H bonds in planar biphenyl or non-covalent interactions (NCIs) such as anion-π, halogen or tetrel bonding. It has also been employed as following the reaction path to disclose the mechanism of different organic reactions, as well as to provide a real-space picture of steric repulsion. Indeed, IQA is not limited to ground states, and as such it has been applied to excited states, too. Nevertheless, and although its popularisation has increased enormously in the past years, there is still room for improvement. The quantum mechanical methods that supply the wave functions from which the RDMs are calculated have been widened since the initial restriction to single-determinant wave functions (as shall be further explained in the next Chapter) and the use of multiconfigurational methods as well as density functional theory (DFT) has been enabled. The set of methods compatible with IQA is nowadays wide, but the treatment of medium-large systems is still challenging. In order to bypass this issue, some tricks have been proposed to reduce the high computational demands the calculations require. As such, localisation techniques and the specific selection of a reduced number of atoms of interest in a larger molecular system provide promising solutions to this major problem.

One field of computational chemistry research that has shown reluctant to the application of IQA thus far is that corresponding to biochemical systems. Their usual large size and the frequent need to also model the surrounding environment, such as the solvent that in many occasions is determinant in the properties a target biomolecule exhibits, have precluded a wider application of this energetic analysis. The present thesis, however, aims to contribute to a turning point in the underuse of IQA in the solution of biochemical problems. The selected systems for which it has been applied constitute valuable examples of the information IQA is capable of providing. However, before addressing biochemically related instances, the extension of the combined IQA-DFT methodology is mandatory for future applications where electron correlation cannot be disregarded. As such, different density functionals have been evaluated in the description of hydrogen-bonded and cooperative systems as water clusters are, what may serve as a benchmark for larger HB-featuring organic molecules. Following this first work is the study of fluorine conformational effects in small and medium-sized organic molecules, that serves to deepen in the understanding of the effect different substituents in interaction with fluorine present, what might be exploited in future design of shape-controlled bioactive peptides. Next, the inclusion of solvation effects by means of a continuous solvation model allows for the IQA decomposition of the corresponding solvation energies into atomic and group contributions. To close this research line, a final study is devoted to the IQA scrutiny of intermolecular electrostatic interactions and the so-called charge penetration effect, that reveals a direct link between the electrostatic energy, of great importance in biochemical modelling, and the IQA descriptors, what permits the subsequent assessment of the energetic effect the interpenetration of molecular densities provokes.

An effort for the extension and a wider application of this methodology to bio-chemical problems is presented. But, what is more, the thesis is intended to also encompass other relevant instances related to the IQA field of expertise: chemical bonding. Accordingly, a second grand direction is towards the assessment of controversial non-covalent interactions such as beryllium, halogen and charge-shift bonding, where IQA, as well as other QCT techniques, may shed new light on the conundrum these bonding classes entail.

Chapter 2

Methodology

This second Chapter aims at introducing the main computational chemistry methods used in the development of the present thesis. Since it is devoted to the wave function analysis and, especially, to the energy decomposition that follows the acquisition of a proper molecular wave function, the first part of the Methodology section focuses on the different methods for obtaining such wave functions. A second part traverses some quantum chemical topology tools for the subsequent analyses, with special attention to the interacting quantum atoms energy decomposition, whose capability of fruitfully studying chemical phenomena on a firm physical basis, regardless of the method used to obtain the wave function, is aimed to be remarked.

2.1 Initial considerations to electronic structure calculations

Henceforth, the atomic unit system will be utilised. This way, the electron mass (m_e), electron charge (e), reduced Planck constant ($\hbar$), Bohr radius (a_0), the atomic unit of energy (Hartree) or the vacuum permittivity (ε_0) are the basis of each corresponding magnitude, thus equalling a unit (the whole factor $4\pi\varepsilon_0$ in the case of the permittivity) [19]. Similarly, since no time dependence will be considered, capital Ψ letters will denote the total wave functions of the systems.

2.1.1 The Born-Oppenheimer approximation

Chemical systems are composed of atoms, and atoms, in turn, are built upon nuclei and electrons. Both constituent particles are interacting mutually, but there exists a major difference that enables a distinct treatment of the two: the large difference in mass. An electron is about 1836 times lighter than a proton, and a single proton is the simplest nucleus. When progressing in the periodic table, more and more neutrons are added, each being 1839 times heavier than electrons. Such huge ratio between the masses of the two particles translates into different orders of magnitude between the electronic kinetic energy and that of nuclei for ground states [9]. This fact was exploited by Born and Oppenheimer, who inaugurated an approximation central to analysing chemical systems and calculating their properties.

The non-relativistic, time-independent Hamiltonian operator associated to a given molecule composed by M atoms —and so nuclei— with N electrons reads

$$
\begin{aligned}
\hat{H} &= -\sum_{\alpha=1}^{M} \frac{1}{2m_\alpha} \nabla_\alpha^2 - \sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 \\
&\quad + \sum_{\alpha}^{M-1} \sum_{\beta>\alpha}^{M} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} + \sum_{\alpha}^{M} \sum_{i}^{N} \frac{Z_\alpha}{r_{\alpha i}} + \sum_{i}^{N-1} \sum_{j>i}^{N} \frac{1}{r_{ij}} \\
&= \hat{T}_n + \hat{T}_e + \hat{V}_{nn} + \hat{V}_{ne} + \hat{V}_{ee},
\end{aligned} \tag{2.1}
$$

where $\hat{T}_n$ and $\hat{T}_e$ represent the nuclear and electronic kinetic energies, respectively, and $\hat{V}_{nn}$, $\hat{V}_{ne}$ and $\hat{V}_{ee}$, the nucleus-nucleus, nucleus-electron and electron-electron interaction energies. The Born-Oppenheimer (BO) approximation takes advantage of the very fast electrons in comparison with the nuclei and decouples both movements. Therefore, for ground and not very high excited states, it is a good approximation to consider electrons as moving around fixed nuclear positions and their dependency as parametrical. The whole wave function can therefore be decomposed as a product of a nuclear wave function Ψ_n (that describes translation, rotation and vibration) and an electronic one Ψ_e:

$$
\Psi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}) = \Psi_n(\{\mathbf{R}_\alpha\}) \Psi_e(\{\mathbf{r}_i\}; \{\mathbf{R}_\alpha\}), \tag{2.2}
$$

where $\{\mathbf{r}_i\}$ and $\{\mathbf{R}_\alpha\}$ denote the set of electronic and nuclear coordinates, respectively. Accordingly, the Hamiltonian is split and two Schrödinger equations are to be solved for both the electrons in the electric field created by fixed nuclei and nuclei immersed in a spatial electron distribution. The electronic equation reads

$$
\hat{H}_e \Psi_e(\{\mathbf{r}_i\}; \{\mathbf{R}_\alpha\}) = E_e(\{\mathbf{R}_\alpha\}) \Psi_e(\{\mathbf{r}_i\}; \{\mathbf{R}_\alpha\}), \tag{2.3}
$$

where

$$
\hat{H}_e = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee}. \tag{2.4}
$$

By adding the internuclear repulsion V_{nn}, the total energy for fixed nuclei is obtained:

$$
U(\{\mathbf{R}_\alpha\}) = E_e(\{\mathbf{R}_\alpha\}) + V_{nn}. \tag{2.5}
$$

$U(\{\mathbf{R}_\alpha\})$ therefore comprises all but the nuclear kinetic energy term from the molecular Hamiltonian and represents the total energy for a given nuclear arrangement. Obtaining $U(\{\mathbf{R}_\alpha\})$ for different nuclear positions gives rise to a potential energy landscape or potential energy (hyper)surface (PES). In reality, when solving the electronic equation, not only is a solution obtained, but a set of wave functions and associated energies accounting for different electronic states. Whenever the PESs are separated enough and do not cross with one another, the BO approximation is applicable; otherwise it fails in describing correctly the systems under consideration.

The second equation is the nuclear Schrödinger equation

$$
\hat{H}_n \Psi_n(\{\mathbf{R}_\alpha\}) = E_n \Psi_n(\{\mathbf{R}_\alpha\}), \tag{2.6}
$$

where the Hamiltonian operator results from the addition of the nuclear kinetic energy operator to the previous $U(\{\mathbf{R}_\alpha\})$:

$$
\hat{H}_n = \hat{T}_n + U(\{\mathbf{R}_\alpha\}). \tag{2.7}
$$

As done with the hydrogenic atoms, the nuclear wave function can be transformed into a product of a wave function describing the translational movement of the molecule under consideration (as accounted for by its centre of mass), and another comprising the internal motion, that is, rotation and vibration. Only when the translational motion is decoupled, the molecular Hamiltonian presents a finite and discrete spectrum.

2.1.2 The many electron wave function

Chemical reactivity, spectroscopic properties related with ultraviolet or visible light or band structure in solids and their conductor-insulator character are properties that can be determined under the fixed nuclei approximation. Indeed, the calculation of the so-called electronic structure is central to quantum chemistry. To this end, a general many-electron wave function (that we shall simply call Ψ, without explicit reference to electrons) is to be calculated, that must fulfil a number of conditions.

The electronic wave function $\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)$ is parametrically dependent upon the nuclear positions (hence the current notation without explicit reference to them), but is a function of the spatial coordinates of all electrons $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N$ plus their spin variable $\sigma_1, \sigma_2, ..., \sigma_N$, giving rise to the joint spatial and spin coordinates $\mathbf{x}_i = (\mathbf{r}_i, \sigma_i)$. The spin is an intrinsic angular momentum of subatomic particles sometimes mistakenly regarded as coming from the rotation (spinning) about some particular axis those particles would experience, but it is actually a purely quantum mechanical effect that is better explained within quantum field theory [39, 40]. In fact, in the non-relativistic formulation used so far, the existence of the spin angular momentum is postulated so as to being able to explain the fine structure of atomic spectra. More natural is its appearance in the relativistic formulation proposed by Dirac when combining quantum mechanics and special relativity [10, 14], though such a framework lies beyond the scope of this thesis.

In atomic units ($\hbar$), the spin can take values $0, \frac{1}{2}, 1, \frac{3}{2}, ...$ and, since it is inherent to the elementary particle under consideration, these can be classified according to their associated spin value. All the particles chemists deal with (electrons, protons and neutrons) present a spin of $\frac{1}{2}$. Others, such as π-mesons (pions), present a zero-valued spin, while that of photons is one [9, 41]. Besides the collective behaviour of ensembles of these particles, the classification in terms of their either integer or half-integer spin also imposes some requirements in their associated wave functions and the incorporation of a fourth quantum number m_s (that can take $+\frac{1}{2}$ or $-\frac{1}{2}$ values in the case of electrons according to their projection along the Z-axis) to the previous three from Eq. 1.18.

Electrons are indistinguishable particles: one cannot label any as being electron 1, 2 or 15 and follow its trajectory, as this would violate the uncertainty principle. They are identical, and thus, indistinguishable. This fact is a property that must somehow be manifested in the wave function describing their behaviour. In fact, when two identical particles, as in an N-electron system, are interchanged (that is, their coordinates, both spatial and spin) the probability density must not be affected, and thus, the squared-norm of the wave function must be kept:

$$|\Psi(\mathbf{x}_1, ..., \mathbf{x}_i, ..., \mathbf{x}_j, ..., \mathbf{x}_N)|^2 = |\Psi(\mathbf{x}_1, ..., \mathbf{x}_j, ..., \mathbf{x}_i, ..., \mathbf{x}_N)|^2. \tag{2.8}$$

The previous requirement is equally fulfilled by two types of wave functions: those that are symmetric with respect to the coordinates interchange and those that

antisymmetric, that is, those whose sign changes when permuting two particle positions and spins [14]. The former is the case of *bosons*, that present also a characteristic integer-valued spin; the electrons, in turn, are *fermions*, a kind of particles that, in addition to their half-integer spin, are described by a wave function that must obey

$$\Psi(\mathbf{x}_1, ..., \mathbf{x}_i, ..., \mathbf{x}_j, ..., \mathbf{x}_N) = -\Psi(\mathbf{x}_1, ..., \mathbf{x}_j, ..., \mathbf{x}_i, ..., \mathbf{x}_N). \tag{2.9}$$

A direct consequence of the antisymmetry of the electronic wave function is the so-called Pauli exclusion principle, that states that no two electrons can pertain to the same state whenever the wave function is represented by a set of monoelectronic functions (i.e., orbitals). The many electron wave function must therefore be continuous, well-behaved, vanishing at infinity and must also incorporate the antisymmetry condition and fulfil the exclusion principle.

2.1.3   Hartree product

The basic requirements that an N-electron wave function must fulfil have been presented. However, no indication has been provided yet to construct such a wave function. As seen in Section 1.3.2, a chemical system composed of more than one electron precludes the Schrödinger equation from being solved analytically due to the interelectronic repulsion term. The simplest approximation is thus omitting this term and considering the system as composed of non-interacting electrons [15]. In such a case, the Hamiltonian can be built upon N monoelectronic Hamiltonians as

$$\hat{H} = \sum_{i=1}^{N} \hat{h}(i), \tag{2.10}$$

where $\hat{h}(i) = \hat{T}(i) + \hat{V}_{ne}(i)$. This leads to N independent eigenvalue equations for which the solutions $\{\psi_j\}$ are available:

$$\hat{h}(i)\psi_j(\mathbf{x}_i) = \varepsilon_j \psi_j(\mathbf{x}_i). \tag{2.11}$$

(Note that the labelling of functions $\{\psi_j\}$ and electrons $\{i\}$ go separately, since each of them can be reallocated in one or another function.) Due to Pauli exclusion principle, each electron must be described by its own wave function $\psi_j(\mathbf{x}_i)$ different to the rest. These are termed *spin-orbitals* and equal the product of a monoelectronic spatial wave function (that is, the orbital) $\phi(\mathbf{r})$ and its spin part $s(\sigma)$,

$$\psi_j(\mathbf{x}_i) = \phi_j(\mathbf{r}_i)s_j(\sigma_i). \tag{2.12}$$

Note that the differing aspect between two monoelectronic wave functions can be either the spatial part, the spin state or both, what implies that two electrons are allowed to share the same spatial orbital whenever they present an opposite spin so that the Pauli exclusion principle is fulfilled. The spin function $s(\sigma)$ can be either α or β, according to the m_s value of $+\frac{1}{2}$ or $-\frac{1}{2}$, respectively. An important property of these functions is that they are orthonormal, that is, their inner product yields one for same-spin functions, since they are normalised, or zero if they represent opposite spin states. This property can be cast into the relation $\langle s_j(\sigma_i)|s_k(\sigma_i)\rangle = \delta_{jk}$.

Consequently with the use of independent one-electron Hamiltonians, the total wave function corresponds to the product of N independent, one-electron wave functions:

$$\Psi^{HP}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) = \psi_i(\mathbf{x}_1)\psi_j(\mathbf{x}_2) \cdots \psi_k(\mathbf{x}_N). \tag{2.13}$$

Such a wave function is called a *Hartree product* and presents an associated eigenvalue

$$E = \sum_{j=1}^{N} \varepsilon_j. \tag{2.14}$$

Of course, this is an imaginary system; electrons are charged particles and charged particles do interact through the Coulomb term in the Hamiltonian. A Hartree product wave function therefore lacks entirely of electron correlation, since the probability density can be decomposed likewise into a product of independent probability densities; in other words, the movement of one electron does not alter that of the others, no matter how close the eventually become. Moreover, the previous assumptions translate into a distinguishable electron model, what contradicts an essential condition that identical particles must satisfy.

2.1.4 Slater determinants

A correct wave function fulfilling the required conditions is obtained by introducing antisymmetry. Let us consider the case of a two-electron system described by the Hartree product of two spinorbitals ψ_1 and ψ_2. One may place electron 1 to ψ_1 and electron 2 to ψ_2 or, conversely, e^- 1 in ψ_2 and e^- 2 in ψ_1. Any of these two solutions is acceptable according to a Hartreee product wave function, but in both cases electrons are distinguishable. If one considers a linear combination of the two possibilities a correct 2-electron wave function may be obtained. In this way, $\Psi(\mathbf{x}_1, \mathbf{x}_2) = 2^{-1/2} \left(\psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2) - \psi_1(\mathbf{x}_2)\psi_2(\mathbf{x}_1) \right)$ (where $2^{-1/2}$ is a normalisation factor) ensures electron indistinguishability —each electron can be in one or another spinorbital— and is antisymmetric upon electron interchange.

The previous 2-electron wave function can be rewritten as a determinant,

$$\Psi^{SD}(\mathbf{x}_1, \mathbf{x}_2) = 2^{-1/2} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) \end{vmatrix}, \tag{2.15}$$

and it leads to a generalisation to N-electron systems as

$$\Psi^{SD}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) = (N!)^{-1/2} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \cdots & \psi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & \cdots & \psi_N(\mathbf{x}_N) \end{vmatrix}. \tag{2.16}$$

These are the so-called Slater determinants (SDs), a fundamental piece in electronic structure methods, and can be succinctly expressed in terms of the main diagonal elements as

$$\Psi^{SD}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) \equiv |\psi_1\psi_2 \cdots \psi_N| \equiv |\psi_1\psi_2 \cdots \psi_N\rangle \tag{2.17}$$

with either a determinantal expression or a ket notation where the normalisation of the wave function is implicit.

There are two important consequences arising from the properties of determinants. On the one hand, the Pauli exclusion principle is fulfilled since a determinant having two equivalent columns (two electrons occupying the same spinorbitals) is

zero. Moreover, the interchange of one row by another (change in electron placement) does not alter the result of the determinant, but its sign (the functions are antisymmetric upon particle interchange).

Besides the basic properties of SDs, these wave functions no longer represent a system of totally uncorrelated particles. When using an SD, the so-called *exchange* (x) or Fermi correlation is introduced [42]. Electron correlation can be classified in two kinds: that corresponding to the interdependence of parallel-spin electrons and that coming from their charged nature. The first kind is implicit in an SD as it ensures that two electrons with same spin cannot be in the same point of space and are very unlikely to be close. This is a consequence of the diminished probability of finding another electron with same spin than a reference one in its vicinity, which is measured by the *Fermi hole*. By contrast, two opposite-spin electrons can in principle occupy the same position in space, although they are repelled by the Coulomb force. This kind of correlation (what is usually understood as *electron correlation*) is accounted for by the *Coulomb hole*, absent in a simple SD representation.

2.1.5 Molecular orbitals and basis functions

The general form of a many-electron wave function fulfilling the basic necessary conditions has been introduced as a Slater determinant, which are constructed from a set of spinorbitals. However, no indication has been given about how those spinorbitals should look like. At this point it becomes worthwhile remarking the fact that there is no unique recipe to calculate neither those single-electron functions nor any general system's wave function and, as could have been averted from these and the prior pages, the type of system under consideration and the approximations taken to address the specific problem of finding a suitable, by definition, non-exact (multielectron) wave function may lead to different kinds of functions yielding all of them accurate results, but obtained by rather different procedures.

The focus of this thesis has been on molecular systems and their properties and associated theoretical grounds. Therefore, non-relativistic, electronic structure calculations have been the means for obtaining the desired data through which to solve the different issues addressed herein. Thus, in order to obtain a proper molecular wave function, the methods used have exploited the fact that molecules are a collection of atoms, and that when two atoms come into contact, the internuclear region undergoes an accumulation of electron density, thus forming a chemical bond. This can be viewed as a constructive interference between the *waves* of each atom [14], leading to a very tempting way of constructing the spatial orbitals as linear combinations of the well-known atomic orbitals. Such a procedure is the basis of the molecular orbital (MO) theory, that, as commented in Section 1.3.3, is the preferred method due to its simply and effective implementation in electronic structure codes, leading to accurate results with an efficient use of computational resources. A further justification of the use of molecular orbitals relies on an observation in the only molecule for which the BO approximation suffices to solve the Schrödinger equation exactly: the hydrogen molecule-ion H_2^+ (and analogous). When its only electron approaches one nucleus, its Hamiltonian can be approximated as neglecting the dependence on the attraction to the other nucleus. This leads to a Hamiltonian similar to that of a hydrogen atom, whose wave function will resemble a hydrogen-like orbital. The fact that it occurs equivalently for one or another nucleus is captured by a linear combination of the two atomic orbitals.

Thus, in MO theory, the set of spinorbitals that gives rise to the SD makes use of spatial orbitals that are spanned over the molecule as a linear combination of atomic

orbitals (LCAO) $\{\chi_\nu(\mathbf{r})\}$ that constitute the basis set of the expansion [19]:

$$\phi_i(\mathbf{r}) = \sum_\nu^K c_{\nu i}\chi_\nu(\mathbf{r}). \tag{2.18}$$

In this way, an energy minimisation with respect to the coefficients of the expansion $\{c_{\nu i}\}$ would lead to the set of MOs describing the electronic structure of the system, as explained in Section 1.3.2. Ideally, an infinite number of atom-centred basis functions should be available for a precise recreation of any molecular orbital (in this case, the basis set is said to be *complete*). However, this is not possible and the size of the basis set must be finite. The type of functions comprising the basis set is not restricted to a particular one or another, and any type can in principle be used: exponential, Gaussian, polynomial, plane waves, etc. The only requisite is that the chosen type be congruent with the physics of the problem. However, from a practical point of view, the chosen set of basis functions must reduce the computational cost as much as possible, while assuring accurate results. Thus, hydrogenic and other related exponential functions such as Slater-type orbitals (STOs), with a functional form

$$\chi_{\zeta n l m_l}^{STO}(r, \theta, \varphi) = NY_{lm_l}(\theta, \varphi)r^{n-1}e^{-\zeta r} \tag{2.19}$$

mimicking exact hydrogen orbitals, can be a good choice from a theoretical point of view, and a small number may give very accurate results. The problem in this case is the treatment of integrals (that shall be presented in the next Section), for which no analytical solution is available and, thus, costly numerical integrations have to be invoked. On the contrary, Gaussian-type orbitals (GTOs)

$$\chi_{\zeta n l m_l}^{GTO}(r, \theta, \varphi) = NY_{lm_l}(\theta, \varphi)r^{2n-2-l}e^{-\zeta r^2} \tag{2.20}$$

are much easier to handle and the integrals in which they are involved frequently present an analytical solution. The problem of this kind of functions emerges as a consequence of their poorer ability to describe the cusp of the wave function on nuclei and their appropriate decay, what makes the use of a higher number of functions with different angular momentum values and exponents necessary. As another example, periodic systems can be well described in terms of periodic functions such as plane waves, that are much more tractable, especially when a large number of atoms (usually also with a large number of electrons each one) is involved.

Since the systems inspected in this thesis are molecular ones with usually atoms of representative elements from the first rows of the periodic table, GTO-based basis sets have been the choice for the electronic structure calculations.

2.2 *Ab initio* methods

The solution of the Schrödinger equation can be achieved with the only previous assumptions, without resorting to additional data other than the fundamental constants and the atomic numbers [14, 19]. Methods of this kind are named *ab initio*, and will be the focus of this Section. By contrast, those that also incorporate adjustable parameters to experimental data are called *semiempirical* and are intended to simplify costly integral calculations that may become a bottleneck for large systems.

The present thesis has dealt with small and medium-sized molecules, and a pure, *ab initio* approach has been preferred. This Section is devoted to the introduction to this family of electronic structure methods.

2.2.1   Hartree-Fock

At the heart of most ab initio methods is the Hartree-Fock (HF) approximation. The ansatz behind HF is the simple use of an SD as the approximate wave function with which to solve the Schrödinger equation [15]. By summarising, this procedure aims to replace the N-body problem by N one-body ones, each of the electrons perceiving an average field created by the fixed nuclei and the averaged positions of the rest of the electrons. This model greatly simplifies the problem at hand, but is also unable to correctly model systems where correlation energy acquires relevance.

In HF there is no approximation in the Hamiltonian operator, that is hence exact and differs from that of the previously introduced independent particle model (the so-called *core Hamiltonian*) by the interelectronic Coulomb operator r_{ij}^{-1}:

$$\hat{H} = \sum_i \hat{h}(i) + \frac{1}{2} \sum_{i \neq j} r_{ij}^{-1}, \qquad (2.21)$$

where the summation runs over the N electron indices i. The approximation is certainly introduced in the form of the trial wave function, that is an SD, and the procedure of solving the Schrödinger equation is based on the variational method. Accordingly, the best trial wave function is the one that minimises the energy. Therefore, when it comes to an SD wave function, it is equivalent to finding the set of spinorbitals $\{\psi_i\}$ for which the (ground state) energy is minimal. This optimisation procedure is subject to the constraint that the spinorbitals be orthonormal, that is, $\langle \psi_i | \psi_j \rangle = \delta_{ij}$, what leads to the HF integro-differential equations, as will be explained in the following lines.

With an SD wave function, the expectation value of the Hamiltonian for the GS takes the form

$$E_0 = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle = \sum_i \left[\psi_i | \hat{h} | \psi_i \right] + \frac{1}{2} \sum_{ij} \left(\left[\psi_i \psi_i | \psi_j \psi_j \right] - \left[\psi_i \psi_j | \psi_j \psi_i \right] \right), \qquad (2.22)$$

where the Slater-Condon rules have been used to reach the final expression.[1]

In the prior equation a specific notation for electronic integrals has been used. For one-electron integrals, it is equivalent to using bra-kets, thus

$$\left[\psi_i | \hat{h} | \psi_i \right] \equiv \langle \psi_i | \hat{h} | \psi_i \rangle = h_{ii} = \int \psi_i^*(\mathbf{x}_1) \hat{h}(1) \psi_i(\mathbf{x}_1) d\mathbf{x}_1, \qquad (2.23)$$

but in the case of bielectronic integrals, spinorbitals at one or another side refer to the same electron and, consequently,

$$\left[\psi_i \psi_j | \psi_k \psi_l \right] \equiv \langle \psi_i \psi_k | \psi_j \psi_l \rangle = \int \psi_i^*(\mathbf{x}_1) \psi_j(\mathbf{x}_1) r_{12}^{-1} \psi_k^*(\mathbf{x}_2) \psi_l(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2, \qquad (2.24)$$

which might be a more convenient representation within the HF theory, as shall be seen below. Note that the previous equation refers to any two electrons 1 and 2. Since they are indistinguishable, we shall adopt this numbering for differentiating the interacting pair of electrons from the spinorbitals in which they are placed.

---

[1]The Slater-Condon rules provide simple expressions for the matrix elements involving SDs. These are based on the orthonormality of the spinorbitals and how either a monoelectronic operator (such as $\hat{h}(i)$) or a bielectronic one (the Coulomb operator r_{ij}^{-1}) act on them (e.g., the spinorbitals are eigenfunctions of the core Hamiltonian, what serves to simplify expressions). For further information, Ref. 15, pp. 68-80, contains full details.

To derive the HF equations the functional variation is used [14]. E_0 is said to be a *functional* of $\{\psi_i\}$, since it is a function of a function, in this case, the SD that in turn is composed of spinorbitals. Resorting to the variational principle, the *best* set of spinorbitals is that for which $E_0[\{\psi_i\}]$ is minimum and approaches the true ground state energy. However, there is a constraint that the set of spinorbitals giving rise to Ψ_0 must fulfil: the set $\{\psi_i\}$ must be orthonormal (i.e., $[\psi_i|\psi_j] \equiv \langle \psi_i|\psi_j \rangle = \delta_{ij}$). Such a constrained minimisation can be solved by Lagrange's method of undetermined multipliers.

In search for the appropriate set of spinorbitals, the Lagrange functional is constructed as

$$\mathcal{L}[\{\psi_i\}] = E_0[\{\psi_i\}] - \sum_{ij} \lambda_{ji}([\psi_i|\psi_j] - \delta_{ij}), \qquad (2.25)$$

where λ_{ji} correspond to the undetermined multipliers. An arbitrary infinitesimal change in ψ must not produce any change in $\mathcal{L}$ and hence $\delta\mathcal{L}[\{\psi_i\}] = 0$. By applying this condition to Eq. 2.25 and expanding the energy term following expression 2.22, one is led to

$$
\begin{aligned}
\delta\mathcal{L} &= \sum_i \left[\delta\psi_i|\hat{h}|\psi_i\right] \\
&+ \sum_{ij} \left(\left[\delta\psi_i\psi_i|\psi_j\psi_j\right] - \left[\delta\psi_i\psi_j|\psi_j\psi_i\right] - \lambda_{ji}\left[\delta\psi_i|\psi_j\right] \right) + \text{cc} = 0, \quad (2.26)
\end{aligned}
$$

after recognising the different complex conjugates (cc) present and that $[\psi_i\psi_j|\psi_k\psi_l] = [\psi_k\psi_l|\psi_i\psi_j]$.

The previous expression can be factorised and simplified after introducing two operators of paramount importance in HF: the Coulomb ($\hat{J}_j$) and exchange ($\hat{K}_j$) operators.

On the one hand, the Coulomb operator as applied to a given spinorbital $\psi_i(\mathbf{x}_1)$,

$$\hat{J}_j(1)\psi_i(\mathbf{x}_1) = \left[\int d\mathbf{x}_2 \psi_j^*(\mathbf{x}_2)\psi_j(\mathbf{x}_2)r_{12}^{-1}\right]\psi_i(\mathbf{x}_1), \qquad (2.27)$$

represents a one-electron potential felt by such electron 1 due to the average interaction with electron 2 over all space and spin coordinates. Its exchange counterpart, in turn, has no classical interpretation. It arises from the antisymmetry of the wave function and acts as *exchanging* electron 1 and electron 2 in spinorbitals i and j with respect to the previous Coulomb term:

$$\hat{K}_j(1)\psi_i(\mathbf{x}_1) = \left[\int d\mathbf{x}_2 \psi_j^*(\mathbf{x}_2)\psi_i(\mathbf{x}_2)r_{12}^{-1}\right]\psi_j(\mathbf{x}_1). \qquad (2.28)$$

With these operators, Eq. 2.26 can be reexpressed as

$$
\begin{aligned}
\delta\mathcal{L} &= \sum_i \int \delta\psi_i^*(\mathbf{x}_1)\left(\hat{h}(1)\psi_i(\mathbf{x}_1)\right. \\
&+ \left.\sum_j \left\{\hat{J}_j(1)\psi_i(\mathbf{x}_1) - \hat{K}_j(1)\psi_i(\mathbf{x}_1) - \lambda_{ji}\psi_j(\mathbf{x}_1)\right\}\right) + \text{cc} = 0, \quad (2.29)
\end{aligned}
$$

and, since $\delta\psi_i^*$ is arbitrary (it is not necessarily zero), each term in the parenthesis must be independently zero, leading to a set of N one-electron equations:

$$\left[\hat{h}(1) + \sum_j \left(\hat{J}_j(1) - \hat{K}_j(1)\right)\right]\psi_i(\mathbf{x}_1) = \sum_j \lambda_{ji}\psi_j(\mathbf{x}_1), \quad i = 1, 2, ..., N. \tag{2.30}$$

The operators (in squared brackets) acting on spinorbital ψ_i constitute the *Fock operator* $\hat{f}$,

$$\hat{f}(1) = \hat{h}(1) + \sum_j \left(\hat{J}_j(1) - \hat{K}_j(1)\right) = \hat{h}(1) + v^{HF}(1), \tag{2.31}$$

an effective monoelectronic Hamiltonian comprising the core-Hamiltonian $\hat{h}(1)$ and an effective one-electron potential: the Hartree-Fock potential $v^{HF}(1)$, that represents the effect the $N - 1$ remaining electrons cause, on average, on a reference electron 1.

With the definition of the Fock operator, the equations for spinorbitals become

$$\hat{f}(1)\psi_i(\mathbf{x}_1) = \sum_j \lambda_{ji}\psi_j(\mathbf{x}_1), \tag{2.32}$$

which reminds of an eigenvalue equation but it is not, since the right-hand side corresponds to a linear combination of scalars multiplying ψ_j and not only one of these. This is the manifestation of a property that SD wave functions have, that there exists a certain degree of flexibility among them so that the spinorbitals they are composed of can be mixed without changing the expectation value (e.g., the energy) of the final wave function. In fact, by applying *unitary transformations* one set of spinorbitals can be transformed into another, leaving the Fock operator unaltered (it is said to be *invariant* upon unitary transformations). One of those sets diagonalises the matrix of Lagrange multipliers λ_{ji} such that

$$\hat{f}(1)\psi_i(\mathbf{x}_1) = \varepsilon_i\psi_i(\mathbf{x}_1). \tag{2.33}$$

These are the standard Hartree-Fock equations (pseudo-eigenvalue equations due to the functional dependence of $\hat{f}(1)$ on $\{\psi_i\}$) and the set $\{\psi_i\}$ that diagonalises the matrix of Lagrange multipliers giving rise to the orbital energies ε_i are called *canonical spinorbitals*. These spinorbitals, however, tend to be delocalised over the entire molecule or several atoms. Nonetheless, as commented a few lines above, once a set of spinorbitals minimising the energy is obtained, an infinite number of new $\{\psi_i'\}$ is available upon unitary transformations, that can be chosen so as to be more localised and giving thus rise to a more *chemical* picture of electronic structure.

The Roothaan-Hall equations

It was in 1951 when C. C. J. Roothaan and G. G. Hall independently proposed the use of a known set of basis functions with which to expand the unknown orbitals (the spatial part of the spinorbitals), thus transforming the HF differential equation into a set of algebraic equations that could be solved by standard matrix thechniques. The way this procedure is implemented differs from being applied to *closed-shell* systems —those with an even number of electrons, each pair placed in one spatial orbital— or to *open-shell* systems —with at least one unpaired electron. We shall consider the former case only.

The restricted Hartree-Fock (RHF) formalism results from the application of the HF method to closed-shell systems. A first step towards the explicit formulation of

this problem is to achieve an HF equation regardless of the spin states of the electrons. By reexpressing the HF equation separating the spatial and spin parts of the spinorbital as $\psi_i(\mathbf{x_1}) = \phi_j(\mathbf{r_1})s_j(\sigma_1)$, integration of the spin coordinates after multiplication on the left by the complex conjugate spin function $s_j^*(\sigma_1)$ reveals that the eigenvalues of the equation (the energies) remain unchanged and are characteristic of the spatial orbitals $\phi_j(\mathbf{r_1})$:

$$\left[\int d\sigma_1 s_j^*(\sigma_1) \hat{f}(\mathbf{x_1}) s_j(\sigma_1) \right] \phi_j(\mathbf{r_1}) = \varepsilon_j \phi_j(\mathbf{r_1}), \tag{2.34}$$

Letting now $\hat{f}(1) \equiv \hat{f}(\mathbf{r_1}) = \int d\sigma_1 s_j^*(\sigma_1) \hat{f}(\mathbf{x_1}) s_j(\sigma_1)$, the left-hand side term becomes

$$\begin{aligned}
\hat{f}(1)\phi_j(\mathbf{r_1}) &= \left[\int d\sigma_1 s_j^*(\sigma_1) \hat{h}(1) s_j(\sigma_1) \right] \phi_j(\mathbf{r_1}) \\
&+ \left[\sum_k \int d\sigma_1 s_j^*(\sigma_1) \left(\hat{J}_k(1) - \hat{K}_k(1) \right) s_j(\sigma_1) \right] \phi_j(\mathbf{r_1}).
\end{aligned} \tag{2.35}$$

Considering the general closed-shell system, the expansion of the former equation in different spin combinations (i.e., of $\alpha(\sigma_1)$ and $\beta(\sigma_1)$ functions) yields

$$\begin{aligned}
\hat{f}(1)\phi_j(\mathbf{r_1}) &= \hat{h}(1)\phi_j(\mathbf{r_1}) \\
&+ \sum_k^{N/2} \left(2\hat{J}_k(1) - \hat{K}_k(1) \right) \phi_j(\mathbf{r_1}) \\
&= \varepsilon_j \phi_j(\mathbf{r_1}),
\end{aligned} \tag{2.36}$$

where the summation over the N spinorbitals has been replaced by the previous one running over the $N/2$ spatial orbitals. It is important to remark that all the expanded terms corresponding to the Coulomb operator survive, whereas in the exchange counterparts half of the terms involving mixed spin states for the same electrons cancel out.

The closed-shell energy thus becomes

$$\begin{aligned}
E_0 = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle &= 2 \sum_i^{N/2} \left[\phi_i | \hat{h} | \phi_i \right] + \sum_i^{N/2} \sum_j^{N/2} \left(2 \left[\phi_i \phi_i | \psi_j \phi_j \right] - \left[\phi_i \phi_j | \phi_j \phi_i \right] \right) \\
&= 2 \sum_i^{N/2} h_{ii} + \sum_i^{N/2} \sum_j^{N/2} \left(2 J_{ij} - K_{ij} \right),
\end{aligned} \tag{2.37}$$

J_{ij} and K_{ij} being the corresponding Coulomb and exchange integrals.

Having eliminated spin, the next step is to find the appropriate set of molecular orbitals (or, more generally, spatial orbitals, since atomic electronic states can also be calculated through this formalism) fulfilling Eq. 2.36. As commented above, this equation can be solved by standard matrix methods when an expansion of the unknown MOs is used, as also anticipated in Section 2.1.5 (recall Eq. 2.18). In this way, the problem of solving the HF integro-differential equation to calculate MOs reduces to calculating the appropriate set of expansion coefficients $\{c_{vi}\}$. Hence, the HF equation becomes

$$\hat{f}(1) \sum_v c_{vi} \chi_v(\mathbf{r_1}) = \varepsilon_i \sum_v c_{vi} \chi_v(\mathbf{r_1}), \tag{2.38}$$

where the summations run over the total K basis functions. Multiplying on the left by $\chi_\mu^*(\mathbf{r}_1)$ and integrating, the integro-differential equation transforms into a matrix equation, which can be concisely represented after introducing two matrices, namely the overlap matrix $\mathbf{S}$, with elements

$$S_{\mu\nu} = \langle \chi_\mu(\mathbf{r}_1) | \chi_\nu(\mathbf{r}_1) \rangle \tag{2.39}$$

and the Fock matrix $\mathbf{F}$ —the matrix representation of the Fock operator in a given basis $\{\chi_\mu\}$—, composed of

$$F_{\mu\nu} = \langle \chi_\mu(\mathbf{r}_1) | \hat{f}(1) | \chi_\nu(\mathbf{r}_1) \rangle \tag{2.40}$$

terms.

With the previous definitions, we arrive at the integrated HF equation of the form

$$\sum_\nu F_{\mu\nu} c_{\nu i} = \varepsilon_i \sum_\nu S_{\mu\nu} c_{\nu i} \tag{2.41}$$

for all the ($N/2$ occupied and $K - N/2$ virtual) K molecular orbitals ϕ_i. These are the Roothaan-Hall equations and can be expressed in a matrix form as

$$\mathbf{FC} = \mathbf{SC}\varepsilon, \tag{2.42}$$

where $\mathbf{C}$ is the matrix of the coefficients and ε a diagonal matrix comprising the orbital energies. These are the Roothaan-Hall equations.

As commented previously, the HF equations are not true eigenvalue equations due to the dependence of the Fock operator on the set of spinorbitals used. Therefore, it is mandatory to resort to some procedure besides the diagonalisation of the corresponding matrix in its Roothaan-Hall form. Such a strategy is called the *self-consistent field* (SCF) method. It involves an initial guess of orbitals (through their coefficients giving rise to $\mathbf{C}$) from which the Fock matrix $\mathbf{F}$ can be constructed ($\mathbf{S}$, in turn, is available from the beginning since it comprises the overlap integrals between each pair of basis functions). Once it is done, $\mathbf{F}$ is diagonalised so as to obtain a new set of coefficients. With it, their associated orbital energies are achieved. But these correspond to the system where the mean field each electron experiences has been previously determined in the initial guess. The new set of coefficients therefore results in a better guess, but it is not necessarily the *best* one or, in other words, the set of coefficients that yield the minimum energy. This is why from an initial set of MOs, several steps involving the calculation and later diagonalisation of $\mathbf{F}$ to obtain new and more accurate MOs are needed. It turns out to be an iterative procedure that lasts until consistency is reached, hence the name self-consistent field; the solution of the Roothaan-Hall equations involves ameliorating the mean field created by the electronic structure of the system until it becomes constant and exact (or, more accurately, the energy changes with the previous step are within a certain small threshold that we consider negligible).

The final set of MOs is usually greater than half the number of electrons, depending on the K number of basis functions. An ordering of the orbitals in increasing energy permits their classification in *occupied*, as the $N/2$ low-lying orbitals representing the GS electronic state, and *virtual*, representing the possible states they can reach upon excitations. The ground state HF wave function Ψ^{HF} comprises the occupied orbitals only and yields an energy E^{HF} that is an upper bound to the exact GS energy E_0 of the system. One of the errors that affect such difference depends

on the adequacy of the basis set used to correctly express the HF-SCF MOs and approximate them to the exact HF ones, that would be achieved in case a complete (i.e., infinite) basis set were used. This error is known as the basis set truncation error. Another one lies in the heart of HF and cannot be solved within this method. Since HF is a mean field method, it does not account for interelectronic repulsion instantaneously, but rather in an average fashion. This is why HF is considered as a non-correlated method, although it does account for Fermi correlation between parallel-spin electrons, but not for the Coulomb one, the type of correlation usually meant by the term electron correlation in electronic structure calculations. Therefore, the HF energy in the basis set limit is taken as the reference to measure the electron correlation recovered by other post-Hartree-Fock methods, that is therefore defined as

$$E_{corr} = E_0 - E^{HF}, \tag{2.43}$$

where E^{HF} corresponds to the energy achieved in the HF limit.

2.2.2 Configuration interaction

The HF mean-field description can be enhanced with the inclusion of more configurations in which electrons are placed in higher-energy spinorbitals, thus corresponding formally to excited electronic configurations. Slater determinants can therefore be used to construct an exact wave function where the deficiencies of a single-determinant representation are corrected by the inclusion of other configurations. Accordingly, the exact GS wave function can be expanded as

$$\Psi_0 = C^{HF}\Psi^{HF} + \sum_{a,p} C_a^p \Psi_a^p + \sum_{\substack{a<b \\ p<q}} C_{ab}^{pq} \Psi_{ab}^{pq} + \sum_{\substack{a<b<c \\ p<q<r}} C_{abc}^{pqr} \Psi_{abc}^{pqr} + \cdots , \tag{2.44}$$

where the different summations in the above expansion collect all the single, double, triple... excitations as those having one, two, three... super- and subscripts in their expansion terms. This way, a singly excited determinant Ψ_a^p differs from the HF one by the interchange of two spinorbitals a and p reflecting the promotion an electron originally located in ψ_a undergoes to the virtual ψ_p:

$$\Psi_a^p = |\psi_1 \cdots \psi_{a-1} \psi_p \psi_{a+1} \cdots \psi_N|, \tag{2.45}$$

The following terms comprise simultaneous double excitations from ψ_a and ψ_b to ψ_p and ψ_q, and so on. The method in which such an expansion is used is called *configuration interaction* (CI).

In a similar fashion than the symmetry of atomic orbitals can be used to construct one or another LCAO, SDs can also be classified according to their symmetry and definite spin, and combined in *configuration state functions* (CSFs). The use of all CSFs for a given basis set is called *full CI* (FCI) and recovers all the correlation energy attainable with such a basis. The exact energy E_0 would only be available if the basis set were infinite, what is not feasible. Nor is using a large number of CSFs. In fact, the total number of determinants grows factorially with both basis functions and electrons. For a single determinant, the computational complexity of solving the SCF equations grows formally with the fourth power of the number of basis functions: $\mathcal{O}(K^4)$. Therefore, CI calculations are usually limited to the inclusion of single and double excitations (giving thus rise to the CISD method) so that the number of determinants included in the expansion (Eq. 2.44) remains tractable (CISD formally scales as $\mathcal{O}(K^6)$). Nonetheless, the penalty paid for these limited CI calculations is

the lack of *size-consistency*, what is manifested in the mismatch between, for example, the energy of a dimer AB in the non-interaction limit and the sum of the monomeric ones, that should coincide.

A related concept is that of *size-extensivity*. A size-extensive method ensures a linear scaling with the number of interacting particles (e.g., electrons). A truncated CI expansion is not size-extensive (nor is it size-consistent) and it thus recovers less and less electron correlation when enlarging the system. In the example of a dimer, a CISD calculation allows up to two simultaneous excitations in the dimer, whereas the separate monomers can independently be doubly-excited, what would imply a quadruple excitation in the complex. Such lack of flexibility prevents a good recovery of the N-fragment system's energy [15, 19].

2.2.3   Multiconfiguration and multireference methods

In a CI calculation the expansion coefficients for the CSFs in Eq. 2.44 are determined. The determinants are constructed from the occupied and virtual orbitals obtained in an HF calculation. However, only the occupied ones are optimal for a given configuration —these are the ones that contribute to the energy—, what leaves the virtual orbitals unoptimised and subject to the only condition of being orthonormal [19, 43]. In multiconfiguration self-consistent field (MCSCF) methods a double optimisation is performed. Since the energy is now also a functional of those previously unoccupied orbitals, this ansatz allows for the variational optimisation of each orbital at the same time that the expansion coefficients of the different configurations are determined in search for a minimum energy. In the closed-shell case, not only *zero* or *two* occupations are possible, but also fractional numbers in between. Accordingly, each occupation number n_i associated with an MO ϕ_i is determined as

$$n_{i,MCSCF} = \sum_j^{CSFs} C_j^2 n_{i,j}, \qquad (2.46)$$

where, since each MO does not appear in general in only one configuration, the summation averages the occupation it presents in every CSF with its relative contribution to the final wave function.

A popular MCSCF method is the *complete active space self-consistent field* (CASSCF). CASSCF restricts the excitations to a subspace of orbitals called the *active* space, where all possible excitations take place (see Figure 2.1), thus being equivalent to a reduced FCI calculation. Complementary, the *inactive* space contains those orbitals whose excitations are not expected to contribute significantly to the improved wave function. Therefore, after ordering the HF orbitals in increasing energy, the inactive space contains the lowest lying ones and the highest virtual MOs that are not expected to mix with the occupied orbitals. The active space is typically defined by the n electrons participating in the excitations and the m orbitals in which they are distributed, leading to the notation [n,m]-CASSCF specifying a given calculation. It is formed by the occupied and virtual orbitals whose mixing would recover a large fraction of the correlation energy and, consequently, CASSCF is intended to provide a size-extensive alternative to CI, thus yielding accurate results when the active space chosen is appropriate. In fact, the selection of such an active space is the main concern when performing a CASSCF calculation. The orbitals involved typically correspond to MOs with a high presence of valence AOs, but that must be carefully examined in each case.
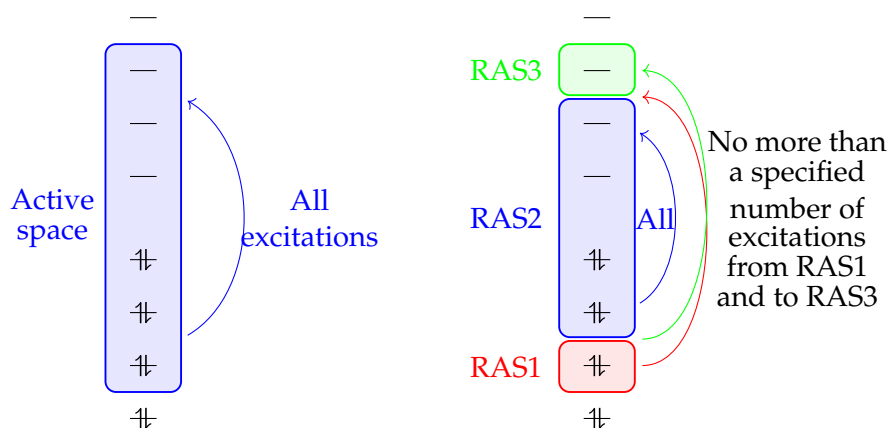
FIGURE 2.1: Schematic representation of the CASSCF (left) and RASSCF (right) procedures. Adapted from Ref. 19.

One problem inherent to the CASSCF formulation is the rapid growth of the computational requirements with the system size, since more and more orbitals must in principle be included in the active space. An alternative to reduce the size of the active space while guaranteeing a given accuracy through the inclusion of the orbitals needed for describing an enough number of excitations is the *restricted active space self-consistent field* (RASSCF) method. In this case, the active space is further split into three parts, namely, RAS1, RAS2 and RAS3. The second one, RAS2, is the equivalent to the CASSCF active space and comprises the doubly occupied MOs that participate in the FCI calculation with the corresponding virtual ones. The RAS1 section precedes the RAS2 in the active space as those orbitals that participate in it, but in a different manner: only a specified number of electrons are allowed to simultaneously participate in the excitations with the RAS2 and RAS3 spaces. Similarly, only a given number of electrons can be promoted to the RAS3 empty orbitals at the same time.

In many cases, the HF wave function provides at least a good qualitatively description of the system near equilibrium. It is not always like this and some molecules necessitate more than one CSF as a reference, i.e., when their contributions become similar. A truncated CI procedure in which the excitations are taken from more than one determinant is named *multireference configuration interaction* (MRCI). One associated problem becomes evident: the inclusion of more reference configurations leads to a higher number of excited determinants, what can be computationally very intensive. However, the size-extensivity error can be reduced significantly.

2.2.4 Møller-Plesset perturbation theory

Configuration interaction, multiconfigurational and multireference methods have the HF ground state as the starting point for a recovery of the correlation energy by means of the inclusion of excited configurations that allow the wave function to be more flexible in the minimisation process. In the Introduction, two different approaches for solving the many-body problem have been presented. On the one hand, the variational method that lies behind the previous electronic structure methods. On the other, there is perturbation theory (PT). The application of PT to Ψ^{HF} as the zeroth-order wave function (i.e., that of the reference, model system) dates back to the 1930s, the early times of quantum mechanics, and was proposed by Møller and Plesset, thus being known as Møller-Plesset perturbation theory (MPPT) [14, 15, 19].

In MPPT, the unperturbed system's ground state therefore corresponds to $\Psi_0^{(0)} = \Psi^{HF}$, with an associate energy $E_0^{(0)} = E^{HF}$. In HF, the Hamiltonian operator is exact and the main approximation behind this method is the wave function as a Slater determinant, which is an eigenfunction of

$$\hat{H}^{HF} = \sum_i \hat{f}(i) = \sum_i \left(\hat{h}(i) + v^{HF}(i) \right) = \hat{H}^0. \tag{2.47}$$

In this way, the perturbation $\hat{H}'$ that leads the model system to the true one becomes

$$\hat{H}' = \hat{H} - \hat{H}^{HF}. \tag{2.48}$$

The 0th-order energy results from

$$E_0^{(0)} = \left\langle \Psi_0^{(0)} \middle| \hat{H}^0 \middle| \Psi_0^{(0)} \right\rangle = \left\langle \Psi^{HF} \middle| \hat{H}^{HF} \middle| \Psi^{HF} \right\rangle = \sum_i \varepsilon_i, \tag{2.49}$$

whereas the addition of the first-order correction[2]

$$\begin{aligned}
E_0^{(1)} &= \left\langle \Psi^{HF} \middle| \hat{H}' \middle| \Psi^{HF} \right\rangle = \left\langle \Psi^{HF} \middle| \hat{H} - \hat{H}^{HF} \middle| \Psi^{HF} \right\rangle \\
&= \frac{1}{2} \sum_{ij} \left\langle \Psi^{HF} \middle| r_{ij}^{-1} \middle| \Psi^{HF} \right\rangle - \sum_{ij} \left\langle \Psi^{HF} \middle| \hat{J}_j(i) - \hat{K}_j(i) \middle| \Psi^{HF} \right\rangle \\
&= \frac{1}{2} \sum_{ij} \left([\psi_i \psi_i | \psi_j \psi_j] - [\psi_i \psi_j | \psi_j \psi_i] \right) \\
&\quad - \sum_{ij} \left([\psi_i \psi_i | \psi_j \psi_j] - [\psi_i \psi_j | \psi_j \psi_i] \right) \\
&= -\frac{1}{2} \sum_{ij} \left([\psi_i \psi_i | \psi_j \psi_j] - [\psi_i \psi_j | \psi_j \psi_i] \right)
\end{aligned} \tag{2.50}$$

leads to the HF energy

$$E_0^{(0)} + E_0^{(1)} = E^{HF}. \tag{2.51}$$

Thus, the first correction to the HF energy comes from second order. By recalling expression 1.37, the second-order energy correction corresponds to a summation over all states but the GS, hence comprising single, double, triple... excitations:

$$\begin{aligned}
E_0^{(2)} &= \sum_{n \neq 0} \frac{\left| \left\langle \Psi^{HF} \middle| \hat{H}' \middle| \Psi_n \right\rangle \right|^2}{E_0^{(0)} - E_n^{(0)}} \\
&= \sum_S \frac{\left| \left\langle \Psi^{HF} \middle| \hat{H}' \middle| \Psi_S \right\rangle \right|^2}{E_0^{(0)} - E_S^{(0)}} + \sum_D \frac{\left| \left\langle \Psi^{HF} \middle| \hat{H}' \middle| \Psi_D \right\rangle \right|^2}{E_0^{(0)} - E_D^{(0)}} \\
&\quad + \sum_T \frac{\left| \left\langle \Psi^{HF} \middle| \hat{H}' \middle| \Psi_T \right\rangle \right|^2}{E_0^{(0)} - E_T^{(0)}} + \cdots
\end{aligned} \tag{2.52}$$

where S states for single (Ψ_a^p), D for double (Ψ_{ab}^{pq}) and T for triple (Ψ_{abc}^{pqr}) excitations.

---

[2]To reach the final expression for $E_0^{(1)}$ the Slater-Condon rules have been used.

Examination of the singly-excited terms leads to the conclusion that they do not contribute,

$$\langle\Psi^{HF}|\hat{H}'|\Psi_S\rangle = \langle\Psi^{HF}|\hat{H}|\Psi_a^p\rangle - \langle\Psi^{HF}|\hat{H}^{HF}|\Psi_a^p\rangle = 0, \tag{2.53}$$

since the first terms is zero according to Brillouin's theorem,[3] and the second due to the fact that Ψ_a^p is an eigenfunction of $\hat{H}^{HF}$ and that the determinants are orthonormal to one another. Moreover, matrix elements comprising triple and higher-order excitations are null for 1- and 2-electron operators, as is the case of the Hamiltonian. Therefore, the only contributions to $E_0^{(2)}$ arise from double excitations.

Taking into account that

$$\begin{aligned}
\langle\Psi^{HF}|\hat{H}|\Psi_{ab}^{pq}\rangle &= \langle\Psi^{HF}|\sum_i \hat{h}(i) + \sum_{ij} r_{ij}^{-1}|\Psi_{ab}^{pq}\rangle \\
&= [\psi_a\psi_p|\psi_b\psi_q] - [\psi_a\psi_q|\psi_b\psi_p] \\
&= \langle\psi_a\psi_b|\psi_p\psi_q\rangle - \langle\psi_a\psi_b|\psi_q\psi_p\rangle \\
&= \langle\psi_a\psi_b||\psi_p\psi_q\rangle
\end{aligned} \tag{2.54}$$

(see Ref. 15 for further details about matrix elements between SDs), and that

$$\hat{H}^{HF}|\Psi_{ab}^{pq}\rangle = (E_0^{(0)} - \varepsilon_a - \varepsilon_b + \varepsilon_p + \varepsilon_q)|\Psi_{ab}^{pq}\rangle, \tag{2.55}$$

the second-order correction to the energy becomes

$$E_0^{(2)} = \frac{1}{4}\sum_{ab}^{occ}\sum_{pq}^{virt}\frac{|\langle\psi_a\psi_b||\psi_p\psi_q\rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_p - \varepsilon_q}, \tag{2.56}$$

where 1/4 has been added to avoid double counting in the summations of the occupied and the virtual spinorbitals.

In MPPT, the second-order energy correction gives rise to the MP2 method. In general, MPPT methods are labelled as MPn, where n indicates the level of truncation in the perturbative expansion of the energy. The least computationally demanding MPPT method is indeed MP2, that usually employs similar timings than HF (with a formal scaling of $\mathcal{O}(K^5)$). It is found that most of the correlation energy is hidden in double excitations, MP2 being able to recover about 80%-90% of E_{corr} [19]. Further corrections include more and more correlation energy: MP3 accounts for a 90%-95%, whereas MP4 reaches about a 95%-98% of the total E_{corr} while maintaining a usually affordable computer time ($\mathcal{O}(K^6)$ formally), similar to CISD. For many cases, MP2 yields good enough results that are not always ameliorated when following the MPn series. Since MPPT is not variational, MPn methods may yield energies above as well as below the exact energy. In the ideal case, there is a monotonical convergence from HF to increasing orders of MPn; however, this is not ensured and depending on the system, the MPn series may oscillate and not converge. Especially dramatic errors may be given when the HF wave function does not provide a close enough description of the true system (e.g., a multireference case). In any case, MPn methods are size-extensive (and consistent), and usually provide accurate results for single-reference systems in their ground states. Among all of them, MP2 represents a rather successful compromise between accuracy and computational cost [16].

---

[3]The Brillouin's theorem states that the reference Ψ^{HF} does not interact directly with singly excited determinants Ψ_S, what makes the corresponding matrix elements $\langle\Psi^{HF}|\hat{H}|\Psi_S\rangle$ vanish.

2.2.5 The coupled cluster ansatz

A different non-variational and size-extensive method that relies on excitations from the HF wave function to account for electron correlation is the coupled cluster (CC) method. CC aims at incorporating corrections of a given type to infinite order, in contrast to perturbation approaches, that include all corrections to a given order (single, double, triple, etc.). [14, 19]. The CC machinery is based on the application of the cluster operator $\hat{T}$,

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_N, \tag{2.57}$$

which comprises all possible excitations in an N-electron system (in this case, each type of excitation is denoted by a number, 1 standing for single, 2 for double and so on). Thus, the application of a given $\hat{T}_i$ operator on an HF wave function induces all ith-order excited determinants, such as

$$\hat{T}_2 \Psi^{HF} = \sum_{a<b}^{occ} \sum_{p<q}^{virt} t_{ab}^{pq} \Psi_{ab}^{pq}. \tag{2.58}$$

These t expansion coefficients are typically called amplitudes.

In coupled cluster theory, the N-electron wave function reads

$$\Psi = e^{\hat{T}} \Psi^{HF}, \tag{2.59}$$

where the exponential operator acting on Ψ^{HF} is expanded as

$$
\begin{aligned}
e^{\hat{T}} &= 1 + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{6}\hat{T}^3 + \cdots \\
&= 1 + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right) + \left(\hat{T}_3 + \hat{T}_2\hat{T}_1 + \frac{1}{6}\hat{T}_1^3\right) + \cdots,
\end{aligned} \tag{2.60}
$$

the latter version grouping all the excitation forms of a given order. As can be appreciated, not only $\hat{T}_i$ originates ith-order excited states, but different combinations of lower-order operators adding up to i also yield excited determinants of the same kind. These are named *disconnected* excitations, such as $\hat{T}_1^2$ double excitations or $\hat{T}_2\hat{T}_1$ triple ones. By contrast, $\hat{T}_2$, $\hat{T}_3$, $\hat{T}_4$... are *connected* double, triple, quadruple... excitations. The difference between the two classes stems from whether the excited electrons interact among all of them or in groups (clusters) of a lower number of electrons. In this way, a connected excitation $\hat{T}_3$ involves three excited electrons interacting simultaneously, whereas $\hat{T}_2\hat{T}_1$ implies two interacting electrons and a third non-interacting one.

The CC energy results from the solution of the corresponding Schrödinger equation:

$$\hat{H}e^{\hat{T}}|\Psi^{HF}\rangle = E e^{\hat{T}}|\Psi^{HF}\rangle \tag{2.61}$$

Multiplication on the left by $\langle \Psi^{HF}|$,

$$
\begin{aligned}
\langle \Psi^{HF}|\hat{H}e^{\hat{T}}|\Psi^{HF}\rangle &= E\langle \Psi^{HF}|e^{\hat{T}}\Psi^{HF}\rangle \\
&= E\langle \Psi^{HF}|(1 + \hat{T}_1 + \hat{T}_2 + \cdots)\Psi^{HF}\rangle,
\end{aligned} \tag{2.62}
$$

and taking into account that the different Slater determinats are orthonormal, one is led to

$$E = \langle \Psi^{HF}|\hat{H}e^{\hat{T}}|\Psi^{HF}\rangle. \tag{2.63}$$

Since the Hamiltonian is a two-electron operator, excitations higher than second-order ones can be disregarded in the expansion of the exponential. Moreover, according to Brillouin's theorem, single excitations do not mix with the HF ground state wave function. Accordingly,

$$
\begin{aligned}
E &= \langle \Psi^{HF} | \hat{H} | \left(1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right)\Psi^{HF} \rangle \\
&= \langle \Psi^{HF} | \hat{H} | \Psi^{HF} \rangle + \langle \Psi^{HF} | \hat{H} | \hat{T}_1 \Psi^{HF} \rangle \\
&\quad + \langle \Psi^{HF} | \hat{H} | \hat{T}_2 \Psi^{HF} \rangle + \frac{1}{2}\langle \Psi^{HF} | \hat{H} | \hat{T}_1^2 \Psi^{HF} \rangle \\
&= E^{HF} + \sum_{a<b}^{occ} \sum_{p<q}^{virt} \left(t_{ab}^{pq} + t_a^p t_b^q - t_a^q t_b^p\right) \langle \psi_a \psi_b || \psi_p \psi_q \rangle.
\end{aligned}
\tag{2.64}
$$

As can be appreciated, the correlation energy in coupled cluster is completely determined by the singles and doubles amplitudes as well as the two-electron integrals.

It is in the derivation of the expressions for the amplitudes where the rest of the excitations appear. However, it is again not feasible to include all excitation levels (from $\hat{T}_1$ to $\hat{T}_N$) and the expansion of the cluster operator must be truncated somewhere. Since no single excitation mixes with the HF ground state, the first method within the CC family is coupled cluster doubles (CCD), that can be improved with single excitations that do mix with doubles in the popular coupled cluster singles and doubles (CCSD) method. The inclusion of single excitations does not entail any significant increase in the computational effort (both CCD and CCSD are formally $\mathcal{O}(K^6)$ methods) and provides a more complete model. In CCD, double excitations and their powers (disconnected quadruples, sextuples, etc.) are included, while the CCSD expansion gives

$$
e^{\hat{T}_1 + \hat{T}_2} = 1 + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right) + \left(\hat{T}_2\hat{T}_1 + \frac{1}{6}\hat{T}_1^3\right) + \left(\frac{1}{2}\hat{T}_2^2 + \frac{1}{2}\hat{T}_2\hat{T}_1^2 + \frac{1}{24}\hat{T}_1^4\right) + \cdots ,
\tag{2.65}
$$

including all combinations of singles and doubles until order N (excitations involving more than N electrons are of course not considered). In this way, and contrarily to CI, a truncated CC method is size-extensive since, provided one or several types of excitations (S, D, T...), all possible orders are generated. It is worthwhile mentioning that connected and disconnected forms of same-order excitations are not equivalent and in principle all of them should be included to achieve full accuracy. Nonetheless, very good approximations can be made by means of disconnected terms in, for example, accounting for quadruple excitations by two disconnected doubles, etc.

2.3 Density Functional Theory

Wave function-based methods rely on the Hartree-Fock approximation, where the target wave function is shaped as a Slater determinant and electron correlation is not included by definition (i.e., each electron experiences the average effect of the others rather than an instantaneous repulsion with each of the remaining electrons in the system). In this way, correlation effects are accounted for *a posteriori*, and improvements of this description involve the mixing of the HF determinant with different excitations to allow the wave function to relax and better adapt to the real situation at hand. Whereas HF may be appealing due to the simplicity of the model and its computational efficiency, its further corrections become untractable for large systems, especially when a high degree of correlation is aimed to be recovered.

At this stage, one may wonder whether other possible alternatives to model chemical systems exist. The answer is yes, and that is the purpose of density functional theory (DFT). Whereas the complicated wave function is the milestone of *ab initio* methods, a much simpler electron density (i.e., the electronic probability density) is the central object in DFT. As has been explained previously, the wave function of a system contains all the information concerning it. If such a function is found, one gains, in principle, access to all the properties that characterise the system. However, the N-electron wave function depends on $4N$ variables ($3N$ spatial ones plus N accounting for spin), what makes it increase enormously in complexity when increasing the number of electrons. Moreover, this mathematical object does not reside in the real space, but in Hilbert's: the wave function is not an observable. A rather different approach is followed by DFT, that relies on the electron density $\rho(\mathbf{r})$ as the central quantity through which to calculate all the properties of interest, what becomes very advantageous from the computational point of view [42].

2.3.1 The electron density

The electron density is simply a probability density function describing the (continuous) distribution of electrons in space (as advanced in Section 1.2). It represents the probability per volume element of finding one of the N electrons of a system, following the probabilistic interpretation of the wave function. This magnitude is therefore closely related to Ψ by Eq. 1.13.

A hypothetical substitution of the complicated wave function by $\rho(\mathbf{r})$ would not only lead to a measurable object, but also to a much simpler one: a function of only three (spatial) coordinates, equally for every system. In fact, the solution of the Schrödinger equation, in principle, only requires the knowledge of the Hamiltonian operator, which is uniquely determined by the number of electrons N, the positions of the nuclei $\{\mathbf{R}_A\}$ (recall we are assuming a time-independent and Born-Oppenheimer regime) and their corresponding charges $\{Z_A\}$. The electron density integrates to the total number of electrons

$$\int \rho(\mathbf{r})d\mathbf{r} = N, \tag{2.66}$$

presents maxima (as cusps) only at nuclear positions and gives their associated $\{Z_A\}$ according to Kato theorem:

$$\lim_{\mathbf{r}\to\mathbf{R}_A} \left[\frac{\partial}{\partial \mathbf{r}} + 2Z_A \right] \bar{\rho}(\mathbf{r}) = 0, \tag{2.67}$$

where $\bar{\rho}(\mathbf{r})$ denotes the spherical average of $\rho(\mathbf{r})$. As can be appreciated, the electron density itself provides all the information required to specify a given system.

2.3.2 Early attempts

The physical basis for the use of the electron density to specify the Hamiltonian is due to Pierre C. Hohenberg and Walter Kohn [44]. The so-called Hohenberg-Kohn (HK) theorems prove that the GS energy E_0 (and wave function) is a unique functional of the corresponding GS density, and that a trial density satisfying the contour conditions yields an upper bound to the true E_0, which is attained if and only if ρ_0 is supplied (it follows the variational principle).

Although the physical justification was not given until 1964, the first attempts to use the electron density as the means for the calculation of properties of interest of chemical systems date back to the late 1920s, when Llewellyn H. Thomas and Enrico Fermi proposed the calculation of the kinetic energy basing on a uniform electron gas model and treating the remaining energy terms classically. Supposing that the energy is a functional of the density, so must be the different energy contributions. Under the BO approximation, the electronic energy $E[\rho]$ of an isolated system becomes

$$E[\rho] = T[\rho] + E_{ee}[\rho] + E_{en}[\rho]. \tag{2.68}$$

Supposing that electrons interact classically (that is, only the Coulomb potential is considered), the two potential energy terms can be expressed in terms of ρ as

$$E_{ee}[\rho] \equiv J[\rho] = \frac{1}{2} \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} \tag{2.69}$$

$$E_{en}[\rho] = -\sum_A Z_A \int d\mathbf{r}_1 \frac{\rho(\mathbf{r}_1)}{r_{1A}}, \tag{2.70}$$

where $J[\rho]$ denotes the classical electronic energy only. The Thomas-Fermi (TF) model centres its attention on the kinetic energy functional and provides the following expression considering a system of constant electron density:

$$T_{TF}[\rho] = \frac{3}{10}(3\pi^2)^{2/3} \int \rho^{5/3}(\mathbf{r})d\mathbf{r}. \tag{2.71}$$

This model resorts to a coarse approximation and hence the errors are usually large. Neither is it able to predict chemical bonding, but it provides a first energy functional in terms of the density [19].

A further step was taken by Dirac, leading to the Thomas-Fermi-Dirac model with the addition of an exchange term (first derived by Bloch):

$$K_D[\rho] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \int \rho^{4/3}(\mathbf{r})d\mathbf{r}, \tag{2.72}$$

later improved by Slater with the inclusion of a semiempirical parameter α, leading to Slater's Xα method

$$E_{X\alpha}[\rho] = -\frac{9}{8}\left(\frac{3}{\pi}\right)^{1/3} \alpha \int \rho^{4/3}(\mathbf{r})d\mathbf{r}, \tag{2.73}$$

that enjoyed popularity among physicists, although was not very accepted by chemists.

2.3.3 The construction of modern DFT: the Kohn-Sham prescription

The first functionals of the density resulted unsatisfactory, mainly due to the imprecise representation of the kinetic energy by the Thomas-Fermi model. With the aim of improving the description, Walter Kohn and Lu J. Sham proposed a scheme mirroring the HF model where this energy is much better represented.

As commented above, the HK theorems prove that the ground-state energy is a functional of the (also GS) density and so must be the terms giving rise to the total energy. These can be categorised as system-dependent or universal according to whether their form is maintained for every system or does change. Recalling Eq. 2.68, it can be seen that the electron-nucleus attraction term will change with

the system under consideration —it explicitly depends on the nuclear positions and charges. By contrast, both the kinetic energy and the electron repulsion energies are independent upon the number of electrons or nuclear scaffold, and therefore can be cast in what is known as a (the) universal functional $F[\rho]$. It follows that if $F[\rho]$ were known, the Schrödinger equation could be solved exactly and the calculated properties of the system would be likewise exact. The HK theorems prove the existence of such a functional but give no indication about how to calculate it.

Looking at the HF method where the lack of accuracy comes from the approximate form of the target wave function (the Slater determinant) but that yields the exact kinetic energy associated to it (i.e., that of non-interacting fermions subject to an effective potential v_{HF}), Kohn and Sham came up with an idea to much better approximate $T[\rho]$. Since the electron density is obtained by quadrature (Eq. 1.13), several wave functions may give rise to the same $\rho(\mathbf{r})$. The Kohn-Sham (KS) scheme considers the target GS density ρ_0 of the system under study equivalent to that of a hypothetical non-interacting electrons one. In this way, under this premise the SD wave function representing such a system can be used to calculate the kinetic energy of the real one:

$$T_S = -\frac{1}{2} \sum_i \langle \phi_i | \nabla^2 | \phi_i \rangle, \tag{2.74}$$

where ϕ_i are the KS orbitals, in analogy with the HF ones. These orbitals are the solution to a set of one-electron equations: the *Kohn-Sham equations*,

$$\hat{f}^{KS} \phi_i = \varepsilon_i \phi_i, \tag{2.75}$$

where the KS operator corresponds to

$$\hat{f}^{KS} = -\frac{1}{2} \nabla^2 + v_S(\mathbf{r}), \tag{2.76}$$

hence, a monoelectronic operator where the potential energy is represented by an effective, local potential $v_S(\mathbf{r})$. The resultant monoelectronic functions of the non-interacting system give rise to the density that equals the real one

$$\rho_S(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2 = \rho_0(\mathbf{r}). \tag{2.77}$$

The kinetic energy calculated as in Eq. 2.74 is not exactly the same as that of a real, interacting system. It is rather an approximation, but a very close one. The remaining kinetic energy as well as the effects that are not captured by the classical potential energy expressions (i.e., by the electronic Coulomb repulsion from Eq. 2.69) are grouped into the exchange-correlation (*xc*) functional $E_{xc}[\rho]$. When adding the electron-nucleus attraction energy to the universal functional, that reads

$$F[\rho] = T_S[\rho] + J[\rho] + E_{xc}[\rho], \tag{2.78}$$

the total DFT electronic energy is obtained:

$$E^{DFT}[\rho] = F[\rho] + E_{en}[\rho]. \tag{2.79}$$

By equating E^{DFT} to the exact energy E_0, the exchange-correlation functional is defined:

$$E_{xc}[\rho] = (T[\rho] - T_S[\rho]) + (E_{ee}[\rho] - J[\rho]) = T_C[\rho] + E_{ncl}[\rho]. \tag{2.80}$$

The exchange-correlation functional therefore comprises the non-classical energy effects associated to the interelectronic repulsion $E_{ncl}[\rho]$ and the remaining kinetic energy $T_C[\rho]$ that is not captured by the non-interacting model system. This term is the only unknown in the decomposition of the energy functional (Eq. 2.79).

The different energy contributions can be readily expressed in terms of the orbitals (all but the exchange-correlation, of course), and their constrained minimisation subject to the orthonormality condition of the orbitals leads to a set of mono-electronic equations similar to those associated with the non-interacting reference system (Eq. 2.75):

$$\left(-\frac{1}{2}\nabla^2 + \left[\int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + v_{xc}(\mathbf{r}_1) - \sum_A \frac{Z_A}{r_{1A}} \right] \right) \phi_i = \varepsilon_i \phi_i. \tag{2.81}$$

The terms in squared brackets can be grouped into an effective potential, that is identical to the previous v_S. Therefore, the KS equations are valid for a real, interacting system that shares the same electron density than the model non-interacting one. The orbitals thus obtained are the means to construct such density and therefore serve to calculate the different energy functional contributions independently. Again, if the form of all these functionals were known, the Schrödinger equation could be solved exactly, which is not the case. Within the universal functional decomposition, it is the xc term which is completely unknown, and the formulation of a (the) correct one has been (and it still is, indeed) matter of intense investigation.

Once Eq. 2.81 is fed with some exchange-correlation potential, it yields a more or less accurate set of KS orbitals and, hence, density depending on the accuracy of v_{xc}. However, Kohn and Sham do not provide any guidance on how to obtain or approximate v_{xc} (or, alternatively, the xc energy functional) other than the definition of the potential as the functional derivative of the corresponding energy with respect to the density:

$$v_{xc} = \frac{\delta E_{xc}}{\delta \rho}. \tag{2.82}$$

2.3.4 Approximations to the exchange-correlation functional

Many different approaches have emerged over the years in an attempt to provide a successful approximation to $E_{xc}[\rho]$. The common ground to all (or at least most) of them is the homogeneous electron gas (for which the exact exchange is known; see Eq. 2.72) that gives rise to the so-called *local density approximation* (LDA).

Local density approximation

Within the xc approximations, it is customary (although not physical) to split the xc energy into exchange (x), on the one hand, and correlation (c), on the other, and to approximate the two contributions independently. Besides, these functionals are usually expressed in terms of the corresponding energy densities (energy per particle) ε_x and ε_c as

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho] = \int \rho(\mathbf{r})\varepsilon_x[\rho]d\mathbf{r} + \int \rho(\mathbf{r})\varepsilon_c[\rho]d\mathbf{r}. \tag{2.83}$$

The LDA model utilises the x energy given by the Dirac expression (Eq. 2.72) for the total density. For open-shell systems, however, an LDA version where spin density —the density associated to electrons of a given spin function— is treated explicitly is preferred: the *local spin density approximation* (LSDA), that accounts for

density polarisation in these systems with unpaired electrons. When considering the correlation energy, however, there is no explicit expression available and alternatives based on fitting procedures from accurate Monte Carlo methods are used instead. It is the case of the Vosko, Wilk and Nusair (VWN) functional that, in combination with Slater exchange, gives rise to the SVWN functional.

The real systems (other than crystalline metallic solids) normally differ significantly from that of a uniform electron gas. The LDA/LSDA approximation not surprisingly results in usually large errors, both from the exchange and the correlation parts.

Generalised gradient approximation

Having LDA as the basis, further improvements accounting for the anisotropy of electron density can be introduced in terms of subsequent derivatives of the electron density (following a Taylor expansion series). A direct expansion does not lead to any improvement, but to even worse results. However, other strategies that also include the first derivatives (gradient) do provide more accurate results. These belong to the generalised gradient approximation (GGA) functionals, such as that proposed by Perdew and Wang that modifies the LSDA exchange (similar expressions are used for the closed-shell case) as

$$\varepsilon_x^{PW86} = \varepsilon_x^{LSDA}(1 + ax^2 + bx^4 + cx^6)^{1/15}, \tag{2.84}$$

where a, b and c are constants and

$$x = \frac{|\nabla\rho|}{\rho^{4/3}} \tag{2.85}$$

is the reduced density gradient (equivalently for each α or β densities), that acts as local inhomogeneity parameter. Other examples comprise the popular Perdew-Burke-Ernzerhof (PBE) functional or Becke's B88 and Perdew-Wang's PW91, that incorporates the reduced density gradient as well as a hyperbolic sine to correct the asymptotic behaviour of the energy density.

Further corrections adding either the Laplacian of the density $\nabla^2\rho$ or the kinetic energy density, that includes derivatives with respect to the orbitals, serve to improve the GGA results and belong to a new class called meta-GGA (or mGGA) [45]. Examples of this family are Perdew's TPSS or Becke-Roussel's BR.

At this stage one should be warned about the physical significance of the different functionals derived as corrections to the homogeneous electron gas. Although inaccurate, the latter has indeed physical grounds, whereas those DFT functionals that typically give much accurate energies and general properties rely on parameterised schemes, mixing different terms from other functionals or introducing simpler corrections, such as in Eq. 2.84, that permit the final functional being more free to fit the desired data with the adjustable parameters. This has been indeed subject of criticism. However, the computational advantages of DFT in front of other quantum mechanical methods due to the well balance between accuracy and computational cost —indeed, its formal scaling is $\mathcal{O}(K^4)$, similar to HF, and can even be reduced by invoking further approximations— have made it one of the most popular methods in computational chemistry.

Hybrid functionals

The exchange energy dominates considerably the exchange-correlation part. Therefore an accurate representation of exchange is a need for the development of any accurate DFT functional. Noticing that the exchange energy in HF is exact, the so-called hybrid functionals arise from the inclusion of some amount of HF-like exchange. One example of this type is Becke 3 parameter functional (B3):

$$E_{xc}^{B3} = (1-a)E_x^{LSDA} + aE_x^{HF} + b\Delta E_x^{B88} + E_c^{LSDA} + c\Delta E_c^{GGA} \qquad (2.86)$$

The a, b and c parameters are fitted to experimental data and control the amount of exact (HF) exchange in the case of a, or the exchange and correlation gradient corrections to LDA in the case of b and c.

A simpler hybrid functional is the half-and-half (HH) one, that combines half of the exact exchange with half of LDA exchange-correlation. Other popular examples cover B3 functional in combination with GGAs or mGGAs, such as B3PW91, B3P86, or the probably most common example of a hybrid DFT funtional, B3LYP (with an explicit form close to Eq. 2.86).

Minnesota functionals

Although the main families of DFT functionals have already been presented, more recent developments due to the Thrular group at University of Minnesota deserve special consideration due to their popularity [46, 47]. These are heavily parameterised functionals that include the semiempirical mGGA M06-L and M11-L, or the global hybrid M06, M06-2X or M06-HF density functionals.

To conclude this Section, it is worth mentioning that recent developments inspired by wave function-based methods that typically combine HF exchange with PT2 correlation energy have resulted in another category named as *double hybrid* functionals.

2.4 Density Matrices

The previous section introduces a quite different paradigm in molecular modelling, replacing the quantum mechanical wave function by a simpler function which resides in the real space and from which all properties can in principle be calculated likewise. Nevertheless, the lack of a universal (or, equivalently, an exchange-correlation) functional precludes the exact calculation of properties in DFT, and approximations based on adjustable parameters are currently the most common assumptions.

The calculation of molecular properties without the direct use of the wave function is not restricted to DFT. A different formalism vital in the present thesis is the one based on the use of the reduced density matrices (RDMs) [48].

2.4.1 Expectation values as functionals of the density matrices

Let us begin by examining how an expectation value looks like. Recalling Eq. 1.14, for a normalised 1-electron wave function, the expectation value of an operator $\hat{\mathcal{O}}$ reads

$$\langle \mathcal{O} \rangle = \langle \Psi | \mathcal{O} | \Psi \rangle = \int \Psi^*(\mathbf{x}) \hat{\mathcal{O}} \Psi(\mathbf{x}) d\mathbf{x}. \qquad (2.87)$$

If $\hat{\mathcal{O}}$ is a multiplicative operator, its expectation value can simply be expressed in terms of the electron density:

$$\int \Psi^*(\mathbf{x})\hat{\mathcal{O}}\Psi(\mathbf{x})d\mathbf{x} = \int \hat{\mathcal{O}}\Psi^*(\mathbf{x})\Psi(\mathbf{x})d\mathbf{x} = \int \hat{\mathcal{O}}\rho(\mathbf{x})d\mathbf{x} = \int \hat{\mathcal{O}}\rho(\mathbf{r})d\mathbf{r}. \tag{2.88}$$

However, if $\hat{\mathcal{O}}$ is not a multiplier, but rather an integral or a differential operator, such as that associated to the kinetic energy, the previous steps cannot be followed and $\langle \mathcal{O} \rangle$ cannot be expressed in terms of the electron density. A trick to circumvent this issue in the use of this formulation is to change the variable Ψ^* is dependent upon when applying the operator and restore the original one before integration. Following this recipe, $\Psi^*(\mathbf{x}) \rightarrow \Psi^*(\mathbf{x}')$, so that the expectation value of any mono-electronic operator can be calculated as

$$\langle \mathcal{O} \rangle = \int_{\mathbf{x}'=\mathbf{x}} \Psi^*(\mathbf{x}')\hat{\mathcal{O}}\Psi(\mathbf{x})d\mathbf{x} = \int_{\mathbf{x}'=\mathbf{x}} \hat{\mathcal{O}}\Psi(\mathbf{x})\Psi^*(\mathbf{x}')d\mathbf{x} = \int_{\mathbf{x}'=\mathbf{x}} \hat{\mathcal{O}}\rho(\mathbf{x};\mathbf{x}')d\mathbf{x}, \tag{2.89}$$

where $\rho(\mathbf{x};\mathbf{x}')$ is termed *density matrix*, or more accurately, *first-order reduced density matrix* (1-RDM).

The prior case was for a 1-particle system. For a general N-electron one, the 1-RDM is defined as

$$\rho_1(\mathbf{x};\mathbf{x}') \equiv \rho(\mathbf{x};\mathbf{x}') = N \int \Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)\Psi^*(\mathbf{x}_1', \mathbf{x}_2, ..., \mathbf{x}_N)d\mathbf{x}_2 \cdots d\mathbf{x}_N, \tag{2.90}$$

where the factor N accounts for the indistinguishability of electrons.

For two-body properties, the analogous second-order matrix (2-RDM) can be utilised:

$$\rho_2(\mathbf{x}_1, \mathbf{x}_2; \mathbf{x}_1', \mathbf{x}_2') = N(N-1) \int \Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)\Psi^*(\mathbf{x}_1', \mathbf{x}_2', ..., \mathbf{x}_N)d\mathbf{x}_3 \cdots d\mathbf{x}_N, \tag{2.91}$$

that, as can be appreciated, integrates to the total number of electron pairs $N(N-1)$.

With this procedure, any n-RDM can be calculated at will from the wave function; however, the non-relativistic Hamiltonian of an isolated system (recall Eq. 2.1) possesses only one- and two-body dependencies, what leads to the 1- and 2-RDMs as the only matrices needed for our current purposes. Moreover, since such a Hamiltonian does not act on the spin coordinates, we can refer to the spin-independent matrices $\rho(\mathbf{r};\mathbf{r}')$ and $\rho_2(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1', \mathbf{r}_2')$ obtained upon further integrating the spin of the remaining electrons in Eqs. 2.90-2.91.

Let us examine the explicit form that the electronic energy acquires under this formulation. The first term in the associated Hamiltonian (Eq. 2.4) is the kinetic energy operator, whose expectation value becomes

$$
\begin{aligned}
\langle T_e \rangle &= \int \Psi^*(\mathbf{x}_1, ..., \mathbf{x}_N)\left[\frac{1}{2}\sum_i \nabla_i^2\right]\Psi(\mathbf{x}_1, ..., \mathbf{x}_N)d\mathbf{x}_1 \cdots d\mathbf{x}_N \\
&= \frac{1}{2}N \int_{\mathbf{x}_1'=\mathbf{x}_1} \nabla_1^2 \Psi(\mathbf{x}_1, ..., \mathbf{x}_N)\Psi^*(\mathbf{x}_1', ..., \mathbf{x}_N)d\mathbf{x}_1 \cdots d\mathbf{x}_N \\
&= \frac{1}{2}\int_{\mathbf{x}_1'=\mathbf{x}_1} \nabla_1^2 \rho(\mathbf{x}_1; \mathbf{x}_1')d\mathbf{x}_1 = \frac{1}{2}\int_{\mathbf{r}'=\mathbf{r}} \nabla^2 \rho(\mathbf{r};\mathbf{r}')d\mathbf{r}.
\end{aligned} \tag{2.92}
$$

The second monoelectronic operator involves the electron-nucleus attraction:

$$
\begin{aligned}
\langle V_{ne} \rangle &= \int \Psi^*(\mathbf{x}_1, ..., \mathbf{x}_N) \left[-\sum_{\alpha i} \frac{Z_\alpha}{r_{\alpha i}} \right] \Psi(\mathbf{x}_1, ..., \mathbf{x}_N) d\mathbf{x}_1 \cdots d\mathbf{x}_N \\
&= -\sum_\alpha \int_{\mathbf{r}_1'=\mathbf{r}_1} \frac{Z_\alpha}{r_{\alpha 1}} \rho(\mathbf{r}_1; \mathbf{r}_1') d\mathbf{r}_1 = -\sum_\alpha \int \frac{Z_\alpha}{r_{\alpha 1}} \rho(\mathbf{r}_1) d\mathbf{r}_1,
\end{aligned} \tag{2.93}
$$

that can be readily expressed in terms of the electron density since it is a multiplicative operator. The third term in the electronic Hamiltonian is the two-body interelectronic repulsion, whose expectation value can be expressed as

$$
\begin{aligned}
\langle V_{ee} \rangle &= \int \Psi^*(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) \left[\sum_{i<j} r_{ij}^{-1} \right] \Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \\
&= N(N-1) \int_{\substack{\mathbf{x}_1'=\mathbf{x}_1 \\ \mathbf{x}_2'=\mathbf{x}_2}} r_{12}^{-1} \Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) \Psi^*(\mathbf{x}_1', \mathbf{x}_2', ..., \mathbf{x}_N) d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \\
&= \int_{\substack{\mathbf{r}_1'=\mathbf{r}_1 \\ \mathbf{r}_2'=\mathbf{r}_2}} r_{12}^{-1} \rho_2(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1', \mathbf{r}_2') d\mathbf{r}_1 d\mathbf{r}_1 = \int r_{12}^{-1} \rho_2(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2,
\end{aligned} \tag{2.94}
$$

although other normalisations are possible. Again, it turns out to be a multiplier, and thus no primed variables are needed so that the 2-RDM can be replaced by the pair density $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$. In fact, both $\rho(\mathbf{r})$ and $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$ are particular cases of either the 1- or 2-RDMs. These are obtained when the primed variables equal the original ones and therefore correspond to the respective diagonal matrices.

2.4.2 The pair density and electron correlation

The significance of the pair density is analogous to that of the electron density but, in this case, referring to two distinct electrons. Hence, $\rho_2(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2$ gives the probability of finding any two electrons in volume elements $d\mathbf{r}_1$ and $d\mathbf{r}_2$ about $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively, with any spin (for two specific spins σ_1 and σ_2, $\rho_2(\mathbf{x}_1, \mathbf{x}_2)$ must be used instead). It also reflects the antisymmetry requirement of the wave function [42] and, accordingly (in this case only applicable to the strict pair density with joint spatial-spin coordinates),

$$
\rho_2(\mathbf{x}_1, \mathbf{x}_2; \mathbf{x}_1', \mathbf{x}_2') = -\rho_2(\mathbf{x}_2, \mathbf{x}_1; \mathbf{x}_1', \mathbf{x}_2'). \tag{2.95}
$$

For the special case of $\mathbf{x}_1' = \mathbf{x}_1$ and $\mathbf{x}_2' = \mathbf{x}_2 = \mathbf{x}_1$,

$$
\rho_2(\mathbf{x}_1, \mathbf{x}_1) = -\rho_2(\mathbf{x}_1, \mathbf{x}_1) = 0. \tag{2.96}
$$

Therefore, $\rho_2(\mathbf{x}_1, \mathbf{x}_1)$ fulfils the Pauli exclusion principle. Moreover, $\rho_2(\mathbf{x}_1, \mathbf{x}_1)$ includes all the information about electron correlation, since it dictates the behaviour of electron pairs, that is the way in which electrons interact (matrices accounting for higher-order clusters are therefore not needed).

The pair density can be expressed as the sum of two terms, namely a product of the independent electron densities of the two electrons considered, representing the totally uncorrelated situation, and an exchange-correlation counterpart, regarded as a correction to the former description:

$$
\rho_2(\mathbf{x}_1, \mathbf{x}_2) = \rho(\mathbf{x}_1)\rho(\mathbf{x}_2) + \rho_{xc}(\mathbf{x}_1, \mathbf{x}_2) = \rho(\mathbf{x}_1)\rho(\mathbf{x}_2)\left[1 + f(\mathbf{x}_1|\mathbf{x}_2) \right], \tag{2.97}
$$

where $f(\mathbf{x}_1|\mathbf{x}_2)$ is the correlation factor that accounts for exchange and Coulomb correlation and introduces corrections to the self-interaction error. Indeed, this term is needed to ensure a correct integration of electron pairs rather than the wrong N^2 obtained from $\rho(\mathbf{x}_1)\rho(\mathbf{x}_2)$. $\rho_2(\mathbf{x}_1, \mathbf{x}_2)$ is also closely related to the *conditional probability* $\Omega(\mathbf{x}_2|\mathbf{x}_1)$, that gives the probability of finding a second electron at $\mathbf{x}_2$ given a first electron placed at the spin-spatial point $\mathbf{x}_1$:

$$\Omega(\mathbf{x}_2|\mathbf{x}_1) = \frac{\rho_2(\mathbf{x}_1, \mathbf{x}_2)}{\rho(\mathbf{x}_1)}. \tag{2.98}$$

The difference between the conditional probability of finding an electron at $\mathbf{x}_2$ and its uncorrelated counterpart constitutes the so-called *exchange-correlation hole* $h_{xc}(\mathbf{x}_1|\mathbf{x}_2)$:

$$h_{xc}(\mathbf{x}_1|\mathbf{x}_2) = \frac{\rho_2(\mathbf{x}_1, \mathbf{x}_2)}{\rho(\mathbf{x}_1)} - \rho(\mathbf{x}_2) = \rho(\mathbf{x}_2)f(\mathbf{x}_1|\mathbf{x}_2), \tag{2.99}$$

so that

$$\rho_{xc}(\mathbf{x}_1, \mathbf{x}_2) = \rho(\mathbf{x}_1)h_{xc}(\mathbf{x}_1|\mathbf{x}_2). \tag{2.100}$$

$h_{xc}(\mathbf{x}_1|\mathbf{x}_2)$ integrates to -1 —it contains the charge of one electron— and, therefore, $\rho_{xc}(\mathbf{x}_1, \mathbf{x}_2)$ does to $-N$ so that the decomposition of the pair density in Eq. 2.97 yields the correct number of electron pairs $N(N-1)$.

As advanced in Sections 2.2.1 and 2.3.4, electron correlation can be classified in two groups, namely Fermi correlation, including same-spin effects (i.e., self-interaction and exchange), and Coulomb correlation, that accounts for the instantaneous repulsion that each electron experiences in the presence of another electron due to their charged nature. Accordingly, the *xc* hole regardless of spin can be split as

$$h_{xc}(\mathbf{r}_1|\mathbf{r}_2) = h_x^{\sigma_1=\sigma_2}(\mathbf{r}_1|\mathbf{r}_2) + h_c^{\sigma_1,\sigma_2}(\mathbf{r}_1|\mathbf{r}_2), \tag{2.101}$$

where, by definition, the Fermi hole $h_x(\mathbf{r}_1|\mathbf{r}_2)$ relates the positions of parallel-spin electrons and the Coulomb term $h_c(\mathbf{r}_1|\mathbf{r}_2)$ is valid for any combination. Among the prior two components, it is the Fermi hole that integrates to -1 and guarantees Pauli exclusion since

$$h_x(\mathbf{r}_2 \to \mathbf{r}_1|\mathbf{r}_1) = -\rho(\mathbf{r}_1). \tag{2.102}$$

It is in fact a negative function that dominates the full $h_{xc}(\mathbf{r}_1|\mathbf{r}_2)$. On the other hand, the Coulomb hole integrates to zero and is therefore not restricted to negative values.

2.4.3   RDMs for single-determinant wave functions

A Slater determinant may become a useful choice of wave function in many situations. It constitutes the basis of the HF method that, in turn, serves as the starting point for further improvements of the simple description it provides.

For an SD wave function expressions 2.90 and 2.91 can be greatly simplified. In fact, since such wave functions are constituted by a set of monoelectronic spinorbitals, it is possible to write the 1-RDM and 2-RDM as expansions in terms of $\{\psi_i\}$ [48, 49]:

$$\rho(\mathbf{x}; \mathbf{x}') = \sum_{i,j} d_i^j \psi_i(\mathbf{x}) \psi_j^*(\mathbf{x}') \tag{2.103}$$

and

$$\rho_2(\mathbf{x}_1, \mathbf{x}_2; \mathbf{x}_1', \mathbf{x}_2') = \sum_{ij,kl} d_{ij}^{kl} \psi_i(\mathbf{x}_1) \psi_j(\mathbf{x}_2) \psi_k^*(\mathbf{x}_1') \psi_l^*(\mathbf{x}_2'), \tag{2.104}$$

where d_i^j and d_{ij}^{kl} are the expansion coefficients in each case of each density matrix.

Within the HF method, the use of an SD to shape the eigenfunction of Hamiltonian 2.21 leads to the expression of the energy in terms of spinorbitals 2.22. A close inspection reveals a simpler form of both density matrices. On the one hand, the 1-RDM results in a summation of the products of same spinorbitals, the expansion coefficients thus corresponding to a Kronecker delta δ_{ij}:

$$\rho^{SD}(\mathbf{x};\mathbf{x}') = \sum_i \psi_i(\mathbf{x})\psi_i^*(\mathbf{x}'). \tag{2.105}$$

The 2-RDM, that participates in the bielectronic part, in turn becomes

$$\rho_2^{SD}(\mathbf{x}_1,\mathbf{x}_2;\mathbf{x}_1',\mathbf{x}_2') = \sum_{ij} \left[\psi_i(\mathbf{x}_1)\psi_j(\mathbf{x}_2)\psi_i^*(\mathbf{x}_1')\psi_j^*(\mathbf{x}_2') \right.$$
$$\left. - \psi_j(\mathbf{x}_1)\psi_i(\mathbf{x}_2)\psi_i^*(\mathbf{x}_1')\psi_j^*(\mathbf{x}_2') \right]. \tag{2.106}$$

By considering the SD form of the 1-RDM, $\rho_2(\mathbf{x}_1,\mathbf{x}_2;\mathbf{x}_1',\mathbf{x}_2')$ can be expressed in terms of the latter:

$$\rho_2^{SD}(\mathbf{x}_1,\mathbf{x}_2;\mathbf{x}_1',\mathbf{x}_2') = \rho(\mathbf{x}_1;\mathbf{x}_1')\rho(\mathbf{x}_2;\mathbf{x}_2') - \rho(\mathbf{x}_2;\mathbf{x}_1')\rho(\mathbf{x}_1;\mathbf{x}_2'), \tag{2.107}$$

an expression that resembles Eq. 2.97, where only Fermi correlation is included. In fact, the previous decomposition in terms of the sometimes called *Fock-Dirac density matrix* (1-RDM) leads to the conclusion that, under HF and the single-determinant approximation, everything is determined by such a density function.

If the primes are removed in Eq. 2.107, the fist part equals that of Eq. 2.97, whereas the second does not loose its dependency on the off-diagonal elements of the 1-RDM and thus

$$\rho_{xc}^{SD}(\mathbf{x}_1,\mathbf{x}_2) = -\rho(\mathbf{x}_2;\mathbf{x}_1)\rho(\mathbf{x}_1;\mathbf{x}_2). \tag{2.108}$$

The previous expressions for both density matrices have been given in terms of spinorbitals. However, as commented above, a Hamiltonian that does not act on the spin of electrons permits the use of $\rho(\mathbf{r};\mathbf{r}')$ and $\rho_2(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}_1',\mathbf{r}_2')$ in the calculation of properties such as the energy. For the closed-shell case, both density matrices can be expanded in terms of the spatial MOs. The 1-RDM needed for the kinetic energy calculation (see Eq. 2.92) reads

$$\rho^{RHF}(\mathbf{r};\mathbf{r}') = 2\sum_i \phi_i(\mathbf{r})\phi_i^*(\mathbf{r}') \tag{2.109}$$

and becomes the simple electron density when using unprimed variables (the diagonal 1-RDM for V_{ne}, Eq. 2.93). In the case of the 2-RDM, only the diagonal matrix is needed for V_{ee}:

$$\rho_2^{RHF}(\mathbf{r}_1,\mathbf{r}_2) = \rho(\mathbf{r}_1)\rho(\mathbf{r}_2) - \frac{1}{2}\rho(\mathbf{r}_2;\mathbf{r}_1)\rho(\mathbf{r}_1;\mathbf{r}_2). \tag{2.110}$$

In the above expression, the xc —in reality pure exchange in the HF case— density, which corresponds to the second term, is calculated as

$$\rho_{xc}^{RHF}(\mathbf{r}_1,\mathbf{r}_2) = -\frac{1}{2}\rho(\mathbf{r}_2;\mathbf{r}_1)\rho(\mathbf{r}_1;\mathbf{r}_2) = -\frac{1}{2}\sum_{ij}\phi_i(\mathbf{r}_2)\phi_j(\mathbf{r}_1)\phi_i^*(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2). \tag{2.111}$$

As has been seen, the RDMs can be easily computed from the HF MOs. However, for correlated methods, expressions 2.109 and 2.111 may adopt a much more complicated form. The use of linear combinations of SDs to account for electron correlation permits the expansion of the 1- and 2-RDM as in Eqs. 2.103 and 2.104, but with non-trivial d_i^j and d_{ij}^{kl} coefficients. These must incorporate the products of the expansion coefficients accompanying the SDs and must also account for the Slater-Condon rules for each product of determinants. Therefore, the major problem in these methods concerns the computational cost of those computations, that are frequently not implemented in quantum chemistry packages. On the other hand, methods based on perturbative approaches such as MPn or CC, that do not yield a wave function consistent with the energy they provide, or DFT, where the wave function is completely absent, have necessarily to rely on some approximation to construct the required RDMs.

Whenever the RDMs are available, however, these constitute an appealing alternative in computational chemistry A remarkable property of density matrices is that they are invariant under orbital transformations. If a set of orbitals is transformed into another, since the final wave function remains unaltered, the sum of orbital products gives rise to the same density function. Such a robustness of density matrices results in an appealing real space alternative to calculate properties, as shall be explained a few pages further on.

2.5   The quantum chemical topology toolbox for wave function analyses

Once the main methods for obtaining a system's wave function have been presented, we are in a position to present the family of tools for analysing the chemically relevant information contained in such a complicated mathematical object. The method for obtaining global properties, such as energy and others, has been presented in terms of their associated quantum mechanical operators as expectation values. Nevertheless, as advanced in the Introduction (Section 1.4), the chemical theory is constructed upon the molecular structure hypothesis and most concepts and ideas in the chemist's world are formulated in terms of atoms interacting with each other, lone pairs or charge deficient regions that make a particular atom more or less prone to being attacked by another, orbital interactions that provide stability to a compound or a given ligand in a particular position of the coordination sphere, particular reactions that a certain functional group may trigger, etc. In all these situations a further dissection of the wave function is needed and with this purpose a set of wave function analyses based on very different approaches have emerged over the years.

With the aim of giving a theoretically sound basis to the different chemical constructs, Bader and coworkers developed a theory called the quantum theory of atoms in molecules (QTAIM) [30], rooted in the topology of the electron density from which a diversity of chemical concepts emanate naturally. It is the case of the aforementioned molecular structure hypothesis, that appears unequivocally when regarding the gradient field of the electron density in terms of the attraction basins in which the whole space becomes divided. The manner in which QTAIM defines atoms in a molecule, as advanced in the Introduction (Section 1.4), comprises more than a mere and accidental definition; it rather presents interesting properties that makes it a physically appealing AIM scheme.

Let us examine how the kinetic energy density looks like. The kinetic energy density is a local property that is well-defined when integrating over the whole system. However, a local definition is not unique [30, 50]. Recalling expression 2.92, the total kinetic energy of an N-electron system is proportional to the Laplacian of $\Psi^*\Psi$ integrated to all space, that can be developed as

$$\nabla^2(\Psi^*\Psi) = (\nabla^2\Psi^*)\Psi + \Psi^*(\nabla^2\Psi) + 2\nabla\Psi^*\nabla\Psi. \tag{2.112}$$

Integration of the previous terms over all coordinates but a spatial one (represented jointly by letter τ') leads to the definition of the following functions:

$$L(\mathbf{r}) = -\frac{1}{4}N\int\nabla^2(\Psi^*\Psi)d\tau' = -\frac{1}{4}\nabla^2\rho(\mathbf{r}) \tag{2.113}$$

$$K(\mathbf{r}) = -\frac{1}{4}N\int(\Psi^*\nabla^2\Psi + \Psi\nabla^2\Psi^*)d\tau' = -\frac{1}{4}(\nabla^2 + \nabla'^2)\rho(\mathbf{r},\mathbf{r}') \tag{2.114}$$

$$G(\mathbf{r}) = \frac{1}{2}N\int\nabla\Psi^*\nabla\Psi d\tau' = \frac{1}{2}(\nabla\nabla')\rho(\mathbf{r},\mathbf{r}'), \tag{2.115}$$

where $K(\mathbf{r})$ and $G(\mathbf{r})$ are two distinct definitions of the kinetic energy density and $L(\mathbf{r})$ represents a function of the Laplacian of the electron density $\nabla^2\rho(\mathbf{r})$. In the previous expressions, ∇ operators act on the spatial coordinate $\mathbf{r}$ not included in τ', the primed analogue ∇' acting on the primed coordinate $\mathbf{r}'$. It follows that

$$K(\mathbf{r}) = G(\mathbf{r}) + L(\mathbf{r}). \tag{2.116}$$

Indeed, the two different kinetic energy densities differ from one another by the $L(\mathbf{r})$ term. Globally, such a difference disappears, since $L(\mathbf{r})$ vanishes when intergrating over the whole space according to Gauss's divergence theorem (it can be proved some lines below), but when considering an arbitrary region it is not in general the case. Among the many different ways in which a molecule can be partitioned into atoms, QTAIM atomic basins do provide well-defined energies.

Considering the gradient of the electron density as the vector field in Gauss's theorem, for an arbitrary volume Ω delimited by a surface S it reads

$$\int_\Omega \nabla\nabla\rho(\mathbf{r})d\mathbf{r} = \oint_S \nabla\rho(\mathbf{r})\hat{\mathbf{n}}dS. \tag{2.117}$$

As can be appreciated, the left-hand side term involves the Laplacian of $\rho(\mathbf{r})$, what can be used to reexpress $L(\mathbf{r})$ in terms of the density flux across the boundaries of Ω. Given that the QTAIM atomic basins are enclosed by a zero-flux surface (recall Eq. 1.52), these regions fulfil $T(\Omega) = K(\Omega) = G(\Omega)$.

These quantum mechanically meaningful fragments shape the grounds of QTAIM, that, in turn, is the origin of a wider family of real space, orbital-invariant techniques for wave function analysis: the so-called quantum chemical topology (QCT). The present thesis has made use of various methods pertaining to the QCT family that have proven to be of great help in the study of the elusive chemical bond or more general interactions, as well as to ascertain the causes of stability/instability of a given compound or aggregate.

2.5.1 Characterisation of the $\rho(\mathbf{r})$ scalar field

The topology of the electron density and, with it, the characterisation of the chemical systems can be preformed in terms of the set of critical points $\mathbf{r_c}$ of $\rho(\mathbf{r})$. These are

TABLE 2.1: Classification of the stable critical points of the electron density.

Type	Acronym	(r, s)	Description
(Non-)nuclear attractor	(N)NA	$(3, -3)$	Local maximum of $\rho(\mathbf{r})$ with three negative curvatures.
Bond critical point	BCP	$(3, -1)$	$\rho(\mathbf{r})$ is maximum in two directions and minimum in one (that connecting two NAs).
Ring critical point	RCP	$(3, +1)$	One curvature is negative, while the other two are positive. It is a maximum in only one direction.
Cage critical point	CCP	$(3, +3)$	Local minimum of $\rho(\mathbf{r})$. Gradient paths pointing to each direction emanate from this critical point.

determined by the vanishing first derivative of the density, thus being those points for which $\nabla \rho(\mathbf{r_c}) = 0$, and comprise the maxima, minima and saddle points of such a scalar field.

The classification of critical points is performed in terms of the second derivatives of $\rho(\mathbf{r})$ at such points through the Hessian matrix, that, for an adequate interpretation, is diagonalised. This particular form is acquired when the coordinate axes coincide with the *principal axes of curvature*, which are nothing but the eigenvectors of this real, symmetric matrix. When inspecting the corresponding eigenvalues, three possibilities appear: they can be either positive, negative or zero and, according to it, the critical points are classified in terms of their rank r and signature s as (r, s). The signature corresponds to the sum of the signs of the three eigenvalues —i.e., the sum of the signs of the curvatures of a particular $\rho(\mathbf{r_c})$—, whereas the rank represents the number of non-zero curvatures, that can be either positive, negative or even zero. It turns out, however, that the rank of critical points in the majority of cases of molecular arrangements at and near equilibrium geometries is three. Any non-accidental or non-symmetry related zero curvature would indicate an unstable structure in the sense that a small change in the geometry would make the critical point vanish or bifurcate into other stable ($r = 3$) ones. Therefore, molecular structure is characterised in terms of three-ranged critical points.

Accordingly, four possible stable critical points emerge, whose description and classification in terms of (r, s) can be found in Table 2.1. In order to provide a visual comparison of the four, Figure 2.2 depicts the tetrahedrane molecule along with all its critical points of the electron density. Starting from the points where all gradient paths terminate, the nuclear attractors (NAs) correspond to the local maxima of $\rho(\mathbf{r})$, that coincide with the nuclear positions (these are the atoms in Fig. 2.2). This is a consequence of the singularity in the Coulomb potential, that makes the electron density present a cusp at the nucleus. In some occasions, however, other attractors different from the nuclei appear, and are thus called non-nuclear attractors (NNAs).

Following the classification in Table 2.1 of increasing s value, the next critical point is probably the most *chemical* one: the so-called bond critical point (BCP). It appears on a particular site of the interatomic surface (IAS) separating two topological atoms where $\rho(\mathbf{r})$ is locally maximum (therefore the point where the gradient paths defining the IAS terminate) and owes its name to its association with chemical bonds as it is the topological manifestation of the charge accumulation that follows
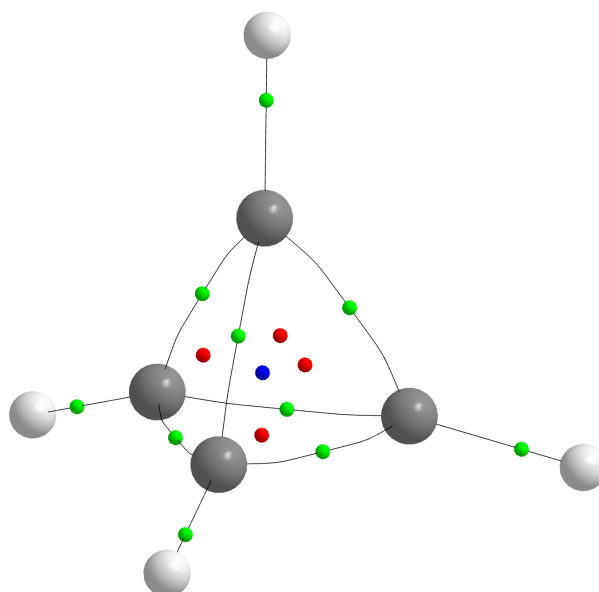
FIGURE 2.2: Critical points of the electron density in the tetrahedrane molecule obtained with AIMAll [24]. BCPs are depicted in green, RCPs in red and the CCP in blue. The bond paths connecting bonded atoms are also shown.

the formation of a bond. In fact, the presence of a bond is usually characterised by the *bond path*, the line connecting two NAs defined by the gradient paths originated at the BCP and terminating at such nuclei subject to some additional conditions. A BCP presents, indeed, a maximum value of the charge density along two directions (it is a region of charge accumulation), but a minimum along that connecting the two nuclei ($\rho(\mathbf{r})$ increases when approaching an NA). Both BCPs and bond paths are depicted in Fig. 2.2. These bond paths correspond to the lines connecting two atoms, and the green points on them are the BCPs marking the limits of each topological atom with respect to its neighbours along such a line closely related with chemical bonds. In the orthodox QTAIM, however, the presence of a bond is determined by an additional condition: the molecule must be in a minimum of its PES, making the forces acting on nuclei vanish. This condition limits the application of this topological characterisation of chemical bonds to equilibrium geometries, but in these situations the values of $\rho(\mathbf{r}_{BCP})$, for example, for a pair of atoms are good indicators of their relative bond strengths in different chemical environments.

A stable critical point with a positive +1 signature is a ring critical point (RCP). As its name suggests, it is found in planar moieties, such as the triangular faces of the tetrahedron formed by the C atoms in Fig. 2.2, or in planar species such as benzene. A RCP presents a maximum along one direction as that approaching the molecule, but it is a minimum in the plane in which it is found, since any movement on it would entail an increase in $\rho(\mathbf{r})$ as it approaches any of the surrounding nuclei.

Finally, the cage critical points (CCPs) are those where the gradient paths emanate to any direction. These are local minima of the electron density and originate gradient lines that may terminate at either a (N)NA, BCP or RCP. Not surprisingly, it appears at the centre of the molecular tetrahedron of tetrahedrane (the blue point in Fig. 2.2), or in similar *cage* molecules such as cubane, from which a displacement in any direction provokes an increase in $\rho(\mathbf{r})$. It is also remarkable that the wave

function and hence ρ and its derivatives vanish at infinity, what results in an infinite number of critical points from which infinite gradient paths emerge. Some are attracted by the nuclei, some by RCPs and others by RCPs, but clearly not by any CCP, since the latter is a local minimum of $\rho(\mathbf{r})$ and presents no negative curvature.

2.5.2   Chemical insight from the Laplacian of $\rho(\mathbf{r})$

Another scalar field whose topological properties can be readily linked with chemical concepts is the Laplacian of $\rho(\mathbf{r})$. This is indeed derived from $\rho(\mathbf{r})$, but contains information that is not directly accessible with the electron density itself. Rather, $\nabla^2\rho(\mathbf{r})$ is able to unveil the charge concentration and depletion regions hidden in a bare $\rho(\mathbf{r})$ representation.

The Laplacian of the electron density provides a measure of the local curvature of such a function in all its three dimensions. It in fact corresponds to the trace of the diagonal Hessian matrix of $\rho(\mathbf{r})$, giving thus rise to an overall picture of the molecular regions of charge accumulation or depletion depending on the value $\nabla^2\rho(\mathbf{r})$ presents. A negative Laplacian indicates that the function (in our case, the electron density) is locally concentrated, whereas when it is positive, the function is locally depleted at such point.
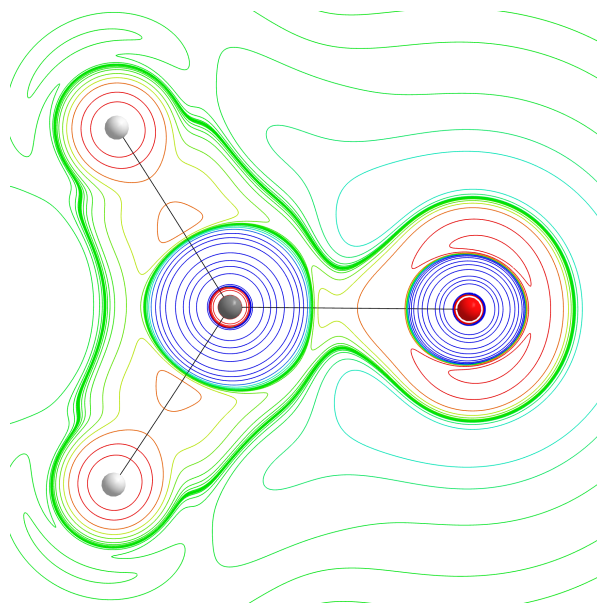
FIGURE 2.3: Laplacian of $\rho(\mathbf{r})$ on the molecular plane of formaldehyde obtained with AIMAll [24]. The values of $\nabla^2\rho(\mathbf{r})$ are colourcoded. Accordingly, green-blue values represent regions with $\nabla^2\rho(\mathbf{r}) > 0$, whereas reddish colours denote $\nabla^2\rho(\mathbf{r}) < 0$ regions.

The fine structure revealed by $\nabla^2\rho(\mathbf{r})$ consists of successive pairs of charge concentration and charge depletion regions corresponding to the quantum shells K, L, M, etc. in the proximities of each atom, as can be observed in Figure 2.3. The most interesting shell from the chemical point of view is the outer one: the valence shell. It is divided in a first inner part characterised by a $\nabla^2\rho(\mathbf{r}) < 0$ and thus called the *valence shell charge concentration* (VSCC), and an outermost region with $\nabla^2\rho(\mathbf{r}) > 0$. A close inspection to the topology of $\nabla^2\rho(\mathbf{r})$ —with its corresponding critical points calculated in a same fashion than the previous $\rho(\mathbf{r})$— not only reveals charge accumulation in the bond regions of a molecule, but also other points about the atoms

where $\nabla^2\rho(\mathbf{r})$ is minimum. These non-bonded charge concentration regions are associated to lone pairs (the two reddish regions about the O atom in Fig. 2.3), as a physical manifestation of purportedly localised electron pairs that the VSEPR (valence shell electron pair repulsion) model predicts. In fact, the empirical VSEPR, that aims to predict the geometry of different compounds according to a few simple rules, finds valuable support in QTAIM. The electron pair model is a pillar of VSEPR since this model rationalises the most likely arrangement that a given compound may present basing on the minimisation of the electron pair repulsion. These electron pairs can be bonding or non-bonding and their relative position seeking a minimum repulsion (therefore maximally separated) is presumed to determine the final geometry of the compound. The successful mapping between the VSEPR model and the critical points of the VSCC therefore provides an appealing physical basis for the former.

The previously seen capabilities of $\nabla^2\rho(\mathbf{r})$ are readily applied to chemical reactivity. This way, the charge concentration regions can be associated with Lewis bases or nucleophiles, or possible protonation sites. On the contrary, Lewis acids or electrophiles are linked with charge depletion regions whose interaction with electron rich moieties could entail a stabilising interaction.

2.5.3 Covalency and ionicity in real space: the delocalisation index

The definition of topological atoms permits an unambiguous calculation of atomic properties by integration of different operator densities over the atomic basins. The study of chemical bonding is not restricted to the topology of $\rho(\mathbf{r})$ and its Laplacian and, as such, other functions can be used to gain further insight into the nature and types of chemical bonds without resorting to any orbital model.

The topological study of the Laplacian of the electron density has been shown to provide real space signatures of the electron pair model by carefully identifying the VSCC regions and its maxima with either bonding or non-bonding localised electron pairs. Thus, the physical basis of the electron pair model may actually be found in the Fermi hole [51].

Given a position $\mathbf{r}_1$ of a reference electron, when a same-spin one approaches it, its probability density is penalised in that region until becoming null at that position (recall Eq. 2.102). If this behaviour persists when moving away from $\mathbf{r}_1$, the hole is said to be localised and no other same-spin electron will be allowed to occupy simultaneously that region of space. Furthermore, if this feature is shared by another electron with opposite spin (such as in closed-shell molecules [52]), the electron pair arises. Therefore, electron pairing emerges as a consequence of Pauli exclusion principle as determined by the Fermi hole. A measure of localisation and delocalisation of electron pairs can in turn be made basing on the extent this function is spread over one or several atomic basins. This way, the real space analogue of bond order between a pair of atoms, termed *delocalisation index* (DI(A, B) or δ^{AB}), is defined as

$$\text{DI}(A, B) \equiv \delta^{AB} = -2 \int_{\Omega_A} \int_{\Omega_B} \rho_{xc}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2, \qquad (2.118)$$

where the *xc* density, accounting for both the Fermi and Coulomb holes, is exploited to measure the number of electron pairs shared by those basins Ω_A and Ω_B [53]. [Note that since $\delta^{AB} = \delta^{BA}$, a factor of 2 has been included in the definition.] When both integrations take place in the same basin, a measure of how localised the *xc*

hole is appears. Accordingly, the *localisation index* (LI(A) or λ^A) corresponds to

$$\mathrm{LI}(A) \equiv \lambda^A = -\int_{\Omega_A}\int_{\Omega_A}\rho_{xc}(\mathbf{r}_1,\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2. \tag{2.119}$$

In general, the localisation index of a topological atom does not account for the whole electron population it presents (i.e., $|\lambda^A| < N(A)$). This fact signals the spread of the xc hole over other atomic basins with which atom A shares electrons. The electron population in A is therefore recovered when accounting for the total hole related to that atom:

$$\lambda^A + \frac{1}{2}\sum_{B\neq A}\delta^{AB} = N(A). \tag{2.120}$$

It follows that

$$-\int d\mathbf{r}_1\int d\mathbf{r}_2\rho_{xc}(\mathbf{r}_1,\mathbf{r}_2) = N = \sum_A \lambda^A + \frac{1}{2}\sum_A\sum_{B\neq A}\delta^{AB} \tag{2.121}$$

when the xc density is integrated over the whole space. Sometimes a different sign criterion is used and the pair density is defined as $\rho_2(\mathbf{r}_1,\mathbf{r}_2) = \rho(\mathbf{r}_1)\rho(\mathbf{r}_2) - \rho_{xc}(\mathbf{r}_1,\mathbf{r}_2)$ (note the different sign accompanying the xc density than that used in Eq. 2.97). Consequently, the xc density would integrate to a positive number of electrons rather than a negative one as described above so that the minus signs preceding the prior integrals would not be needed.

The localisation and delocalisation indices have profound statistical roots. They are closely related to the fluctuation in the average electron population of a given atom $N(A)$ through the variance [52]:

$$\mathrm{var}(N(A)) = N(A) - \lambda^A = \frac{1}{2}\sum_{B\neq A}\delta^{AB}. \tag{2.122}$$

Moreover, the DI is proportional to the covariance between the electron populations of two distinct basins as it is in fact a measure of the correlation between atomic populations [31]:

$$\delta^{AB} = -2\mathrm{cov}(N(A),N(B)). \tag{2.123}$$

All the previous properties suggest that the DI can be related to the higher or lower sharing of electrons between atoms and thus to the bond order between them. When an electron pair is prefectly delocalised between two atoms, its DI equals 1. When the same happens with a couple of them, the number is doubled. This is why the DI is considered as the real space analogue of bond order and serves to characterise quantitively such electron sharing between pairs of atoms. Besides, its relation with the LI and the average number of electrons in a basin (or simply, its electron population) permits the classification of bonds as covalent or ionic, with all possible degrees of polar covalency that lye in between. For instance, if electron correlation is neglected, the DI for a two-center, two-electron bond varies between 0 and 1 and provides an ionicity-covalency ruler. An ideal ionic bond will lead to DI = 0, and an ideal covalent one to DI = 1. Accordingly, the HF description provides a DI = 0.22 for the prototypical ionic bond of the diatomic NaCl, whereas dihydrogen presents a DI = 1.00 [53]. When analysing multiple bonds the correspondence is also found. This way, ethylene shows a DI between its carbon atoms of 1.88 and that of the nitrogen molecule becomes 3.04.

2.5.4 Natural adaptive orbitals: one-electron functions from real space

So far, different QCT descriptors have been presented. These are orbital-invariant and, accordingly, no reference has been made to orbitals other than for the process of their construction from orbital-based wave functions. However, and although at first sight it may seem contradictory, it is also possible to use real space-based orbitals, one-electron functions with direct links to other QCT descriptors. This is the case of the natural adaptive orbitals (NAdOs), a set of orbitals obtained upon diagonalisation of the cumulant densities integrated over different domains [54].

In Section 2.4 it has been shown that all the information required in the calculation of any n-body property is available in the respective nth-order RDM. Accordingly, a measure of bond order (and degree of covalency and ionicity) between a pair of atoms is done in terms of the two-centre integration of the two-body density component accounting for all the correlation between electrons: the xc density. This way, the double integration in the same basin leads to the localisation index, and when two distinct ones were used, to the delocalisation index, a real space (two-centre) bond order proportional to the covariance between those two electron populations. Going a step further, it is possible to generalise the concept of bond orders to account for n-centre correlations, and this is done by means of an, in turn, generalisation of the xc density: the *cumulant* densities (CDs).

As in the two-electron case, generalised *exchange-correlation* densities can be defined so as to represent n-electron fluctuations. This is done in what is known as nth-order CDs $\rho_n^C(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n)$, by means of which the RDMs can be expanded. This way,

$$\rho(\mathbf{r}) = \rho_1^C(\mathbf{r}) \tag{2.124}$$

$$\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \rho_1^C(\mathbf{r}_1)\rho_1^C(\mathbf{r}_2) + \rho_2^C(\mathbf{r}_1, \mathbf{r}_2) \tag{2.125}$$

$$\rho_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \rho_1^C(\mathbf{r}_1)\rho_1^C(\mathbf{r}_2)\rho_1^C(\mathbf{r}_3)$$
$$+ \rho_1^C(\mathbf{r}_1)\rho_2^C(\mathbf{r}_2, \mathbf{r}_3) + \rho_1^C(\mathbf{r}_2)\rho_2^C(\mathbf{r}_1, \mathbf{r}_3) + \rho_1^C(\mathbf{r}_3)\rho_2^C(\mathbf{r}_1, \mathbf{r}_2)$$
$$+ \rho_3^C(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3), \tag{2.126}$$

etc., with $\rho_2^C(\mathbf{r}_1, \mathbf{r}_2) \equiv \rho_{xc}(\mathbf{r}_1, \mathbf{r}_2)$, and where for the sake of simplicity, the diagonal, spin-free RDMs have been presented (similar relationships are found for the more general non-diagonal, spin-resolved RDMs [55]). Important recurrence relationships appear among the CDs, by which knowledge of an n-CD permits the calculation of the $(n-1)$-CD, and so the rest of the lower-order CDs:

$$\int \rho_n^C(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n)d\mathbf{r}_n = (-1)^n(n-1)\rho_{n-1}^C(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_{n-1}). \tag{2.127}$$

Continuing with the formulation introduced in Section 2.4 according to which the xc density integrates to $-N$, the previous relationship permits the correct integration of the 2-RDM to the $N(N-1)$ number of electron pairs, the 3-RDM to $N(N-1)(N-2)$, and so on.

As can be appreciated, each nth-order RDM can be decomposed into lower-order densities plus a remainder: its corresponding same-order cumulant. This is the object that contains the sought nth-order fluctuations. One interesting property of CDs is that, by virtue of Eqs. 2.121 and 2.127, each CD integrates to N multiplied by a given factor. In other words, getting rid of the $(-1)^n(n-1)$ factors in Eq. 2.127 and moving them to the expansions of the RDMs, integration of the cumulant densities yields the total number of electrons in the system and, at the same time, they

carry the nth-order correlation. By generalising what has been done with the delocalisation index, it is possible to define general n-centre bond indices $\langle N_{ab...n} \rangle$ by integration of the nth-order CD as

$$\langle N_{ab...n} \rangle = \int_{\Omega_a} d\mathbf{r}_1 \int_{\Omega_b} d\mathbf{r}_1 \cdots \int_{\Omega_n} d\mathbf{r}_n \rho_n^C(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n), \tag{2.128}$$

from where

$$\sum_{ab...n} \langle N_{ab...n} \rangle = N, \tag{2.129}$$

thus providing an n-centre decomposition of the electron population and a measure of n-centre bonding closely related to the DI. Moreover, integration of all but one coordinates and summation over all possible combinations of basins simply yields the electron density:

$$\rho(\mathbf{r}) = \sum_{ab...n} \rho_{ab...n}^C(\mathbf{r})$$

$$= \sum_{ab...n} \int_{\Omega_a} d\mathbf{r}_1 \int_{\Omega_b} d\mathbf{r}_2 \cdots \int_{\Omega_n} d\mathbf{r}_n \rho_{n+1}^C(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_{n+1}). \tag{2.130}$$

Similarly, when non-diagonal CDs are used, the, in turn, non-diagonal 1-RDM is obtained. As seen in Eq. 2.103, the 1-RDM can be expressed as an expansion in terms of the occupied MOs of the system and, on the same footing, the $\rho_{ab...n}^C(\mathbf{r}; \mathbf{r}')$ cumulant density matrices can be decomposed as

$$\rho_{ab...n}^C(\mathbf{r}; \mathbf{r}') = \boldsymbol{\phi}(\mathbf{r}) \mathbf{D}^{ab...n} \boldsymbol{\phi}(\mathbf{r}'), \tag{2.131}$$

where $\boldsymbol{\phi}(\mathbf{r})$ denotes the vector containing all the occupied MOs and $\mathbf{D}^{ab...n}$ is the matrix comprising the expansion coefficients. If $\rho_{ab...n}^C(\mathbf{r}; \mathbf{r}')$ is diagonalised, a set of multicentre natural orbitals $\{\phi_i^{ab...n}\}$ —the natural adaptive orbitals (NAdOs)— is obtained,

$$\rho_{ab...n}^C(\mathbf{r}) = \sum_i n_i^{ab...n} \left| \phi_i^{ab...n}(\mathbf{r}) \right|^2, \tag{2.132}$$

with their corresponding occupation numbers $\{n_i^{ab...n}\}$ fulfilling

$$\sum_i n_i^{ab...n} = \langle N_{ab...n} \rangle. \tag{2.133}$$

NAdOs therefore provide a one-electron (i.e., orbital) picture of n-centre bonding, their occupation numbers decomposing the corresponding n-centre bond index $\langle N_{ab...n} \rangle$ —which is equivalent to the n-centre DI though following a different normalisation. These orbitals are a means to detect and visualise multicentre bonding. NAdOs being completely localised over n centres are those that describe n-centre bonding. By contrast, those that are only partly localised over those atoms signal the existence of higher-order multicentre bonding.

2.5.5 Electron distribution functions

An alternative way of conceiving bonding is neither based on energetics nor on the delocalised orbitals between two fragments, but on the statistics of the electron distribution in space [56, 57].

Let us consider a general one-electron wave function and an exhaustive, arbitrary partition of space into m three-dimensional domains $\{\Omega_i\}_{i=1,...,m}$ fulfilling

$\cup_i \Omega_i = \mathbb{R}^3$ and $\Omega_j \cap \Omega_k = \varnothing \,\forall\, \Omega_j, \Omega_k \in \{\Omega_i\}$. According to Born's interpretation, the probability of finding the electron described by $\phi(\mathbf{r}_1)$ in one of such regions is

$$p(\mathbf{r}_1 \in \Omega_i) = \int_{\Omega_i} |\phi(\mathbf{r}_1)|^2 d\mathbf{r}_1, \qquad (2.134)$$

and provided ϕ is normalised, the sum of all distinct probabilities associated to each domain yields unity: $\sum_i p(\mathbf{r}_1 \in \Omega_i) = \int_{\mathbb{R}^3} |\phi(\mathbf{r}_1)|^2 d\mathbf{r}_1 = 1$.

For the more general N-electron case, $|\Psi|^2 d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N$ gives the probability of electron 1 being found in infinitesimal volume element $d\mathbf{r}_1$ around $\mathbf{r}_1$, electron 2 in $d\mathbf{r}_2$, and so on, and integration over three-dimensional regions $\Omega_i, \Omega_j, ..., \Omega_m$ gives the probability of those electrons to be found in such particular domains. Since electrons are indistinguishable, an $N!$ factor must be added so that all possible permutations are considered. However, this would correspond to the case in which only one electron is contained in each domain. For a more general situation in which n_1 electrons are assigned to Ω_1, n_2 to Ω_2, etc., only permutations among electrons inside the same region are allowed. For this reason, in an N-electron system, the probability of a given distribution of electrons $S \equiv \{n_1, n_2, ..., n_m\}$ satisfying $n_1 + n_2 + \cdots + n_m = N$ in real-space domains is

$$p(S) = \frac{N!}{n_1! n_2! \cdots n_m!} \int_{\Omega_1} d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_{n_1} \int_{\Omega_2} d\mathbf{r}_{n_1+1} d\mathbf{r}_{n_1+2} \cdots d\mathbf{r}_{n_1+n_2}$$
$$\cdots \int_{\Omega_m} d\mathbf{r}_{N-n_m+1} d\mathbf{r}_{N-n_m+2} \cdots d\mathbf{r}_N |\Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)|^2. \qquad (2.135)$$

These are the electron number distribution functions (EDF) [56] that provide valuable information concerning chemical bonding, being equally applicable to all systems provided their wave functions are accessible.

It is also possible to focus on the population of only one domain. Once all the $p(S)$ have been obtained, the probability of a one-domain population can be calculated by the standard rules of statistics adding up all the relevant probabilities that yield the requested marginal:

$$p(n_i) = \sum_{S|n_i} p(n_1, ..., n_i, ..., n_m). \qquad (2.136)$$

In a similar fashion, the probability of having n_i electrons in Ω_i and n_j in Ω_j can be calculated as

$$p(n_i, n_j) = \sum_{S|n_i,n_j} p(n_1, ..., n_i, ..., n_j..., n_m), \qquad (2.137)$$

and equivalent expressions are found for the subsequent probabilities of three-, fourth-, etc., domain population assignments.

With the one-domain EDFs the average number of electrons within such a region $\langle n_i \rangle \equiv N_i$ is available through

$$\langle n_i \rangle = \sum_{n_j=1}^{N} n_j p(n_i = n_j), \qquad (2.138)$$

Analogously, the average number of electron pairs that can be formed between Ω_i and Ω_j is

$$\langle n_i n_j \rangle = \sum_{n_k, n_l}^{N} n_k n_l p(n_i = n_k, n_j = n_l). \qquad (2.139)$$

This information provides an independent route for calculating the variance and the covariance of the electron populations associated to particular domains as

$$\text{var}(n_i) = \sum_{n_j}^{N}(n_j - N_i)^2 p(n_i = n_j) = \langle n_i^2 \rangle - \langle n_i \rangle^2 \tag{2.140}$$

and

$$\begin{aligned}
\text{cov}(n_i, n_j) &= \langle (n_i - N_i)(n_j - N_j) \rangle \\
&= \sum_{n_k, n_l}^{N}(n_k - N_i)(n_l - N_j)p(n_i = n_k, n_j = n_l) \\
&= \langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle \tag{2.141}
\end{aligned}$$

that, in turn, constitute the statistical basis of the localisation and delocalisation indices (recall Eqs. 2.122-2.123).

EDFs give access to averages of general functions of $n_1, n_2, ..., n_m$

$$\langle f(n_1, n_2, ..., n_m) \rangle \equiv \langle f(S) \rangle = \sum_S f(S) p(S) \tag{2.142}$$

that may serve, for instance, to calculate generalised covariances by employing $f(n_r, n_s, ..., n_t) = (n_r - N_r)(n_s - N_s) \cdots (n_t - N_t)$:

$$\begin{aligned}
\text{cov}(n_i, n_j, ..., n_k) &= \langle (n_i - N_i)(n_j - N_j) \cdots (n_k - N_k) \rangle \\
&= \sum_{n_r, n_s, ..., n_t} f(n_r, n_s, ..., n_t)p(n_i = n_r, n_j = n_s, ..., n_k = n_t). \tag{2.143}
\end{aligned}$$

Such statistical measurements of the electron populations can be readily exploited to construct generalised K-centre delocalisation indices

$$\delta^{ij \cdots k} = \frac{(-2)^{K-1}}{(K-1)!}\text{cov}(n_i, n_j, ..., n_k) \tag{2.144}$$

where K denotes the number of domains involved, thus providing a direct link with K-centre NAdOs through their occupation numbers: $\delta^{ij \cdots k} \propto \langle N_{ij \cdots k} \rangle$. Both indices are two sides of the same coin since both are built upon the K-centre covariance.

Although arbitrary three-dimensional regions can be used, when adopting a QTAIM scheme a given partition S is then called *real space resonance structure* (RSRS). By considering the simple example of the hydrogen molecule, which is a 2c-2e system, three possible RSRSs with their associated EDFs emerge: $(1,1)$, $(2,0)$ and $(0,2)$, thus assigning one electron to each atom, two for the first hydrogen and none for the second, and vice-versa. These correspond to a covalent-like resonance structure (in turn, the result of two possible spin combinations, whose individual probabilities can be calculated through the spin-resolved EDFs) and two ionic-like ones.

Therefore, the EDFs are not only linked with ionic or covalent real space indices (be them 2-, 3- or n-centre quantities), but also provide themselves a neat picture of how electrons tend to be distributed in integer packages.

2.5.6 Interacting quantum atoms: a real space energy decomposition

The closing Section for the Methodology chapter concerns the focus of this thesis. As described in the Introduction (Section 1.5), there is no unique way of assigning atoms in a molecule, so a similar non-uniqueness arises for the decomposition of its quantum mechanical energy. Nonetheless, not all the AIM methods possess an equally sound theoretical scaffold, what has led this thesis to rely on that provided by QTAIM. In a similar fashion, not all energy decomposition analyses (EDAs) present the same invariance properties and some of the most widely used ones may yield different conclusions when varying the set of orbitals used or the intermediate or reference states chosen. In this context, the interacting quantum atoms (IQA) approach presents an orbital-invariant alternative as it is (most frequently) based on the spacial partitioning of the (spinless) 1- and 2-RDMs to calculate atomic or fragment energies [58, 59].

Expressions 2.92-2.94 show that the different energy contributions can be calculated by means of the RDMs. IQA exploits its decomposition into atomic parts to obtain associated atomic energies. Similarly as QTAIM partitions the charge density by using spatial atomic weight functions, an atomic decomposition of the 1-RDM using weights can also be obtained:

$$\rho(\mathbf{r}_1; \mathbf{r}_1') = \sum_A \rho^A(\mathbf{r}_1; \mathbf{r}_1') = \sum_A w_A(\mathbf{r}_1') \rho(\mathbf{r}_1; \mathbf{r}_1'). \tag{2.145}$$

In the case of the 2-RDM, only the diagonal one (i.e., unprimed) is required. Its corresponding partition is obtained in this case by applying two simultaneous weight functions as

$$\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \sum_A \sum_B \rho_2^{AB}(\mathbf{r}_1, \mathbf{r}_2) = \sum_A \sum_B w_A(\mathbf{r}_1) w_B(\mathbf{r}_2) \rho_2(\mathbf{r}_1, \mathbf{r}_2). \tag{2.146}$$

These atomic matrices permit the computation of the different energy contributions by replacing the full $\rho(\mathbf{r}_1; \mathbf{r}_1')$ and $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$ by $\rho^A(\mathbf{r}_1; \mathbf{r}_1')$ and $\rho_2^{AB}(\mathbf{r}_1, \mathbf{r}_2)$ in Eqs. 2.92-2.94. The previous Equations 2.145 and 2.146 show the general form of partitioning the RDMs, valid for different kinds of weight functions. When it comes to Bader's, the computation of the atomic energy terms is equivalent to delimiting the respective integrations to the three-dimensional basins. Therefore, the atomic kinetic and potential energies become

$$T^A = \frac{1}{2} \int_{\mathbf{r}'=\mathbf{r}} \nabla^2 \rho^A(\mathbf{r}; \mathbf{r}') d\mathbf{r} = \frac{1}{2} \int_{\Omega_A} \nabla^2 \rho(\mathbf{r}; \mathbf{r}') |_{\mathbf{r}'=\mathbf{r}} d\mathbf{r}, \tag{2.147}$$

$$V_{en}^{AB} = - \int \rho^A(\mathbf{r}_1) \frac{Z_B}{r_{1B}} d\mathbf{r}_1 = - \int_{\Omega_A} \rho(\mathbf{r}_1) \frac{Z_B}{r_{1B}} d\mathbf{r}_1 \tag{2.148}$$

and

$$V_{ee}^{AB} = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\rho_2^{AB}(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}} = \int_{\Omega_A} d\mathbf{r}_1 \int_{\Omega_B} d\mathbf{r}_2 \frac{\rho_2(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}}. \tag{2.149}$$

Finally, the partition of the nucleus-nucleus repulsion is straightforward since nuclei are considered as point charges:

$$V_{nn}^{AB} = \frac{Z_A Z_B}{r_{AB}}. \tag{2.150}$$

Hence, the total energy E is exhaustively decomposed into intraatomic ($A = B$) and

interatomic energy contributions ($A \neq B$). On the one hand, the intraatomic quantities comprise the kinetic energies T_A plus all the interactions inside a particular basin Ω_A: V_{ee}^{AA} and V_{en}^{AA}. These three energies can be cast into the atomic *self* or *net energy* E_{net}^A:

$$E_{net}^A = T^A + V_{ee}^A + V_{en}^A, \tag{2.151}$$

On the other hand, the interatomic contributions comprise the interaction energies between the particles belonging to a given Ω_A and those ascribed to Ω_B:

$$E_{int}^{AB} = V_{ee}^{AB} + V_{en}^{AB} + V_{nn}^{AB}. \tag{2.152}$$

Note that the order of the sub- and superscripts matters. This way, $V_{ee}^{AB} = V_{ee}^{BA}$ or $V_{nn}^{AB} = V_{nn}^{BA}$, but $V_{en}^{AB} \neq V_{en}^{BA}$. The superscripts indicate the basins the corresponding first and second elements of the subscripts belong to. Accordingly, $V_{en}^{AB} = V_{ne}^{BA}$ and when the two elements in either the superscript or the subscript are the same, a permutation does not alter the significance of such energy term (e.g., $V_{en}^{AA} = V_{ne}^{AA}$). To shorten a little bit the notation, interactions corresponding to a same basin will be referred in the following with an only one letter (e.g., $V_{en}^{AA} \equiv V_{en}^{A}$).

The grouping carried out in Eqs. 2.151 and 2.152 allows for an exhaustive decomposition of E as

$$E = \sum_A E_{net}^A + \sum_A \sum_{B>A} E_{int}^{AB}. \tag{2.153}$$

It is also possible to redissect the total energy in additive atomic contributions as $E = \sum_A E_{add}^A$ by defining this effective quantity, the *additive energy*, as the sum of the net energy of an atom plus half all the interaction energies it presents with the rest of the atoms in the system:

$$E_{add}^A = E_{net}^A + \frac{1}{2} \sum_{B \neq A} E_{int}^{AB}. \tag{2.154}$$

As it can be seen, IQA provides a chemically appealing additive scheme that allows for the decomposition of the total energy in associations of atoms. In other words, the very chemical concept of functional groups is found within IQA. As such, E can alternatively be decomposed (or recovered) as a sum of group net energies and their interactions with the rest:

$$E = \sum_{\mathcal{G}} E_{net}^{\mathcal{G}} + \sum_{\mathcal{G}} \sum_{\mathcal{H}>\mathcal{G}} E_{int}^{\mathcal{G}\mathcal{H}}. \tag{2.155}$$

The group net energies $E_{net}^{\mathcal{G}}$ are analogous to the atomic ones in the sense that they comprise the kinetic energies of the particles confined in their extent plus their interactions. Hence, atomic net energies and interatomic interactions within the group are involved:

$$E_{net}^{\mathcal{G}} = \sum_{A \in \mathcal{G}} E_{net}^A + \sum_{\substack{B>A \\ B \in \mathcal{G}}} E_{int}^{AB}. \tag{2.156}$$

Accordingly, the inter-group interaction energies are defined as the sum of those comprising atoms of one group with those of the other:

$$E_{int}^{\mathcal{G}\mathcal{H}} = \sum_{A \in \mathcal{G}} \sum_{B \in \mathcal{H}} E_{int}^{AB}. \tag{2.157}$$

Additionally, group additive energies can be defined likewise.

Clearly, not only predefined functional groups are possible to be established, but also general groups or molecular fragments. Sometimes, in order to distinguish from the atomic analysis, the use of fragment energies is referred to as *interacting quantum fragments* (IQF). This different naming will be used in the present thesis for those works that present a clear distinction of both levels of analysis.

IQA provides a neat picture of the energetics a molecule presents with a particular atomic arrangement, but in many occasions it is the relative energies which are more insightful. It is the case of the *binding* or *formation energy* E_{form}, that collects all the energy changes undergone in a given chemical process. A related quantity is the *deformation energy* E_{def}, accounting for the inner changes produced during that process in an atom or fragment (or globally in the molecule), and thus collects the effects associated to the geometrical (nuclear) and electronic rearrangements. This quantity corresponds to the net energy change of that particular atom or fragment:

$$E_{def}^A = \Delta E_{net}^A = E_{net}^A - E_{net}^{A,0}, \tag{2.158}$$

where $E_{net}^{A,0}$ denotes the atomic net energy of an initial state (the same expression can be generalised to groups of atoms). In the case of a full formation process from the isolated atoms, this quantity represents the *in vacuo* energy $E^{A,0}$ of those and, consequently, the associated formation energy becomes

$$E_{form} \equiv \Delta E = \sum_A E_{def}^A + \sum_A \sum_{B>A} E_{int}^{AB} = E - \sum_A E^{A,0}. \tag{2.159}$$

Some remarks concerning the latter and the former expressions: on the one hand, the formation energy, that for a full process from the isolated atoms equals the difference between the final energy of the molecule and the sum of the atomic ones in isolation, presents two general sources, namely, the deformation of these atoms (generalisable to fragments) and the interaction they acquire when brought from infinity to the final arrangement (that, when some already belong to a preexistent species, an energy difference is used instead: ΔE_{int}^{AB}). A formation process therefore results from the balance between the usually non-favourable deformation and the frequently energy lowering upon interaction the atoms or fragments finally gain. On the other hand, in search for an energetic descriptor carrying the transferable properties of atoms, the net energy emerges as a promising candidate. It coincides with the total energy of an isolated species and when an atom or a functional group is carried from one chemical environment to another, unless both differ to a large extent, the net energy remains very similar and the energy change due to formation consequently becomes mainly explained by the new interactions with the surrounding groups.

The IQA interaction energies are not only able to distinguish among their different two-body sources (i.e., *e-e*, *n-n* or *e-n*) but, given relation 2.97, they can also be split into a classical or Coulomb part E_{class} and a quantum mechanical *xc* one E_{xc}. Hence, for a given pair of atoms,

$$E_{int}^{AB} = E_{class}^{AB} + E_{xc}^{AB}. \tag{2.160}$$

Moreover, the Coulomb term in the previous equation does not necessarily account only for classical electron repulsion (obtained after a direct substitution of $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$

in Eq. 2.94 by its two-term decomposition from Eq. 2.97), but the classical electron-nucleus and nucleus-nucleus interaction energies can also be incorporated by supplying a total density

$$\rho_{tot}(\mathbf{r}) = \sum_\alpha Z_\alpha \delta(\mathbf{r} - \mathbf{R}_\alpha) - \rho(\mathbf{r}) \tag{2.161}$$

for the Coulomb part, leading to a general partition of the interaction energy among all particles.

To conclude, as commented initially not all EDAs fulfil the same conditions or present the same properties that make them more or less suitable when analysing energetics. IQA fulfils in fact three conditions that give it its particular flavour: 1) IQA is able to unequivocally distinguish between atomic (and group) energy components, giving complete recovering of the total energy, 2) it provides a detailed description of the interactions among them and 3) no approximation or references are needed in its development.

2.5.7   The multipole expansion applied to IQA

Without going into details about the specific implementation IQA has in the original code for this energy decomposition analysis (PROMOLDEN [60]), it is worth mentioning that the costly 2-electron integrals can be more easily tackled by invoking either the Laplace or the bipolar expansions of r_{12}^{-1} [49]. These are generalisations of the *multipole expansion*, an expansion series of the interaction energy between two charge densities by means of multipole moments describing the (usually) anisotropic distribution of electrons and nuclei in space.

Let us consider two molecules A and B located at positions $\mathbf{R}_A$ and $\mathbf{R}_B$ with total interpenetrating charge densities $\rho_{tot,A}(\mathbf{r})$ and $\rho_{tot,B}(\mathbf{r})$. The electrostatic interaction between the two species reads

$$E_{elec}^{A,B} \equiv E_{class}^{A,B} = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho_{tot,A}(\mathbf{r}_1)\rho_{tot,B}(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2. \tag{2.162}$$

[Note that in this case A and B do not refer to topological atoms, but to complete molecules whose charge densities have no boundaries]. Such interaction between two charge densities can also be seen as the interaction experienced by molecule A in the presence of the potential due to $\rho_{tot,B}(\mathbf{r})$ [33] (or vice-versa):

$$E_{elec}^{A,B} = \int_{\mathbb{R}^3} \rho_{tot,A}(\mathbf{r}) V^B(\mathbf{r}) d\mathbf{r}. \tag{2.163}$$

By expanding $V^B(\mathbf{r})$ in a Taylor series about position $\mathbf{R}_A$, one arrives at

$$\begin{aligned} V^B(\mathbf{R}_A) &= q^B \frac{1}{R} - \sum_i^3 \mu_i^B \nabla_i \frac{1}{R} + \sum_i^3 \sum_j^3 \frac{1}{3} \Theta_{ij}^B \nabla_i \nabla_j \frac{1}{R} - \cdots \\ &= \sum_n^\infty \sum_i^3 \sum_j^3 \cdots \sum_p^3 \frac{(-1)^n}{(2n-1)!!} T_{ij\ldots p}^{(n)} \xi_{ij\ldots p}^{B(n)}, \end{aligned} \tag{2.164}$$

where indices $i, j, ..., p$ denote the specific components in Cartesian coordinates of the succinctly expressed multipole moments $\xi_{ij\ldots p}^{B(n)}$ of order n, that are tensors of rank 0 (charge, q), 1 (dipole, μ), 2 (quadrupole, Θ), etc. The contravariant $T_{ij\ldots p}^{(n)}$ tensors

that accompany them represent the successive derivatives

$$T^{(n)}_{ij...p} = \nabla_i \nabla_j \cdots \nabla_p \frac{1}{R} \tag{2.165}$$

of the inverse of the distance between the two molecules ($R = |\mathbf{R}_B - \mathbf{R}_A|$). These derivatives introduce position operators, such as $T_x = -\frac{\mathbf{x}}{R^3}$ or $T_{yz} = \frac{3\mathbf{yz}}{R^5}$, transforming Eq. 2.163 into a series of multipolar interactions between multipoles from $\rho_{tot,A}(\mathbf{r})$ and $\rho_{tot,B}(\mathbf{r})$ charge distributions:

$$
\begin{aligned}
E^{A,B}_{elec} \approx E^{A,B}_{elec,mp} &= q^A \big[Tq^B - \sum_i T_i \mu_i^B + \frac{1}{3} \sum_i \sum_j T_{ij} \Theta_{ij}^B - \cdots \big] \\
&+ \sum_i \mu_i^A \big[T_i q^B - \sum_j T_{ij} \mu_j^B + \frac{1}{3} \sum_j \sum_k T_{ijk} \Theta_{jk}^B - \cdots \big] + \cdots \\
&= Tq^A q^B - \sum_i T_i (q^A \mu_i^B - \mu_i^A q^B) \\
&+ \sum_i \sum_j T_{ij} \big(\frac{1}{3} q^A \Theta_{ij}^B - \mu_i^A \mu_j^B + \frac{1}{3} \Theta_{ij}^A q^B \big) + \cdots ,
\end{aligned}
\tag{2.166}
$$

therefore approximating[4] the electrostatic energy as a summation of charge-charge, charge-dipole, dipole-quadrupole, etc. terms. Formally, the summation extends to infinity, but it is frequently enough to consider multipoles up to the quadrupoles for molecules so that the electron distribution in space becomes sufficiently well described. It is important to note that the higher weight of some particular terms over the others comprises valuable information about how a given charge density is distributed. In the case of a completely spherical system, for instance, only the charges will survive, whereas when the species has cylindrical symmetry, both the total charge and the dipoles will be able to contribute (though they can always be zero if the species considered is neutral in the case of the charge, or has a zero dipole if the one-dimensional distribution is symmetric with respect to the centre of inversion).

The previous formulation has been presented in terms of the usual Cartesian coordinates. However, Eq. 2.166 can be reexpressed in spherical coordinates in a simpler way that is particularly appealing from the programmer point of view:

$$E^{A,B}_{elec,mp} = \sum_{l_1 m_1}^{\infty} \sum_{l_2 m_2}^{\infty} C_{l_1 m_1 l_2 m_2}(\hat{R}) \frac{Q^A_{l_1 m_1} Q^B_{l_1 m_1}}{R^{l_1+l_2+1}}, \tag{2.167}$$

where m_i runs from $-l_i$ to $+l_i$, $C_{l_1 m_1 l_2 m_2}(\hat{R})$ are known coefficients calculated for each $(l_1 m_1, l_2 m_2)$ quartet, that depend on the angular coordinate $\hat{R} \equiv (\theta_B - \theta_A, \phi_B - \phi_A)$ determining orientation and $Q^A_{l_1 m_1}$ and $Q^B_{l_2 m_2}$ are the spherical-coordinate multipoles of each molecular density. These are obtained through real spherical harmonics S_{lm} (see Ref. 49) as

$$Q_{lm} = N_l \int_{\mathbb{R}^3} r^l S_{lm}(\hat{r}) \rho(\mathbf{r}) d\mathbf{r}, \tag{2.168}$$

---

[4]Given that we are focusing on the electrostatic interaction between two interpenetrating molecular densities, Eq. 2.166 is an approximation to $E^{A,B}_{elec}$, but they do not exactly coincide. As explained below, the Taylor series cannot yield an exact expansion of the electrostatic energy if the charge distributions interpenetrate one another.

where N_l is a constant dependent on l, and r and $\hat{r}$ represent the radial and the angular coordinates, respectively.

Since the multipolar energies $E_{elec,mp}^{A,B}$ derive from a Taylor series expansion, they equal $E_{elec}^{A,B}$ whenever some conditions are fulfilled. On the one hand, the Taylor series is formally infinite, although it can be truncated at some point preserving a sufficient accuracy. On the other, the convergence of the series is limited to a certain domain of R, those values for which $R \geqslant |\mathbf{r}_1 + \mathbf{r}_2|$. In the case of two interacting infinite charge densities, $E_{elec,mp}^{A,B}$ does in principle not converge since these densities interpenetrate each other and the convergence condition is not fulfilled. However, although the electron density is spread throughout the whole space, it is mostly concentrated around the nuclei and decays exponentially out of them. Moreover, in case the intermolecular distance is not sufficiently large, the convergence problems can be solved by distributing multipoles throughout the molecule [61]. Moving now to the QTAIM case, $E_{elec,mp}^{AB}$ is obtained as in Eq. 2.167 by summing all the atomic pairs $I \in A - J \in B$, maintaining the notation A, B for the interacting molecules. Although group multipoles can also be constructed, this is a distributed multipole scheme. Moreover, topological atoms are not interpenetrating by definition. Nonetheless, much care has to be put on this fact, since the condition $R \geqslant |\mathbf{r}_1 + \mathbf{r}_2|$ is not always fulfilled. In the QTAIM, a maximum atomic radius R_I from the atomic position (i.e., from the atomic nucleus) to the farthest point of the IAS is taken so as to define the convergence sphere of such atom. Whenever $R \geqslant R_I + R_J$, the multipolar approximation will release an accurate $E_{ele,mp}^{IJ}$ energy (provided the electron distribution is well represented by the finite number of multipoles considered). For those QTAIM atoms whose outmost borders extend to infinity, the same consideration than for the infinite molecular densities can be applied, and R_I can be taken as finite when the portion of electron density not considered is negligible.

As a final remark, the previous application of the multipole expansion has been devoted to Coulomb energies. However, it is not restricted to electrostatics and the computation of the exchange-correlation energy can also take advantage of this approximation [62], although no more details will be given since the present thesis makes use of the information provided by the Coulomb multipole expansion only.

2.5.8   Implementation of IQA in density functional theory

IQA is a powerful and physically sound energy decomposition, but since it resorts to the RDMs to compute the energies, the family of DFT methods is in principle not compatible with IQA. In DFT there is no wave function, but a pseudo-wave function constructed from the KS orbitals, so no density matrix (other than the diagonal 1-RDM, that is, the electron density) can be exactly constructed. It is not a problem for the kinetic energy decomposition —it is calculated as the expectation value of the corresponding operator in a given spacial region—, nor is it for the Coulomb interaction —in this case only the charge density is needed, which can be easily partitioned—, but the assignment of the exchange-correlation energy to a given atom or to an interaction between a couple of them is not an easy task, and, in any case, some approximation must be followed. This issue has been matter of concern due to the predominance of DFT in electronic structure calculations, but a few years ago a promising scaling technique was proposed to by-pass this limitation [63].

Given a particular exchange-correlation functional yielding an energy density $\varepsilon_{xc}[\rho(\mathbf{r}), |\nabla\rho(\mathbf{r})|, ...] \equiv \varepsilon_{xc}(\mathbf{r})$, the total DFT exchange-correlation energy can be easily

computed as

$$E_{xc}^{DFT} = \int_{\mathbb{R}^3} \rho(\mathbf{r})\varepsilon(\mathbf{r})d\mathbf{r} + a_0 E_x^{KS}, \tag{2.169}$$

where for hybrid functionals, the fraction of pure HF exchange E_x^{KS} is modulated by a_0. This quantity is readily obtained from the KS orbitals, that do permit the construction of an exchange matrix following Eq. 2.111.

By delimiting the previous integral to a particular atomic basin (recall that $\mathbb{R}^3 = \cup_A \Omega_A$), the total xc energy is split into additive atomic contributions $E_{xc,add}^{DFT,A}$, involving both intra- and inter-basin energies. To enable a distinction of these two sources of the xc energy, a scaling procedure based on the E_{xc}^{DFT}/E_x^{KS} ratio within each basin can be used. The prescription implemented in PROMOLDEN begins with a first calculation of all these ratios as the λ_A parameters

$$\lambda_A = \frac{E_{xc,add}^{DFT,A}}{E_{x,add}^{KS,A}} = a_0 + \frac{1}{E_{x,add}^{KS,A}} \int_{\Omega_A} \rho(\mathbf{r})\varepsilon_{xc}(\mathbf{r})d\mathbf{r}. \tag{2.170}$$

The *additive* Hartree-Fock exchange is calculated through

$$E_{x,add}^{KS,A} = E_x^{KS,A} + \frac{1}{2}\sum_{B \neq A} E_x^{KS,AB}, \tag{2.171}$$

where the intra-atomic ($B = A$) or inter-atomic ($B \neq A$) exchange energies are computed by means of the exchange density $\rho_x^{KS}(\mathbf{r}_1, \mathbf{r}_2)$ as

$$E_x^{KS,AB} = \int_{\Omega_A} d\mathbf{r}_1 \int_{\Omega_B} d\mathbf{r}_2 r_{12}^{-1} \rho_x^{KS}(\mathbf{r}_1, \mathbf{r}_2). \tag{2.172}$$

Once the set of λ_A parameters is available, these are utilised to approximate the intra- or inter-atomic DFT xc energies as follows:

$$\tilde{E}_{xc}^{AB} = \frac{1}{2}\left[\lambda_A + \lambda_B\right] E_{x,add}^{KS,A}. \tag{2.173}$$

Consequently, the xc energy is reconstructed from (or split into) intra-atomic and inter-atomic contributions:

$$E_{xc}^{DFT} = \sum_A \tilde{E}_{xc}^A + \sum_{B \neq A} \tilde{E}_{xc}^{AB}. \tag{2.174}$$

Although the presented scheme is one of the many that could be used to approximate DFT-IQA energies, the experience with this scaling technique results valuable and demonstrates that this partitioning is robust and consistent, giving rise to very useful energies, comparable with those provided by HF [63–65].

2.5.9   IQA partitioning in continuum solvation models

Gas-phase descriptions are convenient in many situations, but when the purpose is to model a molecule in solution, in which the solute-solvent interactions may change dramatically its properties, a correct representation of the solvent is required. In contrast to explicitly treating solvent molecules as part of the whole QM system, continuum models opt for a continuous description of the solvent medium as an unstructured dielectric fluid [66, 67].

A main feature that characterises the distinct continuum solvent models is the manner in which they treat electrostatics, that is, how the molecular charge density defines the molecular potential that originates the solvent reaction potential (commonly known as *reaction field*), and how it is described. Among them, some models make use of multipoles to describe the solute charge density at the molecular origin, while others distribute them to different centres; some do not change the initial distribution upon interaction with the medium and others consider back polarisation effects (i.e., the polarisation suffered by the solute due to the polarised solvent medium). A second feature to distinguish among continuum solvation models is the definition of the cavity in which the solute molecule is placed. This cavity is enclosed by a *solvent accessible surface* (SAS), the interface between solute and the dielectric medium, whose definition ranges from the use of an isovalue surface of the electron density to the more common superposition of van der Waals-like atomic spheres.

There exist different implementations to cope with the electrostatic problem and thus to calculate the solvent potential. One of them is the apparent surface charges (ASC) method, according to which the SAS is discretised and divided in segments (tesserae) to which a set of point charges $\{q_k\}$ are associated. These charges act as the electrostatic response of the solvent medium to the polarisation exerted by the solute and are thus responsible for the solvent reaction potential. Accordingly, the solute-solvent $(X - S)$ electrostatic interaction energy [68] can be expressed as

$$E_{int}^{X-S} = \sum_A \sum_k \frac{Z_A q_k}{|\mathbf{R}_A - \mathbf{s}_k|} + \sum_k \frac{\rho(\mathbf{r}) q_k}{|\mathbf{r} - \mathbf{s}_k|} d\mathbf{r} = \sum_k V^X(\mathbf{s}_k) q_k, \qquad (2.175)$$

where $V^X(\mathbf{s}_k)$ denotes the total electrostatic potential due to the solute at the kth charge position $\mathbf{s}_k$.

When studying the process of solvation, some energy is spent in polarising the solvent when inserting the solute [69]. By considering a linear response of the medium (whose details can be read in Ref. 69), one half of the total interaction energy will be spent in such process. The corresponding energy associated to the solvated state (*sol*) reads

$$E^{sol} \equiv G^{sol} = E(\{\mathbf{R}_A\}^{sol}) + \frac{1}{2} E_{int}^{X-S}, \qquad (2.176)$$

where $E(\{\mathbf{R}_A\}^{sol})$ is the QM energy of the solute in the presence of one half the solvent potential, therefore accounting for polarisation effects. This energy indeed corresponds to a free energy, as it can be regarded as the work spent in assembling all the constituent particles in a final arrangement within a polarised dielectric. Its difference with the energy in the gas phase $E^{GP} \equiv E(\{\mathbf{R}_A\}^{GP})$ gives the (electrostatic) free energy of solvation:

$$\Delta G_{solv} = E^{sol} - E^{GP}. \qquad (2.177)$$

The previous quantity takes into consideration electrostatic effects only. This is the minimum requisite of any implicit solvation model and it is in fact the main energy contribution to solvation when, for example, polar solvents are used [66]. Improvements can be achieved by considering non-electrostatic effects such as dispersion or cavitation —that is, the formation of the cavity within the solvent medium—, thermal corrections being usually small.

So far, global energies associated to solvation have been presented. Nonetheless, one of the purposes of this thesis is to extend the use of IQA to biochemical systems for which solvent effects are in many cases crucial. Therefore, the use of implicit

solvation methods based on the ASC approach results very appealing from the IQA decomposition point of view. On the one hand, the atomic decomposition of the solute potential is straightforward —it is the potential due to the partitioned molecular density—, what serves to define atomic interaction energies with the solvent as

$$E_{int}^{X-S,A} = \sum_k V^A(\mathbf{s}_k) q_k. \tag{2.178}$$

With the previous partitioning, all the ingredients for a full decomposition of E^{sol} become available,

$$E^{sol} = \sum_A E_{add}^{sol,A} = \sum_A \left(E_{add}^A(\{\mathbf{R}_A\}^{sol}) + \frac{1}{2} E_{int}^{X-S,A} \right), \tag{2.179}$$

and additive energies of the solvated atoms can be defined by adding one half the atomic interaction energy with the solvent to the electronic and nuclear additive energy. Following this scheme, the IQA decomposition of the electrostatic free energy of solvation results in

$$\Delta G_{solv} = E^{sol} - E^{GP} = \sum_A (E_{add}^{sol,A} - E_{add}^{GP,A}) = \sum_A \Delta G_{solv}^A. \tag{2.180}$$

2.5.10 Dispersion-augmented IQA decomposition of HF/DFT energies

Dispersion interactions are crucial in large biomolecular systems and complexes. They involve the long-range correlation of electron clouds between atoms or fragments. As such, wave function-based methods accounting for correlation may release an accurate global energy, although at the expense of high computational demands. To circumvent this issue, several models have been proposed over the years. Special attention deserve those developed as a complement of DFT (and HF) methods, that make accurate results much more accessible. One of the most popular ones is Grimme's third generation dispersion correction (D3), a semiempirical potential based on the London formula, that presents an R^{-6} dependence on the interatomic distance and also depends on the respective atomic polarisabilities [70, 71].

Such a dispersion correction is directly added to the mean-field (i.e., either DFT or HF) energy E_{MF} as

$$E_{MF-D3} = E_{MF} + E_{disp}, \tag{2.181}$$

giving thus rise to the either DFT-D3 of HF-D3 methods. When applied to HF, E_{disp} is a new energy source as it contains all the dispersion energy that HF, by definition, lacks. By contrast, E_{disp} is only a part of the total dispersion energy in DFT methods, as they already account for a fraction of it.

The dispersion term is formally the result of two contributions, stemming from a two- and a three-body interaction; however, in line with the pairwise spirit of IQA, all the attention will be paid to the two-body term. In its original formulation, there is a singularity when two atoms A and B approach one another and hence the interatomic separation R_{AB} tends to zero. To avoid this problem and guarantee a correct asymptotic behaviour reaching a constant value at short distances, a damping function can be added. A popular prescription is that due to Becke and Johnson (BJ), leading to a D3(BJ) dispersion correction of the form (as implemented in the ORCA package [72, 73])

$$E_{disp} = - \sum_{A<B} \sum_{n=6,8} s_n \frac{C_n^{AB}}{R_{AB}^n + f^n(R_0^{AB})}, \tag{2.182}$$

where C_n^{AB} is the nth-order dispersion coefficient of the pair AB, obtained from accurate dynamic polarisabilities, s_n is a scaling factor dependent on the density functional used and $f(R_0^{AB})$ is the BJ damping function, linearly dependent on the ratio $R_0^{AB} = \sqrt{C_8^{AB}/C_6^{AB}}$.

The inclusion of an extra pairwise term in the IQA scheme is straightforward. As such, the IQA-D3 methodology [64] combines the HF or DFT Coulomb and exchange-correlation energies with the corresponding E_{disp}^{AB} term, that can be cast into the total pair interaction energy, giving rise to an *enhanced* energy decomposition

$$E = \sum_A E_{net}^A + \sum_{B>A} E_{int}^{AB} = \sum_A E_{net}^A + \sum_{B>A} \left(E_{class}^{AB} + E_{xc}^{AB} + E_{disp}^{AB} \right), \qquad (2.183)$$

where it is worth remarking that the pairwise dispersion energy used has an empirical origin and hence is not included in the xc term although dispersion is formally a correlation effect. This way, we restrict the E_{xc} term to the pure energy contribution calculated from $\rho_{xc}(\mathbf{r}_1, \mathbf{r}_2)$, that is not affected by the D3 potential.

Chapter 3

Results & Discussion

This Chapter summarises the research work carried out in this thesis and already published in seven different manuscripts fully reproduced in Chapter 5. Two separate lines in the object of research, but complementary in what this thesis pursues —i.e., the popularisation of the IQA methodology, in conjunction with other orbital-invariant techniques—, can be found. On the one hand, the first steps carried out towards the extension of IQA to the computational biochemistry world; on the other, the elucidation and characterisation of controversial chemical bond classes, with special attention to non-covalent interactions, by means of IQA and other QCT techniques.

3.1 Towards a biochemical application of IQA

Biochemistry is full of complex chemical processes where subtle changes in the structure of a molecule can be determinant in enabling or impeding its activity. The focus in computational biochemistry is primarily placed on large systems, whose complete treatment in QM terms is prohibitive and hence, QM/MM mixed methodologies combining the quantum mechanical description of a specific region and simplified models of the environment surrounding it are the dominant choice. In this regard, IQA offers the possibility of introducing in a theoretically sound way an atomic partition of the specific effects that this simplified representation of the environment causes on the QM region.

The four research works included in this first Section of Results address the development and application of IQA to larger and more complex systems than those typically studied so far, focusing on specific problems that commence to be fixed, and paving the way for a general strategy of systematically applying IQA to biochemical systems as a QM/MM-IQA combined description with a QM part encompassing around a hundred atoms. Accordingly, Section 3.1.1 addresses the performance of the IQA implementation of DFT in the most prototypical hydrogen-bonded systems: water clusters. Section 3.1.2, in turn, extends the IQA-D3 combined methodology to the elucidation of the local energetic effects that fluorine-containing organic molecules may provoke in the presence of different functional groups, leading to one or another preferred conformation of medium-sized fluorinated biomolecules (e.g., drug compounds). Environmental effects are introduced in IQA for the first time as those caused by a continuum solvent on a solute molecule in Section 3.1.3, partitioning the electrostatic component of the solvation free energy. Finally, a thorough analysis of intermolecular electrostatic interactions is performed in Section 3.1.4, where the IQA/IQF energy terms are inspected so as to describe non-covalent binding in close comparison with widely-used MM potentials, and where the energetic impact of electrostatics in the short range is carefully investigated.

3.1.1 DFT performance in the IQA energy partition of small water clusters

The applicability of IQA was initially restricted to wave function-based methods such as HF, CASSCF, CISD or CCSD, therefore limiting its use to a large extent if correlation effects were desired to be included. Notwithstanding, the high computational demands of the IQA machinery for correlated wave functions can be overcome resorting to density functional theory. With the aim of paving the way for future IQA analyses of hydrogen bonding (HB) in multiple scenarios, especially for large systems, this work presents an assessment of the performance of sixteen different density functionals in the IQA description of HB non-additive effects.

HB is an ubiquitous non-covalent interaction (NCI). It is present in many biochemical aggregates, determining their conformation and activity; crystalline solids, influencing the packing some compounds exhibit; or water systems, where it largely determines the properties of water both as a solid and as a liquid. Not only is HB a strong local interaction, but it also affects the global properties of the aggregates it forms. For instance, HB is responsible for non-additive properties such as energy in water clusters. This way, the higher or lower stability of water systems when adding more and more molecules to the cluster depends on the HB network that has been formed, which may present both *cooperative* or *anti-cooperative* effects.

Mirroring two previous studies on those effects in small water clusters [74, 75], the archetype of hydrogen-bonded systems, this work focuses on $(H_2O)_n$ systems from $n = 2$ to $n = 6$ [65]. On the one hand, cooperative effects are studied on the basis of *homodromic* clusters, that is, those systems where each water molecule acts both as a hydrogen-bond donor and acceptor. These water clusters range from the water trimer to the ring hexamer (see Figure 3.1). Concerning the latter, it is in the water hexamer where the flexibility of orientations and arrangements leads to different possible isomers where cooperative and anti-cooperative effects compete to yield more or less stable structures. In fact, the ring hexamer is the only six-member cluster that can be considered as homodromic, whereas anti-cooperativity also takes place in the book-, cage- and prism-shaped hexamers. The diversity of HB in all these clusters allows for a classification of HB in a strength hierarchy, which has also been inspected under the DFT prism.

Following this approach, the aforementioned HB effects were studied in the previously reported CCSD/aug-cc-pVDZ geometries from the work of Segarra-Martí [76], and other optimised monomer and cage, prism and book hexamers with the same level of theory. The electronic structure calculations prior to the IQA analysis comprised HF, MP2 (up to the pentamer) and CCSD (only for global energies) calculations, as well as 16 DFT functionals comprising: the LDA SVWN, GGA functionals B97, BLYP, BP86, OLYP, PBE and PW91; the hybrid B3LYP, B3P86 and B3PW91; the mGGA TPSS and the heavily-parameterised Minnesota functionals M06, M06-2X, M062L, M06-HF and M11-L.

First of all, cooperative effects emerge in the formation energies of the aggregates as forces providing further strengthening and stability to the n-clusters formed in the process $(H_2O)_{n-1} + H_2O \rightleftharpoons (H_2O)_n$. They can be evaluated by comparing the formation energy ΔE_n of each cluster relative to that of the dimer. With this prescription,

$$\Delta\Delta E_n = \Delta E_n - \Delta E_2 \tag{3.1}$$

indicates the presence of cooperative effects when negative (the inclusion of another water molecule produces a higher stabilisation than the formation of a single HB), and the more negative its value, the stronger the HB cooperative effects. For the
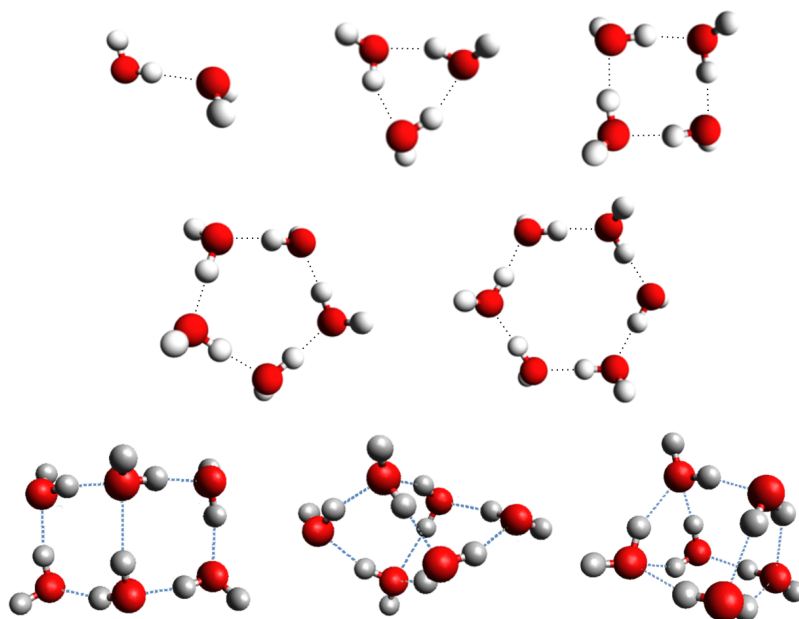
FIGURE 3.1: Water clusters inspected in this work.

cases under consideration, all the $n = 3 - 6$ energies are negative, reaching a minimum at $n = 4$ and diminishing the additional stabilising effect for the pentamer and ring hexamer. It is found that the trends dictated by the reference methods are reproduced by most of the density functionals analysed (see Figure 3.2). Some exceptions are the SVWN functional that overestimates HB cooperativity to a large extent —it is a very simple LDA one—, whereas OLYP underestimates $\Delta\Delta E$ and Minnesota functionals predict very similar $\Delta\Delta E_3$ and $\Delta\Delta E_4$ although CCSD yields $\Delta\Delta E_3 > \Delta\Delta E_4$.

As commented in Section 2.5.6, binding or formation can be viewed as the result of two contributions, namely, the deformation of the fragments (E_{def}) plus the interaction they experience (E_{int}): $E_{form} = E_{def} + E_{int}$. The trends provided by the different DFT functionals for these energies are identical and follow those dictated by the reference methods (Figure 3.3). Nonetheless, SVWN becomes again the most distant density functional as seen in the significantly deviated value of E_{form} per molecule.

The natural decomposition of E_{int} in IQA is into a classical or electrostatic component E_{class} and a QM exchange-correlation one E_{xc}. This decomposition serves to ascertain the energetic preponderance of one or another kind in a given chemical bond. Such a decomposition applied to the covalent O-H bond of water captures the effect that the formation of an H-bond with another molecule causes on it. Figure 3.4 presents the average interaction energies of this bond along with its two components with respect to the monomeric O-H ones. As can be appreciated, all the DFT functionals studied describe the strengthening of the bond due to electrostatics and the lesser degree of covalency dictated by the higher and higher E_{xc}^{O-H} when increasing the cluster size. In this case, although SVWN becomes again separated from the majority of the functionals, it is more remarkable to see how HF and the HF-like M06-HF, along with MP2, result in considerably lower classical energies
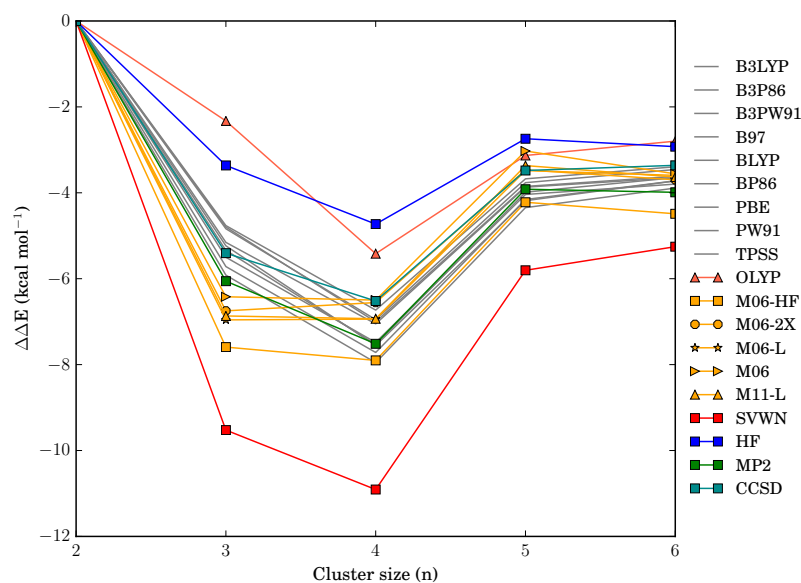
FIGURE 3.2: *n*th cluster formation energies relative to the water dimer.
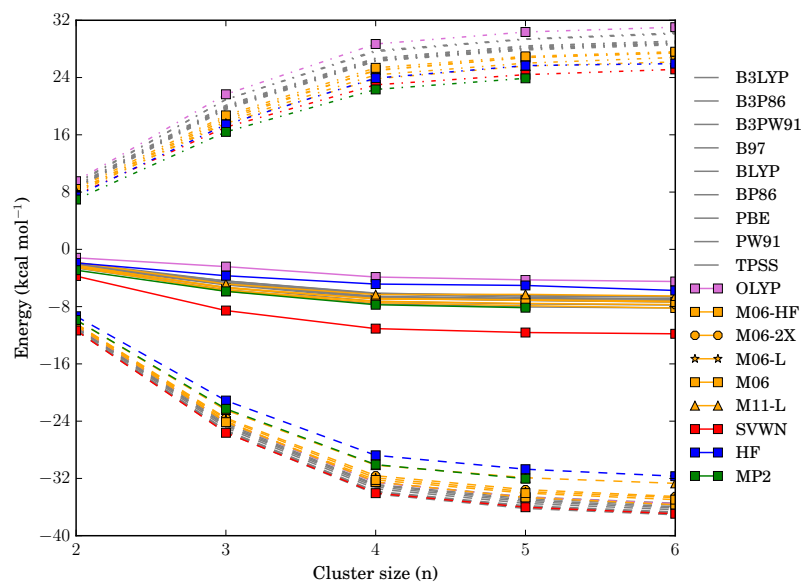
FIGURE 3.3: Formation (solid), deformation (dashed-dotted) and interaction (dashed lines) energies per molecule in the respective *n*th water cluster.
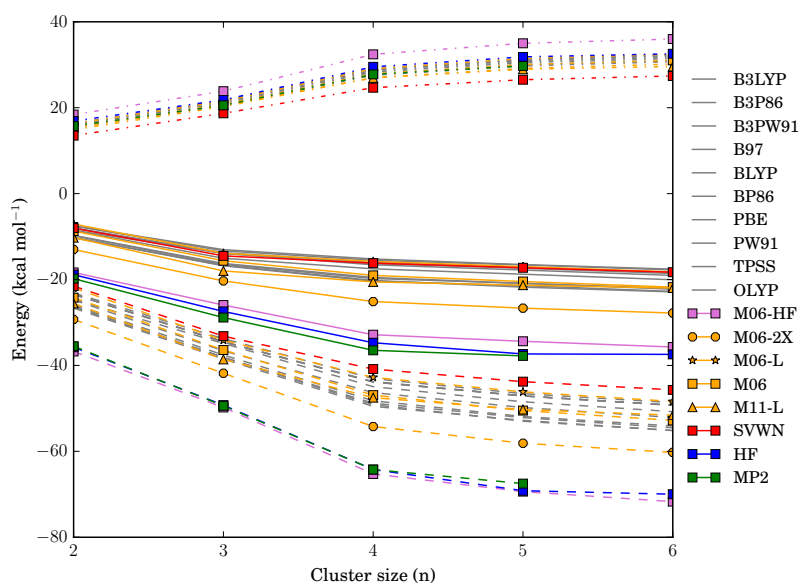
FIGURE 3.4: Average classical (dashed), exchange-correlation (dashed-dotted) and total interaction energies (solid lines) of O-H bonds in which the H atom is an H-bond donor. The energies are relative to the monomer O-H ones.

while overestimating the raise in E_{xc}^{O-H} with n. This might be due to the overdelocalised electron density provided by HF that yields larger atomic charges than those provided by correlated calculations.

Up to this point, it has been seen that most of DFT functionals —in general, those based on GGA and subsequent developments— not only provide correct trends, but also fairly accurate energies related to HB cooperativity. Going beyond the restriction to this kind of non-additive effect, the different water hexamers, such as those studied in Ref. 75, exhibit H-bonds that *cooperate* but also *anticooperate* among them. The hierarchy in strength proposed in the previous work, that classifies HBs according to the number of moieties from which a water molecule is H-bond acceptor or donor, can be also tested under DFT to verify the prior findings. Accordingly, the additive interaction energies defined in Eq. 7 of the publication concerning this work (Ref. 65) provides a valuable energetic descriptor of the performance of DFT in these cases. Nevertheless, let us focus on a non-energetic descriptor but a statistical one: the DI. The HB classification in increasing strength must be followed by an increase in bond order or, in other words, in covalency. This can be measured by the DI that, in light of the results reported in Figure 3.5, supports the mentioned classification. The only discrepancy concerns the values for HB type 3, that do not follow the increasing trend and give a similar and slightly smaller DI than that found for type 2. However, and focusing on the performance of the different DFT functionals tested, the trend dictated by HF is again reproduced by all of them, leading to the double conclusion that i) most DFT functionals behave not only qualitatively (trends) but also quantitatively well (accuracy) when describing HB non-additivities, and that ii) the DFT-IQA implementation based on a scaling technique fruitfully overcomes the limitations of DFT in the lack of a true wave function and provides an appealing and successful scheme for its combination with IQA. The conclusions found provide a firmer basis for applying DFT-IQA to larger H-bonded systems for which *ab initio*
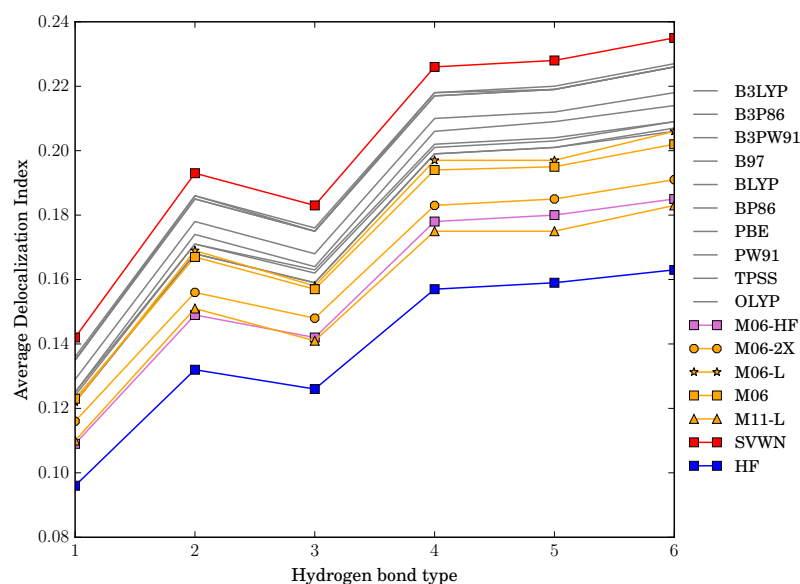
FIGURE 3.5: Average DI for each type of H-bond in water hexamers.

method may result unaffordable.

3.1.2 Fluorine conformational effects characterised by IQA

Fluorine-containing compounds have arisen as promising chemicals in numerous fields. The small size and compactess of fluorine make it a good candidate for hydrogen substitution in organic molecules, with the introduction of new properties due to the high polarity an the lone pairs of the X-F bond. The effects that accompany fluorine insertion are thus electrostatic and stereoelectronic, and their role in conformational preferences has not been fully disclosed yet as it is still matter of controversy. It is the case of the extensively studied *fluorine gauche effect* in 1,2-difluoroethane. In this molecule, the *gauche* conformation turns out to be more stable than placing the two fluorine atoms in opposite positions as found in the *anti* arrangement. This fact has produced extensive literature with not always matching conclusions. For example, natural bond orbital (NBO) analyses assign the *gauche* preference to he $\sigma_{CH} \rightarrow \sigma_{CF}^*$ hyperconjugative interactions (stereoelectronic effects) [77–79], whereas energy decompostion analysis (EDA) finds equally-favourable orbital and electrostatic contributions to the *gauche* effect [80, 81]. Further support for electrostatics comes from IQA analyses, that highlight the Coulomb contribution from the interaction between C and F in positions 1,3 [82].

It becomes clear that no consensus has been achieved on this topic, probably due to the difficulty of directly comparing the results of very diverse methodologies. Paying attention not only to the atomic, but, more importantly, to the group contributions to one or another preferred conformation on the basis of the IQA real space energetic analysis, we have investigated a variety of fluorine-containing molecules of the kind $F-CH_2-CH_2-X$ and $F-CH_2-CO-X$, as well as selected conformers of an α,β-difluoro-γ-aminoacid derivative (see Figure 3.6). For this, HF-D3(BJ)/cc-pVTZ and RI-MP2/cc-pVTZ full optimisations in the gas phase were carried out, followed by DLPNO-CCSD(T)/aug-cc-pVTZ single point calculations on the MP2

converged geometries. In the case of the larger **14** molecule, automatic conformational analyses were performed, from which the most stable conformers were selected and reoptimised at the HF-D3(BJ)/cc-pVTZ level of theory [83].
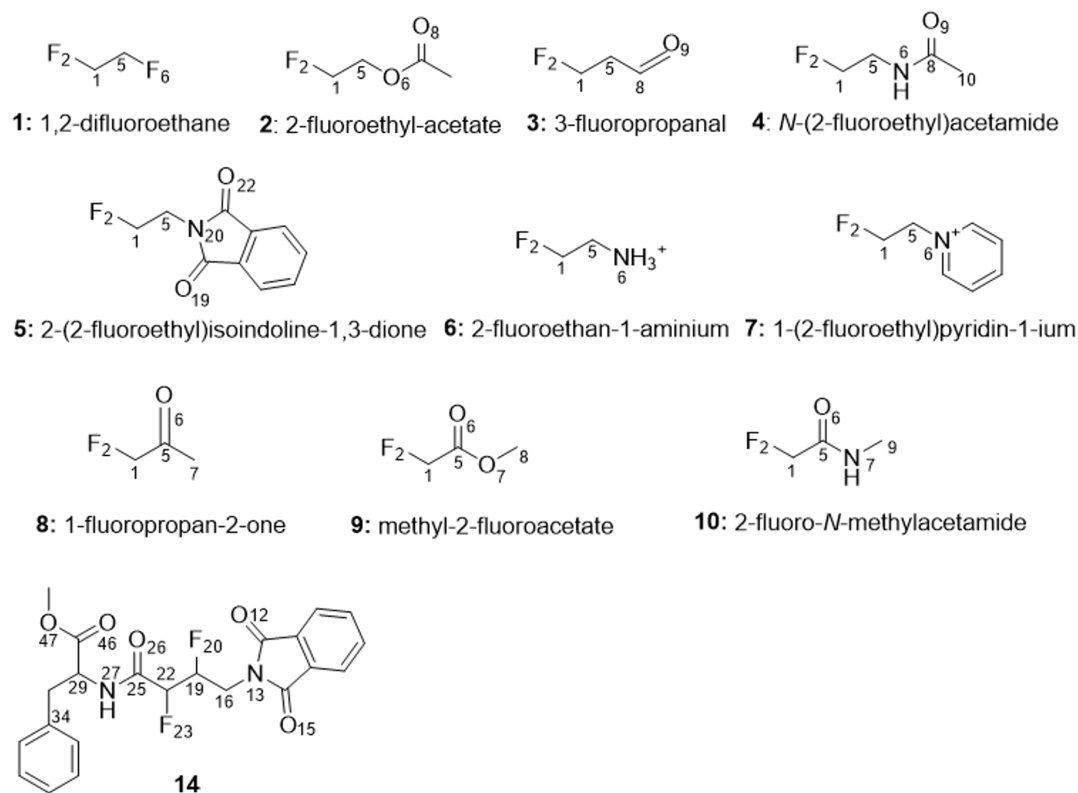
FIGURE 3.6: Fluorine-containing molecules analysed in this Section. These comprise the majority of the full set investigated in Ref. 83, where further details about the IQA conformational analyses can be found.

Figure 3.7 shows the ball-and-stick representation of 1,2-difluoroethane along with the energy difference $\Delta E = E_{gauche} - E_{anti}$ between both conformers. As can be appreciated for the full set of compounds in Ref. 83, Figs. 1-2, ΔE values calculated with HF-D3 are very close to those computed with MP2 and CCSD(T) on the MP2 geometries. These results constitute a validation of the HF-D3 model and, as such, only HF-D3 will be used for the subsequent IQA analyses.

IQA analysis of 1,2-difluoroethane

Starting from the reference case, the energy difference between conformers is very small: -0.3 kcal mol^{-1} in favour of the *gauche* conformer at the HF level. Among the energy contributions favouring this conformation, it is the electrostatic ones ΔE_{class}^{AB} that present the largest differences between isomers. Accordingly, $C_1 \cdots F_6$ and $F_2 \cdots C_5$ interactions favour the *gauche* conformation by -8.0 and -7.9 kcal mol^{-1}, respectively. However, this significant stabilisation of the *gauche* conformer is largely compensated by the strongly repulsive $F_2 \cdots F_6$ and $C_1 \cdots C_5$ contacts, of 12.1 and 6.9 kcal mol^{-1}. Overall, considering the full set of interactions, it is found that both the electrostatic and exchange-correlation components favour similarly the *gauche* conformation ($\Delta E_{class}^{AB} = -2.3$ kcal mol^{-1} and $\Delta E_{class}^{AB} = -1.9$ kcal mol^{-1}).

1: 1,2-difluoroethane ΔE= -0.28 (-0.76) [-0.75]
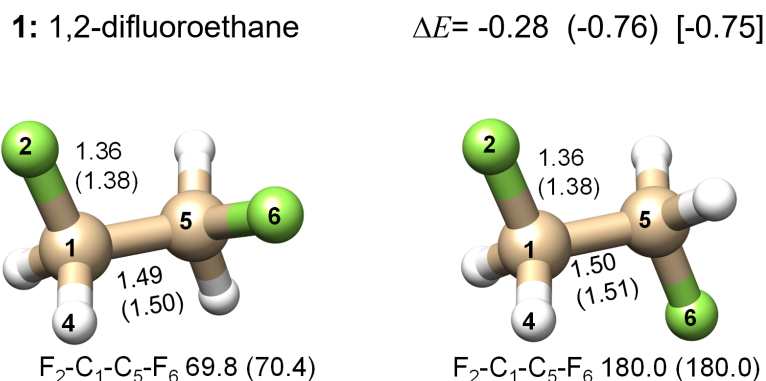
FIGURE 3.7: *Gauche* (left) and *anti* (right) conformations of 1,2-difluoroethane (compound **1**). Distances (Å) comprising C-F and C-C bonds, F-C-C-F dihedral angles (°) and *gauche-anti* energy differences (kcal mol^{-1}) are shown. The numbers without any parenthesis correspond to the HF-D3 results, whereas parenthesis refer to MP2 and brackets to CCSD(T).

An alternative view can be given in terms of interacting quantum fragments (IQF). This way, the four atoms at one and the other side of the rotatable bond can be grouped and their interaction energy studied as that coming from the atoms of one group P interacting with those of the other fragment Q (recall Eq. 2.157). Following this analysis, Table 3.1 shows a more clear picture than the atomic one. On the one hand, the genuine energy contribution placing the *gauche* conformation as the most stable one is the exchange-correlation energy between groups ΔE_{xc}^{PQ} (-3.5 kcal mol^{-1}), accounting for non-classical interactions such as those due to hyperconjugation, in orbital parlance. By contrast, its Coulomb counterpart ΔE_{class}^{PQ} clearly favours the *anti* isomer ($+2.1$ kcal mol^{-1}), partly (55 %) due to the dipole-dipole alignment. It is also worth mentioning that the differences in group net energies ΔE_{net}^{P} and ΔE_{net}^{Q} are very small in comparison with the interaction ones ($0.4 - 0.6$ kcal mol^{-1})[1].

As can be appreciated, alternative views can be obtained by resorting to an atomic or a group description of the system. Both IQA and IQF analyses are probably adequate and useful but, since chemists usually employ a functional group view to explain molecular properties, we believe this higher level of description may be more appropriate, especially when analysing large molecules for which the atomic interpretation may become a mess. Consequently, the following analyses will be carried out in terms of atomic groups.

Substituent effects on conformational preferences of simple fluorine-containing systems

The simplest system turns out to be similarly stable in both *gauche* and *anti* conformations. This fact precludes the usage of vicinal fluorine atoms to modulate conformational preferences in synthetic molecules. Luckily, other conformational effects associated to fluorine can be exploited by changing the groups with which the FCH$_2-$ moiety interacts.

---

[1] ΔE_{net}^{P} and ΔE_{net}^{Q} do not exactly coincide due to some small numerical error introduced during their calculation.

TABLE 3.1: IQF analysis of the energy difference (kcal mol^{-1}) between either *gauche/anti* (for FCH$_2$−CH$_2$X systems, from **1** to **7**) or *cis/trans* (for FCH$_2$−COX, **8** to **10**) conformers. A two-fragment partitioning scheme has been adopted (*P*=CH$_2$F and *Q*=CH$_2$X or *Q*=COX). The IQA reconstructed energies are shown along with their decomposition into net energy variations and different components of the inter-fragment interaction energies. The classical energy is further analysed in low-order multipolar components (charges, c, and dipoles, d).

Sys.	ΔE^{IQA}	$\Delta E_{net}^{P/Q}$	ΔE_{disp}^{PQ}	ΔE_{xc}^{PQ}	ΔE_{class}^{PQ}	ΔE_{cc}^{PQ}	$\Delta E_{cd/dc}^{PQ}$	ΔE_{dd}^{PQ}
1	-0.5	0.6 / 0.4	0.0	-3.5	2.1	0.0	0.0	1.1
2	-0.5	0.7 / 0.5	0.0	-3.1	1.4	0.0	0.0 / 0.0	-0.7
3	-1.2	0.0 / 1.4	0.0	-1.1	-1.5	-0.1	-0.4 / 0.0	-0.4
4	-2.1	-0.6 / 2.1	-0.2	-2.1	-1.4	-0.2	-0.6 / 0.1	-1.0
5	-0.5	2.0 / 1.5	-0.3	-3.6	-0.1	0.0	0.0 / 0.0	0.1
6	-7.3	1.6 / 4.7	-0.6	-6.9	-6.0	2.4	1.3 / -3.5	-4.6
7	-5.1	1.0 / 3.5	-0.3	-5.5	-3.7	0.9	0.2 / -6.1	-0.3
8	-2.8	0.1 / 0.6	0.0	1.5	-4.9	0.0	-0.1 / 0.1	-2.9
9	-0.1	0.6 / 0.5	0.0	0.3	-1.4	0.0	0.7 / 0.2	-2.8
10	-5.8	2.1 / 2.7	-0.2	-4.1	-6.3	0.0	0.2 / 0.1	-5.2

To assess the conformational preferences associated with other groups interacting with the FCH$_2$− unit, the closest systems to FCH$_2$−CH$_2$F with which to start the analysis are those where the second fluorine atom has been replaced by other possibilities. Hence, this Section focuses on CH$_2$F−CH$_2$X species, that correspond to compounds **2-7** in Table 3.1. As can be seen, all *gauche* conformations are energetically favoured, the ionic systems **6** and **7** showing the strongest *gauche* stabilisation. Replacing F by an acetate group (**2**) results in similar energies, with, in this case, a slightly favoured *gauche* conformation by dipole-dipole interactions ($\Delta E_{dd}^{PQ} = -0.7$ kcal mol^{-1}). Aldehyde (**3**) and amide (**4**) groups lead to classical energies favouring the *gauche* conformation (-1.5 and -1.4 kcal mol^{-1}, respectively), similarly as the *xc* component does (-1.1 and -2.1 kcal mol^{-1}), although the ΔE_{net}^Q changes become of the same order than the interaction ones, penalising in this case the *gauche* isomers. On the other hand, the introduction of an isoindole derivative (**5**) gives the ΔE_{xc}^{PQ} energy the preponderance of *gauche* stabilisation, although becomes almost cancelled out by the $\Delta E_{net}^P + \Delta E_{net}^Q$ intra-fragment contributions. Finally, the substitution of F by a cationic group (ammonium in **6** and pyridinium group in **7**) makes the *gauche* conformer undergo a strong stabilisation with respect to the *anti* one. This fact is commonly explained in electrostatic terms but, in reality, the classical energies are surpassed by the *xc* ones in favouring the *gauche* conformation: $\Delta E_{class}^{PQ} = -6.0$ kcal mol^{-1} for **6** and -3.7 for **7**, whereas ΔE_{xc}^{PQ} are, respectively, of -6.9 and -5.5 kcal mol^{-1}. It is nonetheless remarkable that there is a larger energy gap between conformers after substitution by a cationic group, and that the preferences of the classical energies are dominated by charge-dipole and dipole-dipole interactions. In fact, a fraction of the total positive charge is delocalised over the entire molecule, leading to positive charge-charge interactions that increase in the *gauche* orientation. By contrast, it is the charge-dipole and dipole-dipole interactions that dictate the final ΔE_{class}^{PQ} value. All the previous interactions, along with the dispersion ones (that are

quite less remarkable) overcome the also enhanced deformation energies cast into the net energy variations (a total of 6.3 and 4.5 kcal mol^{-1} favouring *anti* in each **6** and **7** compound).

When placing a carbonyl group adjacent to a fluorine atom (FCH$_2$−(C=O)X), *cis* and *trans* minimum energy conformations emerge. In this case, it is assumed that electrostatic effects are responsible for the *trans* preference of the species, led by the antiparallel interacting dipoles of each group (mainly due to the C−F/C=O *trans* bonds). The IQF analysis points toward the same direction: compounds **8** (ketone) and **9** (acetate) present the classical interaction term as the only one favouring the preferred *trans* conformation, the d-d interaction causing a significant contribution to ΔE_{class}^{PQ} where higher order multipole effects seem also relevant. Whereas a high stabilisation of the *trans* conformation is gained after insertion of a ketone ($\Delta E^{IQA} = -2.8$ kcal mol^{-1}), the effect is hampered if another oxygen atom is attached to the carbonyl carbon as in an acetate group ($\Delta E^{IQA} = -0.1$ kcal mol^{-1} only). However, employing a nitrogen in an amide group (compound **9**) leads to an even higher conformational preference ($\Delta E^{IQA} = -5.8$ kcal mol^{-1}), thus enhancing both the d-d interactions (-5.2 kcal mol^{-1}) and leading the *xc* energy to strongly favour the *trans* conformer in this case (-4.1 kcal mol^{-1}, although it is even overcome by the large deformations pointed out by $\Delta E_{net}^{P} + \Delta E_{net}^{Q} = 4.8$ kcal mol^{-1}).

Up to this point, it has been shown that IQA energies capture both stereoelectronic and electrostatic effects, the latter being possible to be further decomposed into multipolar contributions (e.g., c-c, c-d or d-d interaction energies). This is an aspect that has not been considered in other studies, but that may become very insightful as it reveals counterintuituve effects in some cases.

Competition between several conformational effects

The previous studies of model molecules focus on the specific impact that different functional groups have on conformational preferences when interacting with a fluorinated moiety. For larger molecules, however, these effects compete to finally prioritise one conformation over others. In the following lines, an *α,β*-difluoro-*γ*-aminoacid derivative will be analysed: **14** (diastereomer **14a** in Ref. 83, where additional compounds **11**, **12**, **13** and **14b** complete the study herein summarised).

The conformational preferences of **14** are largely determined by intramolecular factors, as the most stable conformer in the gas phase is quite close to the crystallographic ones. This structure presents F−C−C−F and F−C−C−N units in *gauche*, which are generally expected to be energetically favoured. However, there are other cases of low energy conformers that escape from the simple rationalisation according to *gauche* preferences and four of them were considered for the IQF analysis: **conf1**, **conf2**, **conf5** and **conf6**, depicted in Figure 3.8.

Starting from the relative stability (see Figure 3.8 for the HF energies and Table 3.2 for the IQA reconstructed ones) of each conformer, **conf2** is slightly more stable than **conf1**, whereas **conf1** is clearly favoured with respect to **conf5** and **conf6**. At first sight it might seem that minor changes occur when passing from **conf1** to **conf2**, but the IQF analysis (Table 3.2) reveals profound changes affecting both the net energies (deformation) and the interaction between groups, leading to a nearly mutual cancellation reflected in the small value of ΔE. The main appreciable change from **conf1** to **conf2** is the shorter contact between the phenyl group (included in *O*) and the isoindol derivative (in group *R*). Accordingly, ΔE_{disp}^{OR}, ΔE_{xc}^{OR} and ΔE_{class}^{OR} dominate each energy change (see Table 3.3).
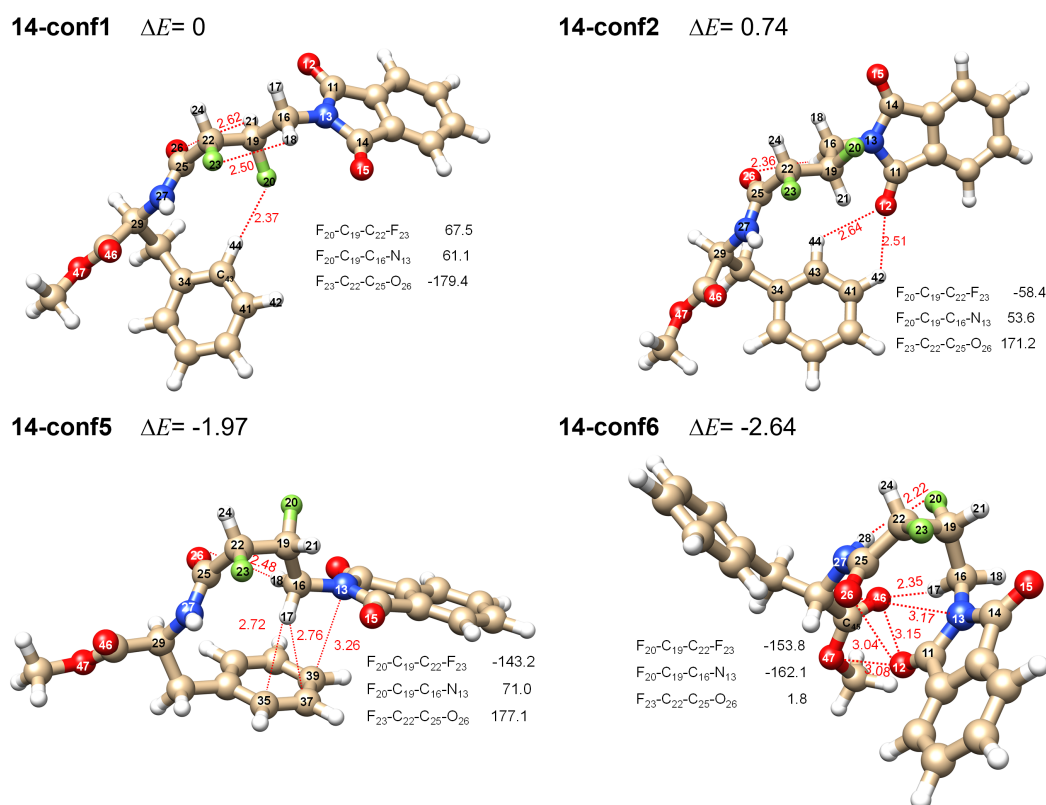
FIGURE 3.8: Selected conformations of compound **14**. HF-D3 Energies with respect to conformer 1 are given in kcal mol^{-1} along with distances (Å) associated with the most remarkable contacts and F−C−C−F, F−C−C−N and F−C−C=O dihedral angles (°).

When assessing **conf5** and **conf6**, notable structural changes are observed. In addition to the *anti* conformation of fluorine atoms, the most remarkable feature in **conf5** is the alignment of groups *O* and *R*, resulting in a high contribution of dispersion energy favouring **conf5** (9.8 kcal mol^{-1}) emanating essentially from ΔE_{disp}^{OR} (12.5 kcal mol^{-1}). Similar results are obtained for the helical **conf6** (a total $\Delta E_{disp} = 9.9$ kcal mol^{-1} for which $\Delta E_{disp}^{OR} = 12.8$ kcal mol^{-1}), although, in this case, the acetate group of *O* is responsible since the phenyl group remains now aside from the other parts of the molecule. The **conf5** and **conf6** conformations are also favoured by electrostatics and, much more importantly, by *xc* interactions, that can become larger than 20 kcal mol^{-1}, favouring both cases with respect to **conf1**. Nonetheless, it is **conf1** that is globally preferred. The total change undergone in a process, such as formation of a new species, is balanced by both the interaction energy and the deformation one, and it is the intra-group energies that largely penalise both **conf5** and **conf6** conformations (especially those affecting the *O* and *R* groups), resulting in the final observed preference of **conf1** geometry.

From the previous analysis, it has become clear that it is the competition of several contributions that yields a final preferred structure. In the case of compound **14** the effects associated to fluorine are diluted in the global energies, for which the interactions between other groups can be of a greater importance. It is also remarkable that a bare explanation of conformational preferences in terms of the interactions that take place between several groups may result insufficient. Rather, deformation energies can play a major role and be determinant in the conformational preferences

TABLE 3.2: IQF energy decomposition for the energy differences (kcal mol^{-1}) of conformers 2, 5 and 6 with respect to 1 of compound **14**. The fragments chosen have been $O=CH_3OCOCH(CH_2Ph)NHCO-$, $P=-CHF-$, $Q=-CHF-$, and $R=-CH_2$-isoindol-1,3-dione.

Conf.	ΔE^{IQA}	ΔE_{net}^{O}	ΔE_{net}^{P}	ΔE_{net}^{Q}	ΔE_{net}^{R}	ΔE_{disp}	ΔE_{xc}	ΔE_{class}
2	0.3	-5.5	-1.2	2.8	-9.7	3.4	8.9	1.7
5	-2.2	-17.5	-1.0	4.5	-19.8	9.8	20.3	1.5
6	-2.1	-16.9	-3.4	3.2	-21.7	9.9	21.4	5.3

TABLE 3.3: Components of the interaction energy differences between groups (kcal mol^{-1}).

Conf.	ΔE_{disp}			ΔE_{xc}			ΔE_{class}		
	OP	PQ	QR	OP	PQ	QR	OP	PQ	QR
	OQ	OR	PR	OQ	OR	PR	OQ	OR	PR
2	0.1	0.3	0.2	1.6	1.5	0.2	-0.8	-0.9	0.8
	-0.9	4.5	-0.8	-7.6	17.6	-4.4	-1.0	4.7	-1.1
5	0.1	0.3	-0.4	2.2	-2.5	-0.5	0.5	2.8	-0.2
	-1.8	12.5	-0.9	-9.6	35.2	-4.5	-3.2	2.8	-1.2
6	-0.4	0.3	-0.7	-2.4	-1.5	-4.6	-6.7	2.5	2.0
	-0.8	10.8	0.7	-4.9	33.7	1.1	1.1	6.5	-0.1

of medium- and large-sized molecules, thus penalising the favoured structure by inter-fragment interactions and giving rise to a probably unexpected preferred conformation. All these conformational effects can be well characterised by IQF analyses with a simple HF-D3 wave function, whose application to larger biomolecules of interest might open new scenarios for the rational design of functionalised derivatives favouring specific conformations.

3.1.3 Assessing atomic and group contributions to the electrostatic free energy of solvation

The interacting quantum atoms energy partition was initially developed for wave function-based methods, from which the 1- and 2-RDM are available. Later on, this formalism has been broadened with the implementation of DFT methods [63] and empirical dispersion corrections [64]. In fact, the essence of IQA relies on a real space partition of the molecule under consideration, but once the atoms have been defined, any additional pairwise potential can in principle be incorporated to the underlying energetic decomposition, as well as the effect of an external potential on each atom, such as that caused by a continuum solvent medium.

In Statistical Mechanics, the free energy of solvation is a non-additive magnitude because its calculation involves the ensemble averaging of an exponential term that cannot be factorised [84]. This is translated into the fact that the coordination spheres of atoms are correlated. Nonetheless, different investigations have been devoted to predicting free energies of solvation from group contributions that obviously rely on an additivity assumption [85–88]. Moreover, as explained in Section 2.5.9, it is possible to decompose the electrostatic interaction energy of a solute with an implicit solvent, what is followed by the *effective* atomic decomposition of the electrostatic contribution to the free energy of solvation.

Having as a final goal the exploitation of IQA in computational biochemistry, it is undeniable that accounting for solvation effects into the IQA partition is mandatory. In this work we have applied the decomposition of the (electrostatic) free energy of solvation presented in Section 2.5.9 both to atoms and groups. For this, a temptative definition of functional groups has been proposed and, with them, the additivity assumption of the electrostatic free energy of solvation has been tested. The chosen continuum solvation model has been the conductor-like screening model (COSMO) [89, 90], an ASC-like method that provides a set of point charges surrounding the molecule and from which the interaction with the solute charge density is readily implemented. Such an IQA descomposition of the solute-solvent interaction has been coded in a new modular version of PROMOLDEN, that provides the separate atomic interaction energies with the COSMO charges. These tesserae point charges, along with their coordinates, are obtained using a locally modified version of the GAMESS code [91] that has been tested for a total set of 412 molecules from the Minnesota solvation database (MNSol) [92, 93], comprising 309 neutral molecules as well as 49 cationic and 57 anionic species. These have been fully optimised both in the gas phase and in the continuum solvent (for which a dielectric constant $\varepsilon = 80$ for mimicking water has been selected) at the HF/aug-cc-pVTZ level and their energies have been compared with the experimental ones.
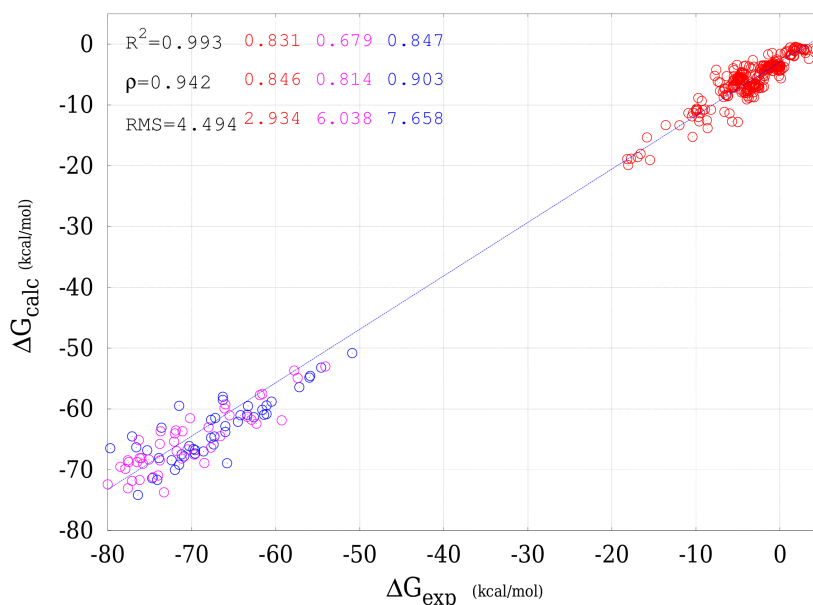
FIGURE 3.9: Comparison between the COSMO-HF hydration energies (ΔG_{calc}) and the experimental ones (ΔG_{exp}) for the selected structures from the MNSol. The correlation between them is studied by means of the coefficient of determination (R^2), the Spearman correlation coefficient (ρ) and the root mean square errors (RMS). Values corresponding to the whole set are given in black, while those ascribed to neutral species are in red, anionic in blue and cationic in magenta.

As can be seen in Figure 3.9, there is a strong correlation ($R^2 = 0.993$) between the experimental free energies and those calculated at the HF level through the COSMO model. In spite of this, the root mean square error is high ($RMS = 4.5$ kcal mol^{-1}), what can be due to the lack of electron correlation and non-polar contributions to the solvation energies, as well as the absence of conformational sampling. In any case,

the HF results are reasonable, capturing the trends shown by the experimental values. It is also remarkable how the accuracy and adequacy of the description depends on the charge state of the systems. Whereas neutral and anionic species show fairly good correlations ($R^2 = 0.831$ and 0.847, respectively), cationic molecules provide worse values ($R^2 = 0.679$). It may be noticed, however, that other levels of theory coupled with more sophisticated solvent models exhibit a similar performance (see Ref. 68 for more details).

IQA decomposition of solvation energy

Before starting with the analysis of the IQA electrostatic solvation energies it is worth evaluating the extent of numerical error these energies are subject to. To estimate it, the IQA-reconstructed energies have been compared with those calculated with COSMO-HF. The IQA numerical error in the decomposition of hydration energies turns out to be similar to the previously reported one for the formation energies of non-covalent complexes [64]: a mean value of 0.76 ± 1.36 kcal mol^{-1}, that corresponds to an average error per atom of 0.06 ± 0.10 kcal mol^{-1}. In the case of fragments, our actual focus, these range from 0.5 to tens of kcal mol^{-1}.
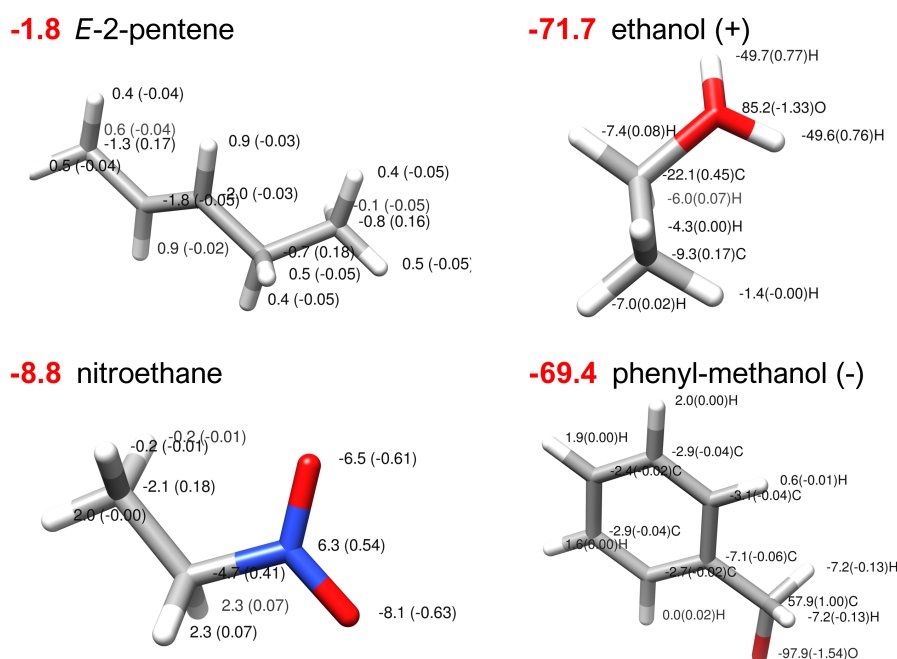
FIGURE 3.10: ΔG^A_{solv} (kcal mol^{-1}) for two neutral (left) and two ionic (right) species. The atomic charges are also indicated (in parentheses) as well as the total ΔG_{solv} (in red) of the molecule.

Figure 3.10 presents a set of four molecules for which the global ΔG_{solv} values as well as the effective atomic ones ΔG^A_{solv} are given. As can be seen, the neutral molecules (on the left) present electrostatic free energies of solvation one order of magnitude smaller (in absolute value) than those reported for charged species (on the right). The atomic contributions of C and H in hydrocarbon molecules are quite small (about $0.5 - 2.0$ kcal mol^{-1} in absolute values), although they increase in strength when attached to polar atoms. As expected, polar groups concentrate stronger hydration contributions, thus inducing larger hydration energies (e.g.,

$\Delta G_{solv} = -8.8$ kcal mol^{-1} for the nitroethane, its polar atoms comprising ΔG_{solv}^A of -6.5, $+6.3$ and -8.2 kcal mol^{-1}). This fact is enhanced in charged species. For example, protonated ethanol presents a ΔG_{solv} of -71.7 kcal mol^{-1}, the most charged atoms contributing -49.7, -49.6 and $+85.2$ kcal mol^{-1}, respectively. We also examined the correlation between atomic charges and ΔG_{solv}^A values: the correlation plot found in Figure 3.11 shows that a very good correlation exists both for the anions ($R^2 = 0.984$) and for the cations ($R^2 = 0.970$). This result reveals that hydration of ionic solutes is closely determined by the distribution of the Bader atomic charges. In turn, electrostatic hydration energies of neutral solutes are probably controlled by other multipolar terms accounting for the anisotropy of the electron distribution, such as dipoles or quadrupoles.
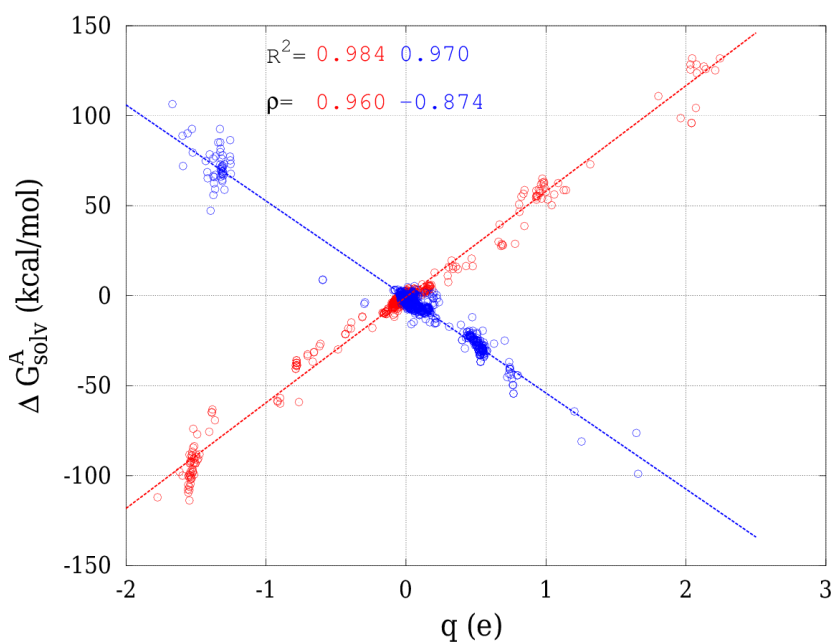
FIGURE 3.11: Correlation between QTAIM atomic charges and corresponding ΔG_{solv}^A atomic contributions to solvation. Data in blue correspond to cations, whereas red represents anions.

Assessment of fragment contributions to solvation energy

Once the main atomic features have been inspected, it is our belief that the best strategy to analyse IQA solvation contributions is to adopt a *united atom* approach. This way, a given organic molecule can be built from these smaller building blocks, be them specific atom types or functional groups. The set of fragments used in this work relies on the 130 fragments defined by Meylan and Howard to assess atom /fragment contribution to a quantity related to ΔG_{solv} [94]. Ours are listed in Table 1 in the original manuscript (Ref. 68) and comprise 51 atom types/functional groups appearing at least 5 times in the MNSol structures chosen.

Focusing on the IQA fragment decomposition of the electrostatic solvation energy (ΔG_{solv}^G), a distribution of group solvation energies is obtained when analysing the different MNSol molecules. Hence, an appropriate manner of studying the distribution of energies obtained is by means of histograms. Figure 3.12 collects six examples of the 51 groups analysed. On the one hand, the methylene group ($-CH-$) shows a distribution of energies from negative to positive values, but concentrated

around its mean value, a positive and thus destabilising contribution to the solvation energy. Different parts in the distribution can be associated to different environments the group finds. Hence, the main peak corresponds to methylene attached to 0 or 1 polar groups, whereas the shoulder is due to the presence of polar atoms at both sides of the group. When incorporating a charge in the molecule the picture may change dramatically. It is the case of methylene groups found in anionic species ($-$CH$-$[-]), that present a much wider and homogeneous distribution of energies, with no central peak representative of the whole distribution (see Figure 3.12). Polar groups may present very different distribution patterns. Chlorine, for instance, shows a wide and flat histogram, the most favourable contributions stemming from monosubstituted aliphatic hydrocarbons. By contrast, the amine group ($-$NH$_2$) — and, in general, polar groups different from halogen atoms— presents a very narrow range of hydration energies. Even when this group becomes protonated, $-$NH$_3$($\equiv$) gives concentrated ΔG^G_{solv} values. Again, the situation is not hold when an anionic group is inspected, and alkoxide ($-$O($-$)) results in scattered data over a wide range of 40 kcal mol^{-1}.
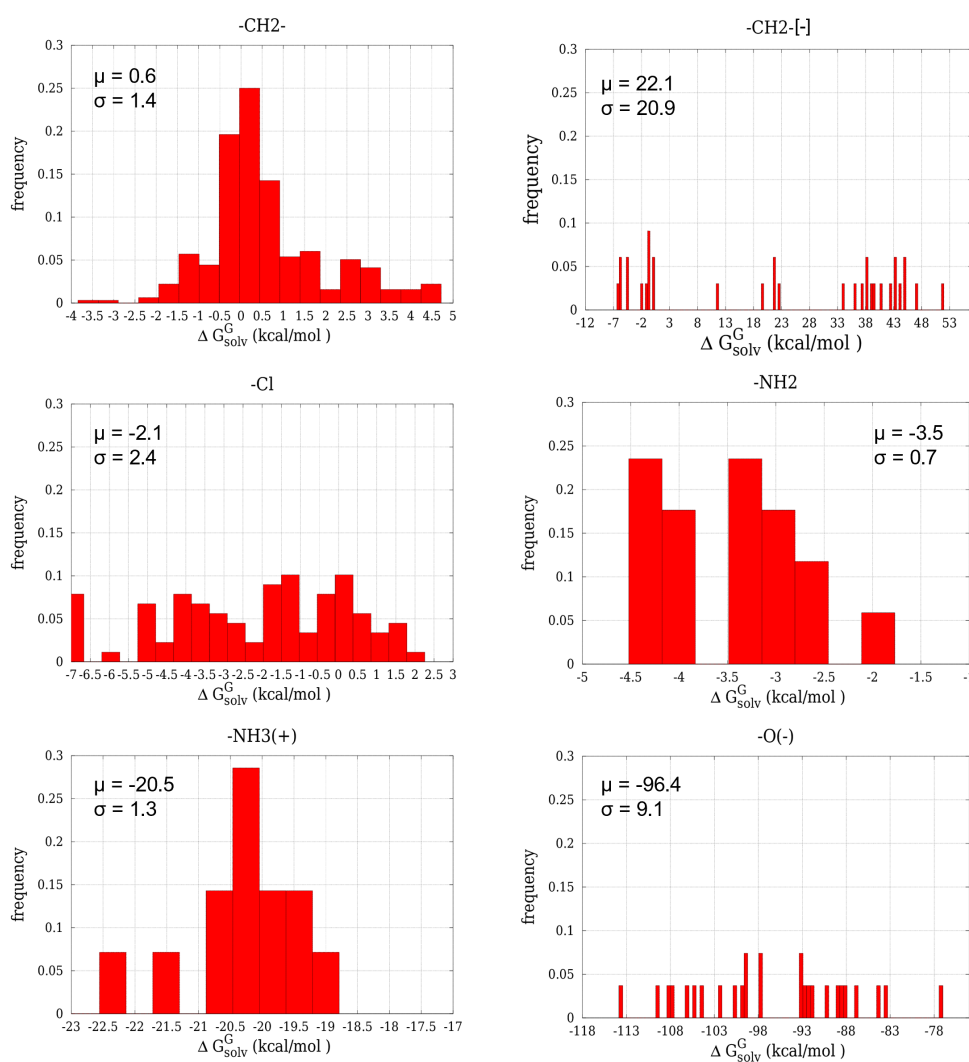
FIGURE 3.12: Histograms showing the distribution of ΔG^G_{solv} in six selected neutral and ionic groups. The mean (μ) as well as the standard deviation (σ) are also shown.

As seen, the distribution of ΔG^G_{solv} energies is heterogeneous: some groups present a narrow energy histogram, whereas others show flat, wide distributions that hamper the association of their mean value with a representative group energy for different situations. This feature is more frequently found in charged species, both in specifically charged groups and also in neighbouring fragments.

Additivity of fragment contributions

The analysis of the ΔG^G_{solv} energy distribution therefore confirms the ability of the IQA energy decomposition of the COSMO-HF solvation energies to provide a detailed assessment of fragment contributions to solvation. Within the different kinds of groups, those belonging to neutral molecules present the least dispersion (low σ values generally < 2 kcal mol^{-1}; see Table 1 in the original publication [68]). Accordingly, it seems reasonable to check whether or not these group energies are additive. With this aim, a final analysis shown in Figure 3.13 relates the IQA-reconstructed free energies ($\Delta G_{estimated} = \sum_G \langle \Delta G^G_{solv} \rangle$) from the different group mean values with the calculated ones with COSMO-HF (ΔG_{calc}) for a set of 32 MNSol molecules not considered initially. The computed and estimated values show only a moderate correlation ($R^2 = 0.705$) with an *RMS* error close to 6 kcal mol^{-1}. It is found that the largest discrepancies correspond to compounds with large aliphatic or aromatic moieties, while those presenting two or more polar groups tend to agree to a larger extent.
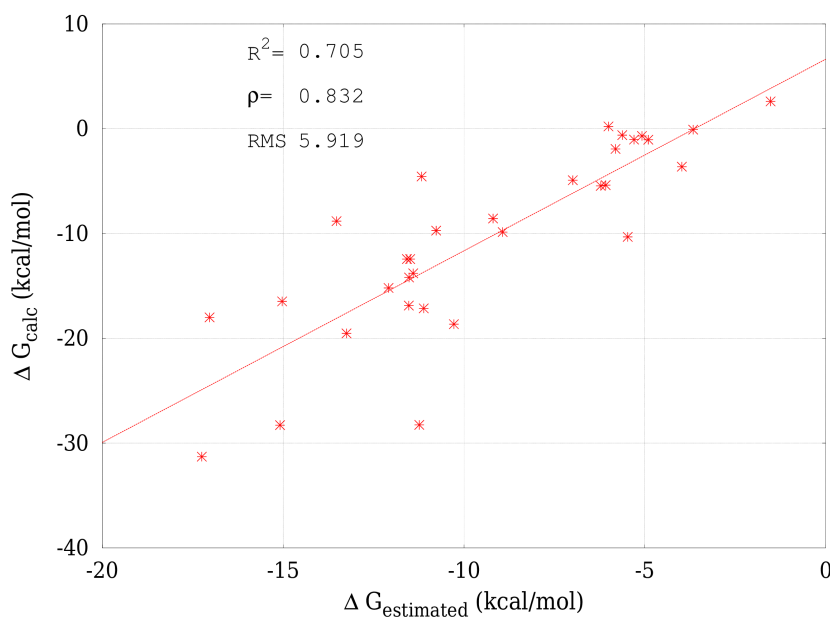
FIGURE 3.13: Correlation plot of the COSMO-HF calculated eletro-static hydration energies (ΔG_{calc}) with those estimated from fragment mean values (ΔG_{estim}).

The model tested herein is not intended to provide a working additive solvation model, but it is rather the result of an initial assessment of IQA-based fragment contributions to the electrostatic free energy of solvation. As has been proven, IQA opens a field to be further explored given that the distribution of the group contributions examined in this work results in some structure-activity relationships. Hence, the detailed analysis induced by the IQA decomposition may become a useful tool

for ascribing solvation energetic contributions to particular functional groups in a consistent manner. Nonetheless, this research is open to more precise atomic grouping (further distinguishing for example between halogen atoms attached to aliphatic carbons or aromatic ones), correction factors accounting for the influence of nearby ionic or polar groups, or the inclusion of non-electrostatic effects such as those provided by the G_{CDS} empirical potential, that belongs to the SMD model [95] and whose incorporation to IQA has been recently accomplished.

3.1.4 A quantum chemical topology picture of intermolecular electrostatic interactions and charge penetration energy

A correct description of electrostatic interactions is essential in molecular modelling. The relative abundance of polar atoms and charged species makes electrostatics play a major role in the control of the stability and activity of many biomolecules such as proteins, nucleic acids and lipids [96, 97]. In this respect, molecular mechanics (MM) potentials relying on point-charge or multipolar descriptions of the anisotropic charge distribution have led to accurate electrostatics in the long range. However, the divergence with the exact computation of the interaction between continuous charge densities at short distances has been a matter of concern. For this reason, corrections aimed at incorporating the so-called *charge penetration* (CP) energy —that is, the energy difference between the exact Coulomb interaction between charge densities and that approximated by a multipolar series— have been proposed [98–102]. These are rooted in the symmetry-adapted perturbation theory (SAPT) electrostatic energy [35] —known as first-order polarisation energy within this model—, an energy decomposition analysis (EDA) that, indeed, somehow mimics the separation of energy contributions done in MM methods. Therefore SAPT seems a reasonable theoretical reference to help parameterise MM potentials. A question on its appropriateness raises, however, when evaluating different EDAs. In fact, different interpretations of the energetic effects involving the overlap of the electron densities of two molecular fragments may be possible depending on the particular EDA of choice.

In the present work [103] we decided to reexamine the nature of intermolecular electrostatic interactions under the prism of an orbital-invariant, reference free EDA, IQA, with a twofold goal. On the one hand, we aim to compare in a consistent and systematic manner the atomic and fragment contributions to the total electrostatic energy through a hierarchy of QM and MM methodologies. On the other hand, we examine the CP concept and establish clear links with our real space methodolody, for which we propose a novel definition of CP based on a joint orbital and real space decomposition.

With the previous purposes, we examined the S66 dataset of non-covalent complexes as well as a selection of 12 complexes from the S66x8 dataset, which constitutes an extension of the S66 set to eight fractions of the equilibrium geometries. HF-D3(BJ)/cc-pVTZ calculations were preformed on the S66/S66x8 geometries, both on the entire complex and on the separate monomers, followed by IQA analyses and the computation of QTAIM multipoles. For the sake of comparison, restrained electrostatic potential (RESP) atomic charges were computed following the general Amber force field (GAFF) prescriptions [104], as well as AMOEBA (atomic multipole optimised energetics for biomolecular simulation) multipoles [105] up to the quadrupoles (for more details, see Ref. 103).

Zeroth-order electrostatic energy and charge penetration

In the Methodology section (Eq. 2.162), the expression for the electrostatic energy between two charge densities has been presented. For weakly interacting compounds, the total charge density of a complex AB ($\rho_{tot,AB}(\mathbf{r}) \equiv \rho(\mathbf{r})$ in the following) can be approximated by the sum of the unperturbed monomer densities as

$$\rho^0(\mathbf{r}) = \rho_A^0(\mathbf{r}) + \rho_B^0(\mathbf{r}), \tag{3.2}$$

that we shall call the *zeroth-order approximation*. For this case, the electrostatic contribution to the formation energy of the complex corresponds to

$$\Delta E_{elec}^0 = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho_A^0(\mathbf{r}_1)\rho_B^0(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2. \tag{3.3}$$

This zeroth-order electrostatic energy also corresponds to the first-order polarisation energy in SAPT, and serves as the reference to define the CP energy E_{pen} through

$$E_{pen} = \Delta E_{elec}^0 - \Delta E_{elec,mp}^0, \tag{3.4}$$

where $\Delta E_{elec,mp}^0$ is the multipolar analogue of ΔE_{elec}^0. Following its formal definition some aspects can be pointed out. On the one hand, and recalling the IQA definition of the formation energy (Eq. 2.159), both intra- and intermolecular contributions are mixed in E_{pen} since a global contribution to the formation energy involves net energy changes (intramolecular) and those associated with inter-fragment interaction energies (intermolecular). A second point concerning E_{pen} is that $\Delta E_{elec,mp}^0$ can be affected by a truncation error (depending on the order at which the expansion series is truncated) that mixes with the energy error due to the inadequacy of applying the multipole expansion for interpenetrating charge densities, what would correspond to a *pure* error of the multipolar description due to penetration.

If the zeroth-order approximation is combined with a Bader partitioning of the three-dimensional space into the basins of attraction of each molecule in the complex ($\mathbb{R}^3 = \Omega_A + \Omega_B$), the total electrostatic energy of AB becomes split in several terms:

$$\begin{aligned}
E_{elec}^0 = {}& E_{elec}^A(\rho_A^0, \rho_A^0) + E_{elec}^A(\rho_B^0, \rho_B^0) + E_{elec}^A(\rho_A^0, \rho_B^0) \\
+ {}& E_{elec}^B(\rho_A^0, \rho_A^0) + E_{elec}^B(\rho_B^0, \rho_B^0) + E_{elec}^B(\rho_A^0, \rho_B^0) \\
+ {}& E_{elec}^{AB}(\rho_A^0, \rho_A^0) + E_{elec}^{AB}(\rho_B^0, \rho_B^0) + E_{elec}^{AB}(\rho_A^0, \rho_B^0) + E_{elec}^{BA}(\rho_A^0, \rho_B^0),
\end{aligned} \tag{3.5}$$

each term representing the electrostatic energy due to the interaction between the fraction of the densities in parentheses contained in the respective basins, represented by the superscripts in the order given. This way, $E_{elec}^{BA}(\rho_A^0, \rho_B^0)$ stands for $\int_{\Omega_B} d\mathbf{r}_1 \int_{\Omega_A} d\mathbf{r}_2 \rho_A^0(\mathbf{r}_1)\rho_B^0(\mathbf{r}_2) r_{12}^{-1}$ and $E_{elec}^B(\rho_A^0, \rho_B^0) = \frac{1}{2} \int_{\Omega_B} d\mathbf{r}_1 \int_{\Omega_B} d\mathbf{r}_2 \rho_A^0(\mathbf{r}_1)\rho_B^0(\mathbf{r}_2) r_{12}^{-1}$, for which the single superscript denotes a double integration over the same basin.

When applying the same decomposition to one of the fragments (such as A; analogously for B) within the same Bader partitioning of the real space (i.e., $\Omega_A + \Omega_B$), the total electrostatic energy of the original species results from the sum

$$E_{elec}(\rho_A^0, \rho_A^0) = E_{elec}^A(\rho_A^0, \rho_A^0) + E_{elec}^B(\rho_A^0, \rho_A^0) + E_{elec}^{AB}(\rho_A^0, \rho_A^0), \tag{3.6}$$

therefore encompassing both intramolecular (same-basin interactions) and intermolecular (between different molecular basins) contributions in the partition induced by $\rho^0(\mathbf{r})$.
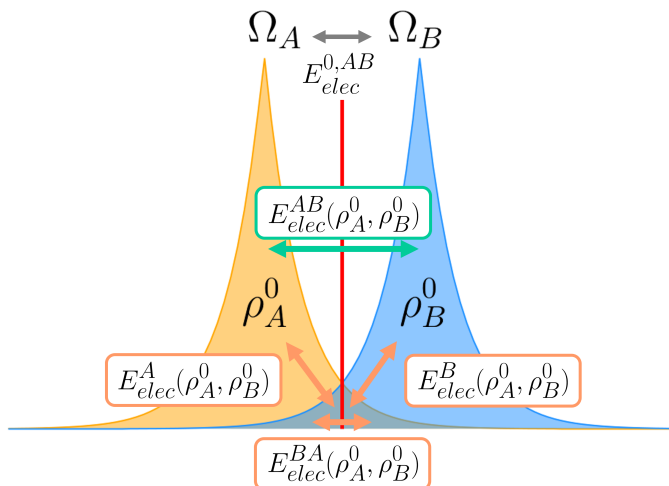
FIGURE 3.14: Schematic representation of the four contributions giving rise to ΔE^0_{elec} from the double Bader's and zeroth-order density partitioning. The IQF electrostatic interaction term between two molecular basins Ω_A and Ω_B is also depicted to highlight its difference with $E^{AB}_{elec}(\rho^0_A, \rho^0_B)$, as $E^{0,AB}_{elec}$ accounts for the interaction between the total densities inside each basin (the original ρ^0_A or ρ^0_B and the tail of the opposite one that has penetrated into either Ω_A or Ω_B).

By subtracting the fragment energies $E_{elec}(\rho^0_A, \rho^0_A)$ and $E_{elec}(\rho^0_B, \rho^0_B)$ from E^0_{elec} in Eq. 3.5, the electrostatic contribution to the formation energy is recovered:

$$\Delta E^0_{elec} = E^A_{elec}(\rho^0_A, \rho^0_B) + E^B_{elec}(\rho^0_A, \rho^0_B) + E^{AB}_{elec}(\rho^0_A, \rho^0_B) + E^{BA}_{elec}(\rho^0_A, \rho^0_B). \qquad (3.7)$$

Among the four surviving terms (whose graphical representation is shown in Figure 3.14), $E^{AB}_{elec}(\rho^0_A, \rho^0_B)$ is the ordinary interaction term between density ρ^0_A associated with molecule A and density ρ^0_B corresponding to basin B. It indeed matches ΔE^0_{elec} at long distances ($\lim_{R_{AB}\to\infty} E^{AB}_{elec}(\rho^0_A, \rho^0_B) = \Delta E^0_{elec}$), whereas the other three terms present an opposite behaviour, becoming more and more relevant when shortening R_{AB}. As such, those terms can be directly related to the energy associated with the interpenetration of molecular densities, thus being grouped in an IQF-like —we shall distinguish between IQA and IQF when dealing with either atomic or fragment decompositions— electrostatic charge penetration energy

$$E^{IQF}_{elec,pen} = E^A_{elec}(\rho^0_A, \rho^0_B) + E^B_{elec}(\rho^0_A, \rho^0_B) + E^{BA}_{elec}(\rho^0_A, \rho^0_B), \qquad (3.8)$$

that satisfies $\lim_{R_{AB}\to\infty} E^{IQF}_{elec,pen} = 0$ and so its three components.

IQF-D3 partition and pairwise approximation

In search for the most appropriate IQF descriptor of binding, it is found that the pairwise electrostatic term in conjunction with D3 dispersion ($E^{AB}_{elec} + D3$) yields the best correlation and lowest errors with the benchmark formation energies (see Figure 3.15, left), a global trend that is widely maintained in the three main groups in which the S66 complexes can be classified, namely H-bonded (represented in magenta), dispersion-dominated (turquoise) and, to a slightly lesser extent, mixed (in yellow) complexes. In fact, the use of the total electrostatic contribution ΔE_{elec}

(comprising both the inter-fragment interaction terms and the intra-fragment deformation ones) results in a poorer correlation and higher *RMS* errors ($R^2 = 0.888$, $RMS = 17.3$ kcal mol^{-1}), what is explained on the basis of the unbalanced description provided by the inclusion of intramolecular energies that tend to cancel out with the kinetic and *xc* energy changes in the formation process and that are absent in the present description (see Figure 3.15, right).
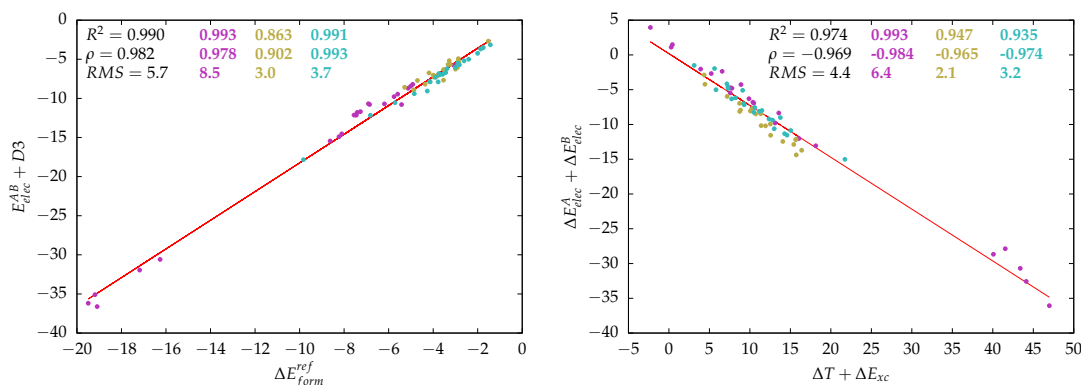
FIGURE 3.15: Left: correlation between the dispersion-augmented IQF intermolecular electrostatic energy $E_{elec}^{AB} + D3$ and the S66 reference binding energies ΔE_{form}^{ref}. Right: anticorrelation featured by the intrafragment electrostatic contribution to formation $\Delta E_{elec}^{A} + \Delta E_{elec}^{B}$ and the total kinetic plus exchange-correlation contributions $\Delta T + \Delta E_{xc}$. The statistical measurements comprise the coefficient of determination R^2, Spearman's rank correlation coefficient ρ and the root mean square error *RMS*. All the energies are in kcal mol^{-1}. Data corresponding to the whole set of complexes is depicted in black, that ascribed to the H-bond group in magenta, while mixed and dispersion complexes are in yellow and turquoise, respectively.

When approximating the wave function of the system as the Hartree product of the independent monomeric ones (i.e., the zeroth-order approximation), the resulting zeroth-order IQF terms plus dispersion $E_{elec}^{0,AB} + D3$ produce similar results than the previous ones for the fully relaxed systems. Indeed, the *RMS* errors remain low (3.1 kcal mol^{-1}) and the correlation fairly good ($R^2 = 0.971$), although that of the mixed subset becomes somewhat deteriorated ($R^2 = 0.755$). This happens at the equilibrium geometries, but seemingly the same conclusions can be extended to other intermolecular separations R_{AB}, in light of the similar trends exhibited by both descriptions in Figure 3.16 (the complete set of 12 S66x8 complexes investigated can be found in the original article 103). In spite of this, significant deviations anticipating CP (and other effects such as charge transfer or charge polarisation) emerge for short distances. Moreover, the magnitude of these effects as well as the shape and slope of the E_{elec}^{AB} and $E_{elec}^{0,AB}$ curves are clearly system-dependent, thus revealing further details about the role of electrostatics in these complexes.

Comparison between $E_{elec}^{0,AB}$ and the pairwise MM energies

The proved validity of both the pairwise and the zeroth-order approximations provide an insightful theoretical support for the construction and analysis of MM electrostatic potentials. Accordingly, QTAIM and AMOEBA multipolar energies (up to the quadrupolar interactions) along with those provided by RESP atomic charges have been compared with the IQF terms. According to the statistical descriptors,

the three multipolar/point-charge energies correlate considerably well with $E_{elec}^{0,AB}$, yielding $R^2 > 0.95$ and $RMS \approx 1$ kcal mol^{-1} for the whole S66 set and also similar values for the H-bond and dispersion groups. The less abundant mixed complexes present a lower agreement, QTAIM/AMOEBA multipoles yielding a higher correlation degree than the RESP point charges ($R^2 \simeq 0.6 - 0.8$ for the former vs. 0.5 for the latter). Insightful is also the evolution of these energy terms with the intermolecular separation. As can be appreciated in Figure 3.16 (and the rest of the systems in Ref. 103, Figure 3), the point-charge and multipolar energies evolve quite closely to $E_{elec}^{0,AB}$, what suggests that the fitting procedure carried out for the derivation of the RESP charges may incorporate in an effective way higher order effects, given the similarity between these monopoles with the, in principle, more complete descriptions that consider dipoles and quadrupoles.
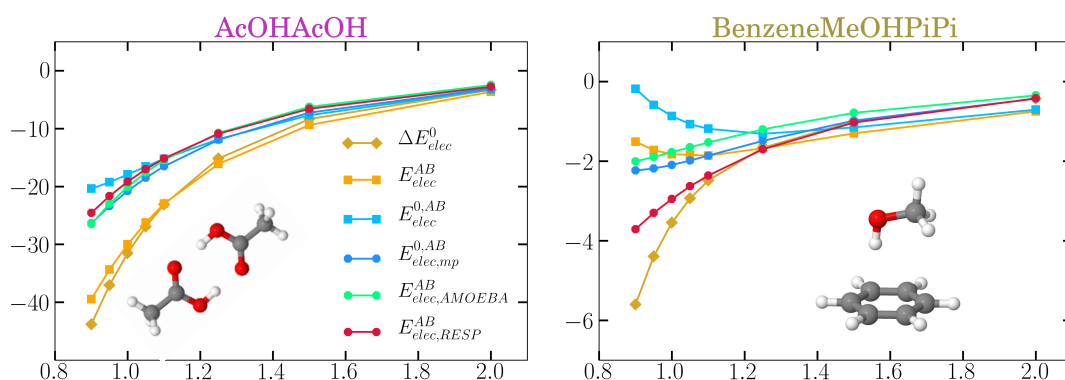
FIGURE 3.16: Intermolecular electrostatic interaction energies (in kcal mol^{-1}) as a function of the distance (relative to the equilibrium distance, R_{AB}/R_{eq}) for the acetic acid dimer (H-bond class) and the benzene-methanol complex (mixed group). For comparative purposes, the total zeroth-order electrostatic contribution to formation is also included.

Comparison of diatomic electrostatic interactions

So far, global molecular energies have been inspected. However, a more detailed analysis in terms of interacting atomic pairs is feasible thanks to the IQA ability to unambiguously decompose global energies into atomic ones. QTAIM multipoles, as they follow the same atomic partitioning, also provide coherent multipolar descriptions. In turn, the RESP/AMOEBA potentials are based on distributed charges/-multipoles that are located at nuclear positions, thus making a comparison of the different methodologies at this level very pertinent.

Establishing the IQA term $E_{elec}^{0,IJ}$ as the *reference* to compare with, Figure 3.17 shows a non-surprising almost perfect match between the IQA energies and the QTAIM multipolar ones. As mentioned, both methods follow the same partition of the density, giving rise, on the one hand, to exact energies by integration of the continuous densities contained inside the atomic basins and, on the other hand, those approximated by the multipoles derived from exactly the same densities. Only significant deviations are found for polar contacts, especially for some HO$\cdots$H H-bonds, the rest of the interactions presenting much lower differences (Figure 3.17, top right). On the contrary, AMOEBA multipoles (and also RESP charges that can be found in Ref. 103, Fig. 4) provide significantly less correlated energies ($R^2 = 0.7$) and large RMS errors (34.2 kcal mol^{-1}, while QTAIM multipoles present a value of

only 0.6 kcal mol^{-1}). When comparing $E_{elec}^{0,IJ}$ and $E_{elec,AMOEBA}^{IJ}$, the discrepancies are in this case of an order of magnitude higher than the previous ones, and amount to hundreds of kcal mol^{-1}, involving not only short polar contacts, but methyl C atoms too. These discrepancies, which are also found in the RESP charges, are not entirely unexpected given that (a) the AMOEBA multipoles are obtained by means of the distributed multipole analysis (DMA) [106] and (b) the RESP charges are fitted against the molecular electrostatic potential. Hence, AMOEBA/RESP rely on a very distinct charge assignment.
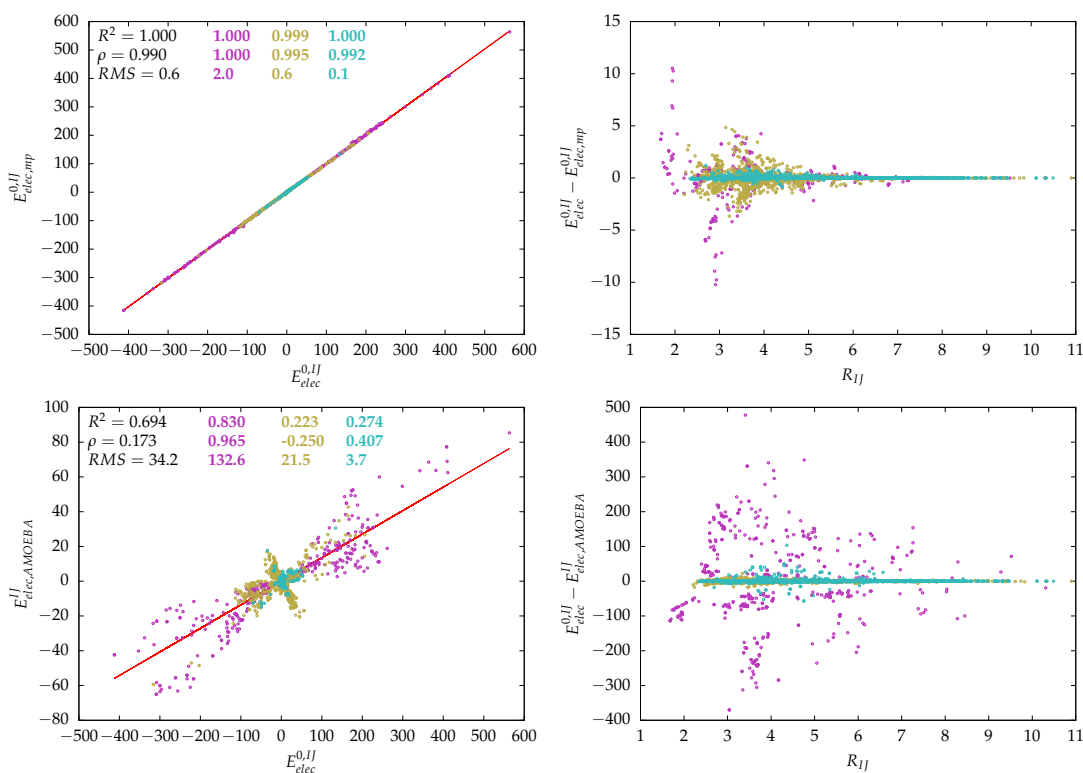
FIGURE 3.17: Comparison of the QTAIM ($E_{ele,mp}^{0,IJ}$) and AMOEBA ($E_{ele,AMOEBA}^{0,IJ}$) multipolar atomic descriptions with the IQA one ($E_{ele}^{0,IJ}$). On the left are the correlation plots and, on the right, each difference as a function of the interatomic distance (Å). The set of atomic contacts has been divided in polar (magenta), non-polar (turquoise) and mixed (yellow). All the energies are given in kcal mol^{-1}.

Charge penetration under the QTAIM scrutiny

Once the intermolecular and interatomic electrostatic interactions have been analysed, further insight into intermolecular electrostatics can be gained by inspecting the CP energy. Following the previously-derived double decomposition of ΔE_{elec}^0 (Eq. 3.7), Figure 3.18 includes the energies emanating from such decomposition for the H-bonded compounds —this is the most representative group since they show the strongest interactions and more pronounced CP energies; for further details about the rest of complexes, see Fig. 5 from Ref. 103.

As can be appreciated, $E_{elec}^{AB}(\rho_A^0, \rho_B^0)$ is stabilising in all cases but is very far from accounting for all ΔE_{elec}^0. It is the IQF penetration energy that recovers the remaining fraction of ΔE_{elec}^0 not captured by $E_{elec}^{AB}(\rho_A^0, \rho_B^0)$, which in H-bond complexes amounts
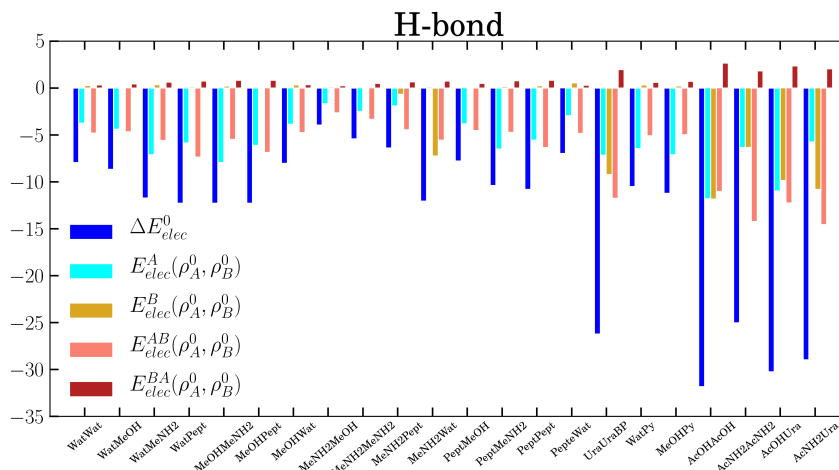
FIGURE 3.18: Decomposition of ΔE_{elec}^0 into $E_{elec}^{AB}(\rho_A^0, \rho_B^0)$ and the three
IQF penetration terms $E_{elec}^A(\rho_A^0, \rho_B^0)$, $E_{elec}^B(\rho_A^0, \rho_B^0)$ and $E_{elec}^{BA}(\rho_A^0, \rho_B^0)$ for
the H-bond group of complexes. Energies are given in kcal mol^{-1}.

to ca. 50 % of the total electrostatic formation energy. What is more, among the
three terms giving rise to $E_{elec,pen}^{IQF}$, the other intermolecular energy, $E_{elec}^{BA}(\rho_A^0, \rho_B^0)$, gives
no significant contribution at all and is even destabilising. The missing stabilising
effect is thus accounted for by the two intramolecular contributions $E_{elec}^A(\rho_A^0, \rho_B^0)$ and
$E_{elec}^B(\rho_A^0, \rho_B^0)$, representing the attraction between the tail of each density ρ_A^0 and ρ_B^0
and the nuclei of the molecule in whose domains those densities have penetrated.

Among the many formulations of CP-corrected potentials is AMOEBA+ [107]. A
combined picture of the CP energy supplied by this model and that defined for IQF
is shown in Figure 3.19, where both ΔE_{elec}^0 energies reconstructed from IQF quanti-
ties, and the total $\Delta E_{elec}^{AMOEBA+}$ obtained from the addition of the AMOEBA+ pene-
tration energies $E_{pen}^{AMOEBA+}$ (calculated with the corrections found in Ref. 107) to the
corresponding multipolar ones $E_{elec,AMOEBA}^{AB}$, match almost perfectly. This result is
nothing but the verification that both energies are the same and, consequently, the
comparison between the penetration energies defined in both models, $E_{pen}^{AMOEBA+}$
and $E_{elec,pen}^{IQF}$, is very pertinent. A close look at the evolution of the latter energies
reveals similar trends in both definitions of the penetration energy, that in H-bond
complexes, such as the H-bonded uracil dimer (Fig. 3.19 left), gradually splits when
decreasing R_{AB}. The intensity of the electrostatic effects in the dispersion complexes
is, by contrast, much less pronounced and all the energies behave very close such as
in the benzene dimer (Fig. 3.19 right). The evolution of the penetration energies is
also mirrored by the intramolecular terms $E_{elec}^A(\rho_A^0, \rho_B^0)$ and $E_{elec}^B(\rho_A^0, \rho_B^0)$, providing
further evidences of the intramolecular nature of CP energy corrections.

Concluding remarks

As has been shown, the zeroth-order IQF interfragment energies $E_{elec}^{0,AB}$ provide a
faithful description of intermolecular electrostatic interactions, the HF-D3 $E_{elec}^{AB} + D3$
energies being a good approximation to the binding energies of non-covalent com-
plexes. This fact also supports the pairwise approximation adopted by MM po-
tentials and a direct comparison between $E_{elec}^{0,AB}$ and the intermolecular electrostatic
energies supplied by the, on the one hand, QTAIM multipoles and, on the other
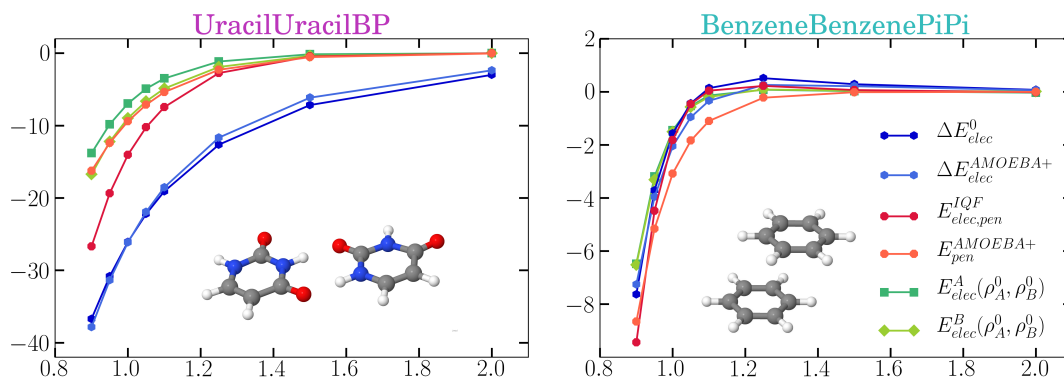
FIGURE 3.19: Comparison between the AMOEBA+ energies and the zeroth-order IQF ones for an H-bond complex (left) and a dispersion one (right) from the S66x8 dataset. Energies (Y-axis) are given in kcal mol^{-1} and distances (X-axis), relative to the equilibrium ones (R_{AB}/R_{eq}).

hand, RESP point-charge model and the multipole-based AMOEBA potential reveals a similar description of intermolecular electrostatics according to the different models. However, this does not hold at a deeper atomic level, where the different methodologies employed to assign a portion of the charge distribution to each atom evidence their discrepancies. In fact, these MM potentials are parameterised so as to reproduce global energies, as they do, but more questionable is the local atomic description they provide, in contrast to the QTAIM multipolar description, whose robustness may inspire the creation of novel and more insightful MM potentials (such as FFLUX [108, 109]).

On the other hand, IQF emerges as a powerful tool of analysis to elucidate the nature of the so-called charge penetration energy. The proposed double decomposition of zeroth-order densities and QTAIM three-dimensional partitioning leads to an insightful IQF penetration energy describing the effect that the penetration of a molecular density into the domain associated to another molecule entails. This effect has been characterised in intramolecular electrostatic terms and the analysis of the penetration energy calculated in the CP-corrected potential AMOEBA+ has pointed to the same direction. This aspect therefore warns about the appropriateness of those CP-corrected potentials mimicking ΔE_{elec} that become only partly *corrected* and forget other quantum-mechanical effects with which the CP energy mainly cancels out.

3.2   Real space elucidation of controversial chemical bonds

Since the emergence of quantum mechanics and the first proposals to approximate many-electron wave functions resorting to hydrogen-like ones, chemical imaginary has been forged on an orbital image of bonding. Needless to say, orbitals are not physical observables, and their use in setting up a wave function is not mandatory, but a choice among others. Thus, orbital-free interpretations based upon measurable objects and reference-independent arguments probably constitute a need for the understanding of the most fundamental ideas in chemistry.

   Chemical bonding is indeed one of those cornerstones of the chemical edifice for which a convincing precise definition has remained elusive. A general agreement is that a chemical bond must entail some energy lowering with respect to non-interaction, some stability gain that provokes a given pair of atoms (in the simplest case) to remain close to each other. It is with these energetic arguments where IQA can deploy its artillery and help clarify the role played by the different atoms and energy sources. A chemical bond is established upon the (strong) interaction between at least two partner atoms. But interactions exist among all the constituent particles of a system, or, after having defined atomic entities, among atoms. On some occasions, especially when different species take part, the strength of the interaction is not sufficient to be considered as a real bond and the term *non-covalent interaction* (NCI) is preferred. From the archetypal hydrogen bonds to the more recently characterised halogen bonds, NCIs have focused researchers' attention given the large variety of specific atomic interactions and their potential exploitation in supramolecular chemistry.

   Three research works constitute the second and third parts of the Chapter of this thesis devoted to Results. The first one, in Section 3.2.1, is devoted to the real space characterisation of beryllium bonding, a quite recent bonding category found between a Be-containing species and a Lewis base. Section 3.2.2 focuses on the well-established halogen bonding, for which naive interpretations in terms of the purportedly attractive electrostatics dictated by a positive-valued electrostatic potential region of a halogen and an attacking electron-rich atom still persists. A final effort with which this thesis concludes explores the charge-shift bond (subsection 3.2.3), given the extraordinary claim there exist to give it the same status in chemical bonding theory as covalent and ionic bonding. IQA as well as other quantum chemical topology techniques are therefore not only used to clarify conundrums in the literature concerning the previous NCIs, but they are also tested so as to check the outreach of these orbital-invariant methods in dissecting the different facets the above-mentioned bonding situations may present.

3.2.1   Beryllium bonding in the light of modern quantum chemical topology tools

Among the many categories of non-covalent interactions (NCI) we find beryllium bonding. Known to be very good Lewis acceptors, beryllium derivatives were proposed to form complexes with Lewis bases (LBs) [110]. The success of this idea can be recognised in the numerous works that, since then, have studied the nature of these weak bonds. Beryllium bonds have been found in both σ- [110] and π-complexes [111], and the set of theoretical methods used to assess their nature, that

comprise orbital-based EDAs such as LMOEDA, natural bond orbitals (NBO) analyses, QTAIM or electron localisation functions (ELF), among others, point to electrostatics as a major source for their stability, although non-classical contributions cannot be ignored [112].

In order to revisit this class of NCI, we explore in this work [112] beryllium bonds using different methods, all of them belonging to the QCT family. These comprise i) IQA energy decompositions so as to gain access to the prevailing energy components, ii) natural adaptive orbitals (NAdOs) to provide a one-electron image including correlation effects, and iii) electron distribution functions (EDFs) to describe the distribution of electron populations. With this aim, σ-complexes (BeX$_2$−LB systems) and π-complexes constituted by BeX$_2$ moieties in combination with ethylene and acetylene have been optimised at the B3LYP/def2-tzVPD level, followed by B3LYP/def2-qzVPD single point calculations for the subsequent analyses. Since Be is a compact atom whose interactions are not expected to be very influenced by dispersion, no such corrections have been considered. In addition, one-electron bonds such as those found in BeLi, BeNa or BeLiBe have also been considered, for which CASSCF/aug-cc-pVQZ calculations have been preferred.

Beryllium σ-complexes

The first set of complexes studied comprise BeX$_2$ units (for X=H, F and Cl) in interaction with H$_2$O, NH$_3$ and CO Lewis bases. When the very ionic BeX$_2$ species interacts with an LB, it undergoes a charge transfer (CT) process by virtue of which it acquires a higher electron population or, on the contrary, it transfers electron density to the counterpart moiety, as evidenced by the QTAIM charges in Table 3.4. The CT values there shown are generally larger than in typical H-bonded systems (e.g., 0.08 e in X$_2$Be$\cdots$CO), and some of the LBs become negatively charged, what points towards the existence of back-donation channels between the beryllium compound and the Lewis base.

Regarding the energetic analysis, both classical and *xc* components contribute similarly to the stability of the complex. When inspecting the deformation energies, it turns out that these are not overcome by the *xc* energy, a characteristic feature of very ionic bonds explained by electrostatic bonding models. However, as mentioned, E_{class}^{AB} is not capable of explaining the stability of the complexes, but requires the assistance of its E_{xc}^{AB} counterpart. In fact, covalency emerges as a very relevant contribution in view of the unusually high DIs (between $\sim 0.3 - 0.6$), that are much larger than in, for example, hydrogen bonds (HBs) —for instance, the strongly bound FHF$^-$ system presents a $\delta^{AB} \approx 0.2$. These DIs are the result of various contributions or delocalisation channels. An orbital picture for them is provided by the 2-centre NAdOs, whose occupation numbers sum up to half the DI between two centres, and correspond to the individual contributions of the 2-centre delocalisation channels to the total δ^{AB}. As a representative example, H$_2$Be$\cdots$OH$_2$ presents a $\delta^{AB} = 0.467$ that can be decomposed into four main channels accounting for a total $\delta^{AB} = 0.459$, as depicted in Figure 3.20. The dominant channel in this and the rest of the systems (see Ref. 112) is due to the σ lone pair of the LB. The purportedly second most relevant channel involving donation to the p$_{Be}$ orbital according to the NBO analysis corresponds to the fourth component in the current NAdO description, contradicting the NBO result. More relevant than the latter are the back-donation channels found, that involve the Be−X σ bonds (Be−H in this case). This back-donation capacity depends on both the Be-containing system and the LB. In

TABLE 3.4: Binding energy of the complex, deformation energies of the fragments and classical and exchange-correlation contributions to the interaction energy between the interacting species (kcal mol^{-1}). *A* refers to the beryllium-containing compound and *B* denotes the LB the former faces, whereas for the π-bonded complexes, *a* denotes a parallel alignment and *b* and orthogonal one. QTAIM charges (*e*) are also shown as well as the DI between both units (electron pairs). Only the charge of the BeX$_2$ units are given since $Q_A = -Q_B$ for guaranteeing global charge neutrality.

System	E_{bind}^{AB}	E_{def}^{A}	E_{def}^{B}	E_{class}^{AB}	E_{xc}^{AB}	Q_A	δ^{AB}
H$_2$Be$\cdots$OH$_2$	-25.31	35.27	50.45	-54.30	-56.73	0.014	0.467
H$_2$Be$\cdots$NH$_3$	-32.15	34.17	52.45	-60.89	-57.88	-0.010	0.481
H$_2$Be$\cdots$CO	-12.34	44.94	45.90	-38.16	-65.02	0.076	0.593
H$_2$Be$\cdots$C$_2$H$_4$^{a}	-15.38	52.67	51.48	-39.71	-79.82	0.080	0.755
H$_2$Be$\cdots$C$_2$H$_4$^{b}	-3.39	10.81	21.95	-7.42	-28.73	-0.030	0.311
H$_2$Be$\cdots$C$_2$H$_2$^{a}	-19.57	60.25	59.51	-49.00	-90.34	0.103	0.826
H$_2$Be$\cdots$C$_2$H$_2$^{b}	-0.62	5.32	9.48	-1.84	-13.57	-0.014	0.162
F$_2$Be$\cdots$OH$_2$	-29.75	27.35	44.99	-52.19	-49.90	-0.024	0.387
F$_2$Be$\cdots$NH$_3$	-37.79	26.83	48.83	-59.95	-53.51	-0.045	0.422
F$_2$Be$\cdots$CO	-10.83	20.90	34.61	-23.72	-42.62	-0.020	0.372
F$_2$Be$\cdots$C$_2$H$_4$^{a}	-9.62	16.84	30.94	-16.33	-41.07	-0.047	0.399
F$_2$Be$\cdots$C$_2$H$_4$^{b}	-11.44	16.21	32.67	-19.17	-41.15	-0.050	0.394
F$_2$Be$\cdots$C$_2$H$_2$^{a}	-11.87	19.55	34.17	-21.35	-44.24	-0.041	0.416
F$_2$Be$\cdots$C$_2$H$_2$^{b}	-8.11	13.54	28.48	-14.48	-35.64	-0.045	0.346
Cl$_2$Be$\cdots$OH$_2$	-32.35	39.26	55.75	-67.44	-59.93	-0.001	0.502
Cl$_2$Be$\cdots$NH$_3$	-40.56	37.21	59.54	-74.89	-62.43	-0.027	0.527
Cl$_2$Be$\cdots$CO	-13.15	39.47	47.78	-40.29	-60.11	0.029	0.563
Cl$_2$Be$\cdots$C$_2$H$_4$^{a}	-10.94	25.53	42.44	-24.96	-53.95	-0.028	0.557
Cl$_2$Be$\cdots$C$_2$H$_4$^{b}	-11.84	22.43	42.41	-25.83	-50.85	-0.037	0.521
Cl$_2$Be$\cdots$C$_2$H$_2$^{a}	-14.05	31.52	47.64	-33.89	-59.32	-0.017	0.590
Cl$_2$Be$\cdots$C$_2$H$_2$^{b}	-7.70	19.98	36.77	-20.29	-44.16	-0.035	0.457

this respect, NAdOs serve to order the LB donating and BeX$_2$ back-donating capacity, that becomes CO $\approx$ NH$_3$ > H$_2$O for the former and BeH$_2$ > BeCl$_2$ > BeF$_2$ for the latter.

π-bonded complexes

π-beryllium complexes are generally weaker than σ-bonded ones, but still their binding energies are comparable to those of HB systems (see Table 3.4). In addition to the variability in the Be species (by changing X in BeX$_2$ by H, F and Cl) in combination with different LBs (in this case, ethylene and acetylene), we have studied two distinct conformations between the two interacting species. On the one hand, BeF$_2$ and BeCl$_2$ prefer a parallel alignment (*a*) with acetylene, whereas an orthogonal arrangement (*b*) becomes more stable for ethylene acting as an LB, although both conformations remain energetically close. On the other hand, BeH$_2$ presents more pronounced differences, favouring the parallel alignment in each case.

$\delta^{AB} = 0.154$ $\delta^{AB} = 0.121$ $\delta^{AB} = 0.119$

$\delta^{AB} = 0.048$ $\delta^{AB} = 0.017$
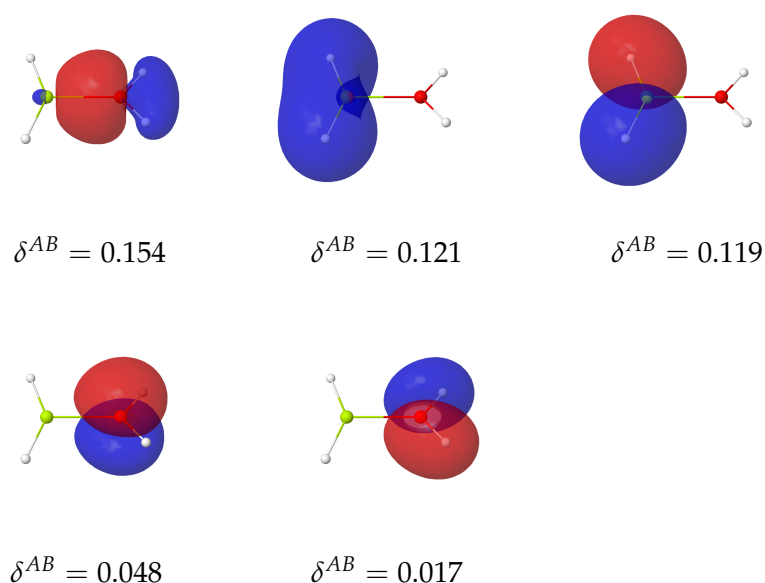
FIGURE 3.20: $H_2Be \cdots OH_2$ highest occupied NAdOs showing their corresponding contribution to the delocalisation index δ^{AB}.

CT in these π-systems is also very relevant. In general, it is the BeX_2 unit that increases its electron population, but when it comes to BeH_2 in its most stable arrangement, either the ethylene or acetylene becomes more populated. In fact, in these complexes, the BeH_2 moiety becomes more deformed than the Lewis base, a feature that is greatly inverted for the rest of the systems (that is, all but the parallel $BeH_2 \cdots LB$ π-complexes present $E_{def}(BeX_2) << E_{def}(LB)$). The characteristic high deformations of BeH_2 anticipate the important role that hydrogens play in the stabilisation, leading again to back-donation issues.

Inspection of the most relevant NAdOs in $BeH_2 \cdots LB$ π-systems (Figure 3.21) reveals that indeed back-donation, very relevant in the parallel complexes (central NAdOs in Figure 3.21), is largely hampered in the perpendicular conformation. Moreover, not only is the σ back-donation impeded, but also the π-channels from ethylene and acetylene to BeH_2 are greatly restricted. Concerning the geometries of the interacting species, as mentioned above, a larger distortion of the BeH_2 unit is related to a higher degree of electron delocalisation with the LB as reflected by the bent geometries in the first two rows of Figure 3.21 and the more linear ones in the last two.

Contrarily to BeH_2, its halogen-substituted counterparts BeF_2 and $BeCl_2$ show delocalisation channels with similar contributions to δ^{AB} in both the parallel and orthogonal conformations (see Figure 3 in the original publication 112). This can be ascribed to the higher electronic flexibility that halogen atoms provide as compared to hydrogen, as the former present more flexible electron densities than H due to higher angular momentum contributions.

Finally, returning to the energetic analysis of Table 3.4, it becomes evident (in view of E_{class}^{AB} vs. E_{xc}^{AB}) that from the energetic point of view, the π-complexes are worse described by Coulomb models alone and the non-classical xc contribution acquires a higher relevance than in the σ-complexes inspected.

$\delta^{AB} = 0.348$ $\delta^{AB} = 0.248$ $\delta^{AB} = 0.093$

$\delta^{AB} = 0.393$ $\delta^{AB} = 0.236$ $\delta^{AB} = 0.099$

$\delta^{AB} = 0.047$ $\delta^{AB} = 0.158$ $\delta^{AB} = 0.080$

$\delta^{AB} = 0.019$ $\delta^{AB} = 0.079$ $\delta^{AB} = 0.048$
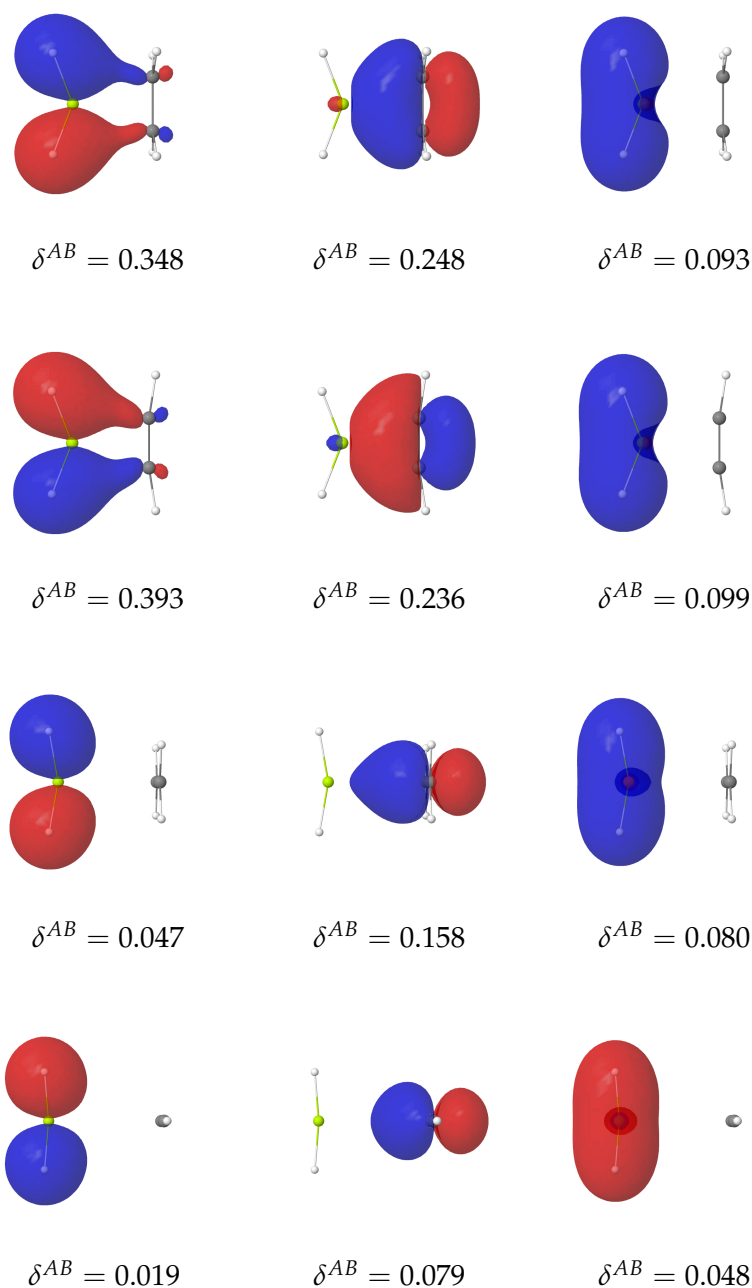
FIGURE 3.21: Highest occupied NAdOs of the $BeH_2 \cdots LB$ systems, LB being the parallel $C_2H_4{}^a$ and $C_2H_2{}^a$ in the first two rows, respectively, and the orthogonal $C_2H_4{}^b$ and $C_2H_2{}^b$ in the lower half of the Figure. These delocalisation channels account for a ~ 90 % of the total δ^{AB}.

One-electron beryllium bonds

Besides the σ and π bonding, Be has been also observed to form one-electron bonds. To study this bonding pattern, BeLi, BeNa, BeLiBe and the 1,8-BeH-disubstituted naphtalene anion have been analysed in their lowest doublet electronic state. In these systems, and according to the electronegativity differences, Be becomes negatively charged in combination with Li and Na. This translates into a progressive destabilisation of the atomic net energies with the accommodated negative charge

(see Table 3.5). In the diatomic molecules, it is seen that the classical interaction energy largely depends on the magnitude of the opposite-sign charges of the interacting atoms. Accordingly, BeLi system is characterised by $E_{class}^{AB} << E_{xc}^{AB}$, a relation that is inverted in BeNa following the lowering of $|Q|$ values of each atom. At the same time E_{xc}^{AB} are not affected and remain almost the same, also in accordance with a similar DI of about 0.5 electron pairs, thus corresponding to a one-electron bond.

TABLE 3.5: IQA and related properties of one-electron bonded LiBe, NaBe, and BeLiBe linear molecules and the $C_{10}H_6Be_2H_2^-$ anion. For the triatomic species BeLiBe (*a*), both the Be−Li and Be$\cdots$Be bonds are inspected, whereas for the poliatomic $C_{10}H_6Be_2H_2^-$ system (*b*) only the Be$\cdots$Be bond analysis is presented. E_{net}^A and E_{net}^B are given in atomic units, E_{xc}^{AB} and E_{class}^{AB}, in kcal mol^{-1}, Q_A values are in electrons and δ^{AB}, in electron pairs. *A* and *B* indices denote the first and second atom of the first column, respectively.

System	E_{net}^A	E_{net}^B	E_{xc}^{AB}	E_{class}^{AB}	Q_A	Q_B	δ^{AB}
BeLi	-14.557	-7.285	-37.47	-95.89	-0.707	0.708	0.466
BeNa	-14.594	-161.796	-35.40	-19.52	-0.330	0.330	0.550
LiBe^a	-7.286	-14.584	-19.27	-49.46	0.738	-0.369	0.223
Be$\cdots$Be^a	-14.584	-14.584	-13.74	14.17	-0.369	-0.369	0.323
Be$\cdots$Be^b	-14.114	-14.114	-38.75	177.18	1.258	1.258	0.293

The BeLiBe case is directly related to the BeLi results. As Table 3.5 shows, the sum of both E_{xc}^{LiBe} energies adds to approximately the value it presents in the corresponding diatomic molecule. In line with this, the DI of each LiBe pair is *halved* with respect to that found in the diatomic species, in accordance with a 3c-1e (three-centre one-electron) bond. A generalisation of the 2-centre delocalisation index used so far permits the study of *n*-centre bonding (recall Eq. 2.128 from Section 2.5.4). By doing so, the δ^{BeLiBe} index found presents a value of 0.186, one of the largest ever reported.

Further insight about bonding can be gained by inspecting the EDFs, the probability of a given real space resonance structure (a given integer distribution of electrons among QTAIM atoms). Table 3.6 shows that the electron distribution is concentrated in two resonance structures in the BeLi and BeNa cases, the delocalised electron being heavily polarised towards the Be atom in BeLi, whereas BeNa presents the opposite situation, albeit less pronounced. BeLiBe concentrates, in turn, the variability of distributions about three structures, highlighting the polarisation experienced towards the two terminal Be atoms.

TABLE 3.6: Electron distribution functions (EDFs) for the one-electron bonded LiBe, NaBe, and BeLiBe linear molecules. The atoms are labelled in the order in which they are written.

	BeLi			BeNa				BeLiBe	
n_A	n_B	$p(n_A, n_B)$	n_A	n_B	$p(n_A, n_B)$	n_A	n_B	n_C	$p(n_A, n_B, n_C)$
5	2	0.710	4	11	0.619	4	2	5	0.370
4	3	0.275	5	10	0.354	5	2	4	0.370
3	4	0.009	3	12	0.025	4	3	4	0.219
6	1	0.004	6	9	0.001	5	3	3	0.013
						3	3	5	0.013

FIGURE 3.22: Intramolecular delocalisation channel between the two
Be atoms in the $C_{10}H_6BeH_2^-$ naphthalene derivative anion.

The last system considered is the naphthalene derivative depicted in Figure 3.22. This anion has been reported as presenting a relatively strong one-electron intramolecular Be-Be bond [113]. According to the values gathered in Table 3.5, this homonuclear interaction is characterised by a strongly destabilising electrostatic interaction, in line with the high positive charges (1.26 e) each atom bears. The xc interaction only partially compensates this feature. To properly understand this bonding situation an IQF perspective (i.e., based on fragments) has to be adopted. Accordingly, the interaction between the two $C_{naph}BeH$ triatomic fragments reveals a total E_{int}^{AB} energy of about -62 kcal mol^{-1}, the E_{class}^{AB} contribution accounting for only -8 kcal mol^{-1}. It becomes clear that a rather important stabilisation comes from the Be-Be bond formation. A neat image of this bond is provided by a unique NAdO representing the major part of the delocalisation between both atoms, as depicted in Figure 3.22.

Concluding remarks

So far, the interactions comprising regular chemical bonds and other NCIs have succesfully been explained in the energetic terms provided by the IQA descriptors. Nevertheless, when it comes to more exotic bonds, such as Be bonding, some relevant effects are not directly captured within the IQA framework and other complementary techniques are thus required.

Without any loss of consistency, these complementary techniques can be retrieved from the QCT toolkit IQA belongs to, allowing thus for a systematic analysis of different bonding situations, no matter the strength, properties of the atoms involved or class of the interaction (e.g., intramolecular covalent bond or non-covalent interaction).

As shown for the specific case of Be bonds, IQA assisted by the real space-based natural adaptive orbitals, that complement the energetic description with an appealing one-electron picture of electron delocalisation, is able to extract the most relevant features that characterise this class of chemical bonding. Accordingly, these tools have shown that electrostatics indeed play a significant role, in line with previous

studies, but also that it is the energetic contribution due to electron delocalisation or sharing (i.e., the *covalent* contribution) that properly complements the overall bonding picture. A further analysis of this contribution reveals interesting patterns that characterise the different bonding situations. For example, the NAdOs picture provides a neat orbitalic image of electron delocalisation channels describing forward- and back-donation that can provide a great support in the analysis of chemical bonding with traditional chemical concepts.

3.2.2 Challenging the electrostatic σ-hole picture of halogen bonding: an IQA analysis

Halogen bonding (XB) has traditionally been explained in electrostatic terms, resorting to the concept of the σ-hole: a charge-depleted region that is able to interact with electron-rich moieties. According to this interpretation, electrostatics is the driving force in the formation of XBs, whose fingerprint is in the positive value of the electrostatic potential (ESP) around a halogen atom that points towards the electron-rich species with which it interacts. Albeit chemically appealing for its simplicity, such a traditional view has been challenged in the last few years by complementary theoretical techniques. As such, XB has been analysed by orbital-based methods that include natural bond orbitals (NBO), Kohn-Sham based EDAs, block-localised wave functions (BLWs) within the valence bond (VB) realm, or symmetry-adapted perturbation theory (SAPT) [114–121]. These methods have already warned about the importance of non-electrostatic contributions, such as charge transfer, polarisation or dispersion, that have their mirror image in the exchange-correlation energies of IQA. It is the latter orbital-invariant description which may provide more robust conclusions, as advanced in previous works [120, 122–130].

To help clarify the real role of electrostatics in XB as well as the limitations behind the bare ESP description, we have carried out a thorough IQA analysis in combination with other QCT techniques on a set of triatomic linear halogen anions comprising different combinations $[X\cdots Y\cdots X]^-$ of Cl, Br and I atoms [131]. This choice is the result of an attempt to capture the most relevant features of XB while maintaining simplicity. In this respect, M06-2X/x2c-TZVPPall D3-corrected calculations have been performed for $[X\cdots Y\cdots X]^-$ species with X, Y = Cl, Br; iodine-containing systems being analysed with the scalar relativistic RESC approach. The optimisation procedures have led to symmetric $[X_1-Y_2-X_3]^-$ systems, and in search for a geometry closer to the crystallographic ones, minimum energy $X_1-Y_2\cdots X_3^-$ species with frozen X_1-Y_2 distances (as those found in the isolated diatomic molecules) have been added to the set of systems analysed.

Atomic charges and interatomic interaction energies

Halogen bonds are expected to be formed with a significant contribution arising from electrostatics, at least this is what we should expect from the σ-hole model. However, a positive area on a given isosurface of the ESP of one species facing a negative one of its counterpart is not necessarily a representative picture of the interaction that takes place between both moieties. Instead of resorting to a two-dimensional distribution (i.e., that found on a surface), IQA makes use of the total three-dimensional one comprised in a given volume. Consequently, IQA provides a complete energetic picture that allows for a precise assignment of the stability of a given interaction to a particular energy term.

TABLE 3.7: Atomic charges (electrons), delocalisation indices (electron pairs) and interaction energies (in kcal mol^{-1}) for each atomic pair in the chlorine series (the rest of the systems can be found in Table 1 in Ref. 131). Negative charges are coloured in red, positive in blue and those close to zero in green.

Complex	Pair	DI	E_{int}	E_{xc}	E_{class}
$Cl_1^{-0.30}-Cl_2^{+0.04}\cdots Cl_3^{-0.74}$	1-2	1.278	-161.44	-182.49	21.05
	2-3	0.542	-66.78	-60.61	-6.17
	1-3	0.149	6.86	-5.44	12.31
$[Cl_1^{-0.50}-Cl_2^{+0.01}-Cl_3^{-0.50}]$	1-2	0.872	-104.72	-107.16	2.44
	2-3	0.872	-104.50	-107.17	2.67
	1-3	0.211	7.86	-7.65	15.51
$Cl_1^{-0.46}-Br_2^{+0.16}\cdots Cl_3^{-0.70}$	1-2	1.169	-159.84	-156.46	-3.39
	2-3	0.674	-94.12	-76.97	-17.15
	1-3	0.152	15.02	-5.44	20.46
$[Cl_1^{-0.56}-Br_2^{+0.12}-Cl_3^{-0.56}]$	1-2	0.877	-113.75	-104.49	-9.27
	2-3	0.877	-113.92	-104.55	-9.37
	1-3	0.179	13.50	-6.23	19.73
$Cl_1^{-0.61}-I_2^{+0.34}\cdots Cl_3^{-0.73}$	1-2	1.061	-174.50	-130.45	-44.04
	2-3	0.696	-115.65	-75.89	-39.76
	1-3	0.118	25.32	-4.00	29.32
$[Cl_1^{-0.65}-I_2^{+0.30}-Cl_3^{-0.65}]$	1-2	0.836	-128.55	-93.22	-35.33
	2-3	0.835	-128.07	-93.11	-34.96
	1-3	0.129	22.94	-4.23	27.17
$Br_1^{-0.22}-Cl_2^{-0.10}\cdots Br_3^{-0.69}$	1-2	1.243	-142.25	-165.85	23.60
	2-3	0.590	-58.30	-64.18	5.87
	1-3	0.192	-0.49	-6.57	6.08
$[Br_1^{-0.45}-Cl_2^{-0.10}-Br_3^{-0.45}]$	1-2	0.883	-93.22	-103.46	10.23
	2-3	0.881	-93.03	-103.24	10.20
	1-3	0.251	2.31	-8.48	10.79
$I_1^{-0.06}-Cl_2^{-0.28}\cdots I_3^{-0.66}$	1-2	1.208	-141.70	-147.38	5.69
	2-3	0.542	-37.72	-53.69	15.98
	1-3	0.24	-9.26	-7.41	-1.85
$[I_1^{-0.37}-Cl_2^{-0.26}-I_3^{-0.37}]$	1-2	0.853	-80.02	-92.05	12.03
	2-3	0.851	-79.78	-91.86	12.08
	1-3	0.317	-3.62	-9.73	6.12

Table 3.7 gathers the IQA interaction energies as well as the atomic charges and delocalisation indices between each pair of atoms in the set of chlorine-containing systems. Before discussing the relative contribution of each classical and non-classical energy to the total interaction one, there is a feature that attracts one's attention: the large charge transfer (CT) undergone by the attacking halide X_3^- populating the diatomic species X_1-Y_2, that is reflected in the charge borne by the terminal atom X_1. Indeed, the charge differences are $\Delta Q(X_3) \approx \Delta Q(X_1)$ with respect to each isolated species (see Table 3.8). Moreover, the central atom in both $[X_1-Y_2-X_3]^-$ and $X_1-Y_2\cdots X_3^-$ systems is seen to show a net charge very similar to that found in the

TABLE 3.8: IQA interaction energies (kcal mol^{-1}), atomic charges (electrons) and delocalisation indices (electron pairs) for the *in vacuo* diatomic molecules combining chlorine with itself, bromine and iodine. Negative charges are coloured in red, positive in blue and those close to zero in green.

Diatomic molecule	DI	E_{int}	E_{xc}	E_{class}
$Cl^{0.00}-Cl^{0.00}$	1.440	-170.07	-199.99	29.92
$Cl^{-0.14}-Br^{+0.14}$	1.407	-164.86	-181.86	17.01
$Cl^{-0.33}-I^{+0.33}$	1.323	-178.78	-155.43	-23.35

diatomic species, leading to a highly polarised edge-to-edge situation that is maximal in the symmetric $[X_1-Y_2-X_3]^-$ systems.

Moving to the energetic analysis, it is found that contrary to the common assumption that the covalent part of an interaction is not significantly involved in XB, IQA demonstrates that this is not true for the model systems considered and the *xc* contribution to the interaction energy is of great importance in all the studied cases. This is in line with previous assessments of the role of each interaction energy type in different σ-hole instances [120, 122–130]. A closer look at the atomic pair energies reveals that the classical contribution to the interaction between the 2-3 pairs is many times destabilising (following in general the trends marked by the magnitude and sign of the atomic charges involved). Moreover, even when E_{class} energies are favourable, the non-classical ones dominate.

The total interaction energies between a couple of atoms also depends on the distance. Accordingly, the 2-3 pair interaction becomes reinforced in the symmetric systems with respect to the asymmetric $X_1-Y_2\cdots X_3^-$ ones. Also, the 1-2 interactions are largely affected by the presence of the third, attacking atom. $|E_{xc}(X_1, Y_2)|$ decreases upon interaction, the $[X_1-Y_2-X_3]^-$ species presenting the largest *xc* destabilisations with respect to the diatomic molecules. $E_{class}(X_1, Y_2)$ energies show, in turn, a richer behaviour. In some cases, the Coulomb interaction between the two atoms weakens, but in others it strengthens upon complexation. This is due to CT that takes place and ends up increasing the differences between the opposite-sign charges the two atoms exhibit. For instance, Cl presents a charge of -0.14 e in the ClBr molecule. When the attack of a chloride leads to the $Cl_1-Br_2\cdots Cl_3^-$ complex, Br maintains roughly the same slightly positive charge, while Cl gains 0.16 extra electrons, yielding an electrostatic stabilisation of ca. 20 kcal mol^{-1}.

In general, the electrostatic interactions between halogen atoms are very dependent on their polarisable-polarisant character. Thus, the most stabilising $E_{class}(Y_2, X_3)$ is found in $Cl-I\cdots Cl^-$ and $[Cl-I-Cl]^-$ systems, followed by $Br-I\cdots Br^-$ and $[Br-I-Br]^-$ (see Table 1 in the original article [131]). The opposite situation is led by $I-Cl\cdots I^-$ and $[I-Cl-I]^-$, where the most polarisable atom is now placed at the attacking position, the least polarisable (and most polarisant) one being the atom that has to adapt its electron cloud to the incoming anion.

The last bonding descriptor is the delocalisation index. Probably, the most relevant feature is the relatively large DI for the edge atoms that, in conjunction with the previous findings about the charge distribution and relevance of the *xc* energy, is fully consistent with a purely orbital model to explain bonding in these systems: the Pimentel-Rundle model for 3c-4e (3-centre 4-electron) bonds [132, 133].

Charge transfer along the $Y_2-X_3^-$ stretching

One of the most striking results highlighted above is the large CT from the attacking X_3^- to the diatomic species with which it interacts. The resulting distribution concentrates the charge on the farthermost atom X_1, leaving the central atom almost unaltered. To check whether this observation follows a concerted CT mechanism where the central atom is not affected or it is a stepped process in which CT and polarisation are decoupled, several scans over the Y_2-X_3 distance have been performed with frozen X_1-Y_2 separation.
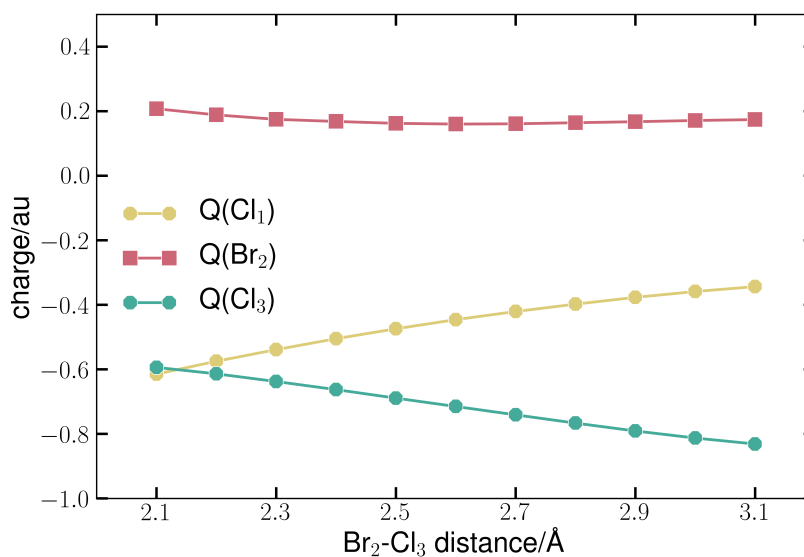
FIGURE 3.23: Evolution of the atomic charges in $Cl_1-Br_2\cdots Cl_3^-$ with the Br_2-Cl_3 distance.

In the distance range sampled in each case, the central atom is indeed not affected by the proximity of the attacking X_3^- halide, and hence displays a constant charge regardless of its separation with the X_3^- atom for distances > 2.1 Å. On the contrary, the terminal atoms present mirror evolutions. These results demonstrate that the CT seen is a one-step process involving the most external atoms, an observation that further highlights the relevance of electron delocalisation and non-classical effects in these systems.

Fragment contributions to the formation energies

The formation of a bond does not only affect the atoms directly involved; the rest of the molecule also adapts to the new situation. Therefore, it may be more insightful to focus on energies concerning the halide, on one side, and the diatomic species, on the other. To this end, IQA formation energies and their decomposition serve to address the energetics of the XB formation. By considering again the Cl series, Table 3.9 shows how penalising the deformation energies become, especially those ascribed to the attacking anion. These higher deformations can be understood by recognising the larger number of relaxation channels in the diatomic molecule that permit an easier accommodation to the final arrangement within the complex. Interestingly, X_1-Y_2 in $[I-Cl-I]^-$ (and also in $[I-Br-I]^-$, see Table 3 from Ref. 131) shows a negative E_{def}, that is, the I−Cl fragment is internally stabilised upon complexation. This

is in reality a non-surprising result given the high electron affinity of iodine, that becomes more electronically populated.

TABLE 3.9: IQA formation energies (in a very small numerical error with respect to the electronic structure calculation ones) for the chlorine-containing complexes obtained from the balance between the deformation of the fragments X_1-Y_2 and X_3, and their mutual interaction. The exchange-repulsion energies (XCR) are also included. All the energies are in kcal mol^{-1}.

Complex	E_{form}^{IQA}	$E_{def}(X_1-Y_2/X_3)$	E_{int}	E_{xc}	E_{class}	XCR
Cl−Cl···Cl⁻	-21.61	14.18 / 24.13	-59.92	-66.06	6.14	-27.75
[Cl−Cl−Cl]⁻	-28.76	14.29 / 53.59	-96.64	-114.82	18.17	-46.94
Cl−Br···Cl⁻	-32.64	14.10 / 32.35	-79.10	-82.41	3.31	-35.95
[Cl−Br−Cl]⁻	-39.28	11.38 / 49.76	-100.42	-110.79	10.37	-49.65
Cl−I···Cl⁻	-35.47	21.59 / 33.27	-90.33	-79.89	-10.44	-25.03
[Cl−I−Cl]⁻	-40.28	21.02 / 43.84	-105.13	-97.34	-7.79	-32.49
Br−Cl···Br⁻	-22.14	10.67 / 25.99	-58.80	-70.75	11.95	-34.09
[Br−Cl−Br]⁻	-29.01	6.75 / 54.96	-90.72	-111.71	20.99	-47.00
I−Cl···I⁻	-16.53	3.93 / 26.53	-46.99	-61.05	14.06	-30.59
[I−Cl−I]⁻	-21.56	-0.09 / 61.93	-83.40	-101.60	18.20	-39.76

The formation energies are in all cases favourable, and this is the result of the highly stabilising interactions. Among the two components of E_{int}, the electrostatic energy E_{class} is seen to correlate with E_{form}. However, and again modelled by the polarisant-polarisable character of the atoms directly involved in the interaction via the σ-hole, E_{class} is in general destabilising, presenting negative values only for the *extreme σ-hole bearing* Cl−I···Cl⁻ and [Cl−I−Cl]⁻ systems. As can be seen, although E_{class} may contain valuable information about the interacting species, it is not enough to describe XB in the cases studied herein (it becomes even unfavourable). E_{xc}, in turn, are the dominant energetic contribution and the one that mostly explains bonding in these cases (contrary to the frequently positive E_{class} values, E_{xc} is always negative and very strong). These findings were also pointed out by Wolters and Bickelhaupt [134], who reported stabilising electrostatics for the trihalides here considered. This is a not surprising outcome from an orbital EDA methodology, that computes the energy associated to interpenetrating electron densities, thus incorporating further stabilising effects coming from the interpenetrating electrons of one species and the nuclei from the other they find.

The final energetic descriptor, the exchange-correlation-repulsion energy (XCR), results from the sum of the deformation energies and the xc ones and is usually linked to Pauli repulsion in the long range. The strongly negative character of this energy points out the non-perturbative regime of the interactions and the relevant role that orbitalic (or xc in IQA parlance) interactions play, that overcome (and to a large extent) the deformation energies, thus signaling that equivalently, Pauli repulsion is overcome by the effect of electron delocalisation. This last descriptor remarks the conclusion that has been emerging throughout the whole analysis, that is, halogen bonding as exhibited by the [X···Y···X]⁻ model systems cannot be described by an electrostatic model alone. By contrast, it is the effect of electron sharing and delocalisation that accounts for the most part of bonding, as evidenced by numerous descriptors such as E_{xc} energies, the charge distribution or the anomalously high DIs

between the edge atoms, that point towards a 3c-4e orbital model as the most fruitful description.

3.2.3 Electron-pair bonding in real space. Is the charge-shift family supported?

The nature of a (bicentric) chemical bond is normally regarded as neither purely ionic nor purely covalent, but rather as a mixture of the two extreme cases. Thus, it is customary to speak about bonds as having a certain degree of covalency and a remaining one of ionicity —a range in which the so-called polar covalent bonds and other possible classifications find their way to be characterised. Notwithstanding, the existence of these two families to cover the different bond situations between pairs of atoms has been challenged in the last years with the appearance of a purportedly new category that should not be classified as any of the former two: *charge-shift* bonding (CSB) [135, 136].

The original proposal was conceived within non-orthogonal valence bond theory (NOVB) after realising that neither the ionic nor the spin-paired (i.e., covalent) valence bond (VB) structures (Ψ_{ion} and Ψ_{cov}) were able to explain bonding in molecules such as F_2, H_2O_2 or H_3SiF. Rather, it is the resonance of the two (as a linear combination giving rise to Ψ_{VB}) that provides most of the stabilising energy that enables bonding. The stabilisation this mixture leads to is called CS resonance energy (RE), computed as the remaining part of the bond energy that is not accounted for by the principal VB structure (either one of the possible Ψ_{ion} or Ψ_{cov}). This RE is associated to the shift of only one electron between the constituent fragments, hence its name charge-shift resonance energy that leads to CSB when this is the major energetic contribution to bonding. A qualitative representation of the VB description of F_2, the prototypical example of a CSB molecule, can be found in Figure 3.24.
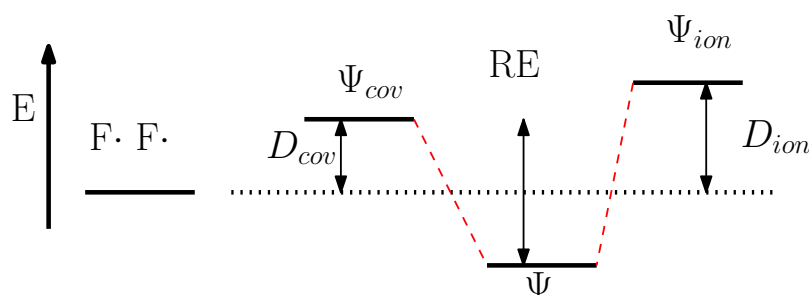
FIGURE 3.24: Schematic representation of the F_2 energy diagram. On the left is the energy of the separate F atoms, followed by the VB covalent structure Ψ_{cov} that, in the case of F_2, is the closest one to the true wave function Ψ, although its associated energy lies above that of the separate atoms. The ionic structure is even less stabilised, being impossible to describe a bound state between the two fluorines with only one resonance structure. It is however the mixing of the two that provides the bonded situation with an associated energy lower than that of the independent atoms. D_{cov} and D_{ion} are the energy differences between the either *cov* or *ion* VB structures and the non-interacting atoms, whereas RE denotes the resonance energy measured from the final VB wave function to the covalent one. The relative position of each energy level is qualitative.

However striking this evidence might be, resonance is representation-dependent, that is, a change in the basis for constructing Ψ_{ion} and Ψ_{cov} wave functions changes the RE value. Since the existence of a third distinct grand category of chemical bonding must be followed by characteristic signatures regardless of the theoretical (or experimental) method used to study it, efforts for proving the distinct nature of CSB have been made resorting, for example, to two-configuration self-consistent field in MO theory, or the Laplacian of the electron density within QTAIM.

In order to examine whether this new category of chemical bonding can be considered as such, we have analysed several CSB-featuring systems [137], although, for the sake of simplicity, only the most representative one shall be considered in this summarised Section. With the aim of providing arguments on a simple basis but without loss of reliability, an easy-to-interpret two-electron, two-orbital model has been the choice for carrying out this work. Accordingly, CAS(2,2)/6-31G* calculations have been performed for both difluorine and dihydrogen molecules for comparative purporses (the rest of the systems calculated and the arguments behind the consideration of this simple model can be found in Ref. 137). The model system wave functions are constructed from two atomic orbitals a and b ($2p_z$ in the case of F_2 and $1s$ in H_2).

The MOs (labelled by symmetry) therefore correspond to $g, u = N_{g,u}(a \pm b)$, $N_{g,u}$ being the normalisation constant of either g or u, leading to a CASSCF wave function of the form $\Psi_{CAS} = \lambda|g\bar{g}| + \mu|u\bar{u}|$. In NOVB, the Heitler-London structures are, in turn, defined by $\Psi_{cov} = N_{cov}(|a\bar{b}| + |b\bar{a}|)$ and $\Psi_{ion} = N_{ion}(|a\bar{a}| + |b\bar{b}|)$, so that the NOVB wave function becomes $\Psi_{VB} = c\Psi_{cov} + i\Psi_{ion}$. For H_2, Ψ_{cov} provides a fairly accurate potential energy curve, whereas in the F_2 case, neither Ψ_{cov}, Ψ_{ion} nor Ψ_{HF} describe bound states. This occurs only when moving to Ψ_{VB} and Ψ_{CAS}. The use of this simple model permits a one-to-one correspondence $(c, i) \leftrightarrow (\lambda, \mu)$, being possible to travel through Ψ_{cov}, Ψ_{ion}, $\Psi_{HF} \equiv |g\bar{g}|$ and $\Psi_{CAS} \equiv \Psi_{VB}$ when orbitals are not allowed to relax and the wave function $\Psi = c\Psi_{cov} + i\Psi_{ion} = \lambda|g\bar{g}| + \mu|u\bar{u}|$. Moreover, easy-to-interpret expressions for the 1- and 2-RDM are achieved (see Ref. 137), showing that the $|g\bar{g}|$-$|u\bar{u}|$ resonance (g-u mixing) lies only on the two-electron term of the 2-RDM, pointing towards e-e repulsion as the driving force of CSB, in opposition to the role usually ascribed to the kinetic energy.

Real space signatures

In real space, various arguments have been adduced for signaling a distinct behaviour of CSB. On the one hand, deformation densities $\Delta\rho$ show a density depletion instead of density accumulation in the bonding region. On the other hand, Laplacian values at the BCP, that are usually negative for homonuclear covalent bonds, are positive instead. Both arguments, while true, do not necessarily point to a new bonding category. $\Delta\rho$ is, by definition, reference dependent, so a change from the spherically-averaged atoms to other valence-prepared ones may cause a change of sign in the internuclear region. Let us now focus on the $\nabla^2\rho$ function. For the simplest (neutral) molecular system, H_2, each atomic density takes the form $\rho(r) = Ne^{-\zeta r}$, so that the Laplacian becomes $\nabla^2\rho(r) = Ne^{-\zeta r}(\zeta^2 - 2\zeta/r)$. Such an analytical expression anticipates an interesting property: that, after a critical distance, even dihydrogen's $\nabla^2\rho$ becomes positive, a condition met at 3.2 a_0 of internuclear separation. Thus, a positive $\nabla^2\rho(r_{BCP})$ may be just an indication of a longer than usual interatomic distance. Moreover, concerning the orbitals, u has a node at the central interatomic point in homonuclear systems (i.e., the BCP), so a large mixing of states $|g\bar{g}|$ and $|u\bar{u}|$ makes the Laplacian tend to positive values, no matter the distance when $\mu \to 1$. Of

TABLE 3.10: Electron density and Laplacian of the electron density at
the BCP obtained from different MOs of F_2. All values in au.

Ψ	$\rho(\mathbf{r}_{BCP})$	$\nabla^2\rho(\mathbf{r}_{BCP})$
$\lvert g\bar{g}\rvert$	0.22	0.46
$\lvert u\bar{u}\rvert$	0.06	2.68
CAS		
18 e	0.21	0.66
16 e	0.06	0.90
2 e	0.15	-0.23

course, these remarks must not be surprising since the previous descriptors derive from the 1-RDM, and it is the second-order one that incorporates resonance (i.e., *g-u* mixing), as commented before.

Concerning real calculations, Table 3.10 shows the $\rho(\mathbf{r}_{BCP})$ and $\nabla^2\rho(\mathbf{r}_{BCP})$ values for the F_2 molecule. As can be seen, the electron density at the BCP is dominated by the 2 e active subsystem, but $\nabla^2\rho(\mathbf{r}_{BCP})$ is not. Rather, the Laplacian presents a positive value at this point that is due to the participation of the other non-bonding electrons; the 2 e subsystem even provides a negative $\nabla^2\rho$, as in the H_2 molecule.

Energetic evidence

Paying now attention to the energetic factors, CSB is characterised by a large RE, as mentioned above, what has been linked to the mechanism behind bond formation. Thus, the contraction experienced by orbitals during typical covalent bond formation leads to an increase in the kinetic energy T, which is not entirely compensated by the lowering in the potential energy V. Hence, an additional mechanism must take place to restore the virial ratio $V/T = -2$ at equilibrium, and here is where the RE comes into play. This is a common feature, but it is in CSB-exhibiting systems where the RE becomes crucial, especially when the additional effect of e-e repulsion between bonding electrons and lone pairs (the so-called lone pair bond weakening effect, LPBWE) has to be overcome. In the VB view, this is accomplished through the mixing of the *cov-ion* configurations, what ultimately results in a T reduction. In a fixed-orbital framework, however, T is a one-electron property and is thus not affected by *g-u* (or, equivalently, *cov-ion* mixing).

Table 3.11 collects the energy components of F_2 and H_2 according to different wave function descriptions. As can be appreciated, in the F_2 case, HF ($\lvert g\bar{g}\rvert$) displays the smallest T, as it corresponds to symmetric single determinants but its density is not compact enough, so that the total one-electron energy ($h = T + V_{ne}$) is higher than in the $\lvert u\bar{u}\rvert$ state, contrarily to what happens in dihydrogen. In the VB description, the covalent structure turns to have a very large T that, in conjunction with V_{ee}, leads to an energy lowering when mixing with Ψ_{ion}. By separating V_{ee} into core-core (c-c), core-active (c-a) and active-active (a-a) e-e interactions it is seen that c-a dominates the total V_{ee} in the covalent structure, with specially large values in the *ion* and $\lvert u\bar{u}\rvert$ states (for further details, consult Ref. 137). In the absence of lone pairs (such as in H_2), no c-a interactions are present, and the one-electron energy decreases from $\lvert u\bar{u}\rvert$ to $\lvert g\bar{g}\rvert$, with an additional stabilisation gained by a lower V_{ee} when mixing both states (or, analogously, Ψ_{cov} and Ψ_{ion}). The presence of lone pairs leads to the contrary behaviour, namely, a one-electron stabilisation towards $\lvert u\bar{u}\rvert$ and a large mixing that seeks to decrease V_{ee}, for which the c-a interaction is dominant.

TABLE 3.11: Kinetic and potential energies of F_2 (left) and H_2 (right) for different wave functions. h represents the one-electron energy $(T + V_{ne})$. All data in au. In the F_2 case, 533 au have been added to V_{ne} and 198 and 108 au have been subtracted to T and V_{ee}, respectively.

Ψ	T^{F_2}	$V_{ne}^{F_2}$	h^{F_2}	$V_{ee}^{F_2}$	T^{H_2}	$V_{ne}^{H_2}$	h^{H_2}	$V_{ee}^{H_2}$
cov	1.229	-2.078	-0.849	0.482	1.202	-3.367	-2.165	0.594
ion	1.229	-2.078	-0.849	1.048	1.202	-3.367	-2.165	0.721
$|g\bar{g}|$	-0.034	-0.373	-0.407	0.025	1.133	-3.622	-2.489	0.655
$|u\bar{u}|$	3.475	-5.110	-1.635	2.083	3.050	-4.036	-0.986	0.697
CAS	0.290	-0.811	-0.521	0.044	1.156	-3.619	-2.463	0.619

Therefore, it is not the kinetic energy that drives the F_2 bond formation —arguments based on the large value in Ψ_{cov} do not even consider the also large T found in the *ion* counterpart—, but the reduction of the extremely large V_{ee} energy that results in a lengthening of the F-F distances.

Intra- and interatomic considerations

The previous analyses have made use of global energetic terms. A more insightful study can, however, be carried out by means of the IQA energy decomposition. Table 3.12 shows some relevant IQA terms along with the DIs of the model H_2 and F_2 molecules in which we have focused so far. Since both are homodiatomic molecules, the corresponding atomic basins coincide, and they thus show similar DIs and covalent xc energies. As expected, H deformations are larger as H acquires a higher hydride character, thus growing from *cov* to *ion*, where $E_{def}(H_2)$ is maximum. In H_2, the optimal CAS (or optimal NOVB wave function) is close to *cov*, and the deformations are mainly compensated by the covalent energy term E_{xc}. The F_2 molecule presents much larger atomic E_{def} energies than dihydrogen (in the case of the *cov* structure its E_{def} is about 20 times larger), but a lower stabilisation gain by interaction (due to the very long internuclear distance and the compactness of F_2 electron distribution). Mixing increases the favourable effect of interaction and, at the same time, decreases deformation owing to a considerable lowering in the intraatomic e-e repulsion (V_{ee}^{AA}) from the *cov* configuration. In fact, the unusually large F-F distance is a consequence of this intraatomic effect, which grows faster when approaching the F atoms (see Ref. 137). Apart from this characteristic feature of F_2 bonding, the xc energies display the standard behaviour. Therefore, the abnormalities found in F_2 can be linked with the intraatomic electron repulsion that, through a lengthening of the internuclear separation, seeks to be minimised. This fact can be ascribed to the compactness of F and the LPBWE, that seems to play a major role in leading F_2 bonding to its actual situation.

Concerning the electron distribution in space, it is found that only one electron pair is delocalised in F_2 and thus the non-negligible EDFs are $p(9,9)$ and $p(10,8) = p(8,10)$. The CAS solution shows no abnormal fluctuation in F_2, with a $p(10,8)$ even smaller than its equivalent in H_2. This is compatible with a proto-bond with hindered delocalisation. Therefore, in contrast to one of the main arguments used to defend the CSB as a bonding category different from the ionic and covalent ones, orbital-invariant analyses show an even smaller electron-pair fluctuation than in normal covalent bonds.

TABLE 3.12: Relevant IQA energy terms for different two-state wavefunctions of H_2 and F_2, as well as the delocalisation indices. All the data are given in au.

	E_{def}^A	T^A	V_{ee}^{AA}	E_{int}^{AB}	E_{xc}^{AB}	V_{ee}^{AB}	δ^{AB}
H_2							
cov	0.005	0.601	0.136	-0.153	-0.194	0.321	0.715
ion	0.132	0.601	0.263	-0.280	-0.321	0.195	1.284
HF	0.043	0.567	0.198	-0.230	-0.262	0.269	1.000
CAS	0.013	0.578	0.165	-0.189	-0.222	0.296	0.833
F_2							
cov	0.092	99.614	40.435	-0.084	-0.122	27.611	0.397
ion	0.526	99.614	40.870	-0.387	-0.425	27.381	2.040
HF	0.173	98.983	40.270	-0.263	-0.297	27.484	1.206
CAS	0.077	99.145	40.237	-0.164	-0.199	27.569	0.713

Concluding remarks

With all the previous analyses it has been shown that the original arguments given according to NOVB theory results cannot translate to the real space elucidation carried out in this work. The specific bonding features exhibited by F_2 (and other molecules) can be well described in terms of the large LPBWE experienced that leads to proto-bonding, but whose real space signatures do not differ significantly from typical covalently-bonded systems. This topic, however, still generates controversy [138], so wider and more profound analyses will be required to clarify the scope of the CSB category.

Chapter 4

Conclusiones

La presente tesis doctoral pretende establecer nuevas vías de aplicación del método de átomos cuánticos interaccionantes (IQA) a sistemas de interés biológico, al mismo tiempo que continúa mostrándose útil a la hora de esclarecer la naturaleza del enlace químico en situaciones complejas.

La primera parte del trabajo expuesto en esta tesis (Capítulo 3.1) se ha centrado en la extensión de IQA a la modelización biomolecular, para la que, de manera exitosa, se han dado los primeros pasos. En primer lugar, y en previsión de una futura necesidad de incorporar efectos de correlación a sistemas de mediano y gran tamaño, se ha llevado a cabo un estudio de la idoneidad de diferentes funcionales de DFT en la descripción de los enlaces de hidrógeno (HB) en pequeños cúmulos de agua y de los efectos de cooperatividad y anticooperatividad asociados a estos. Esta ubicua interacción no covalente (NCI), tan presente en sistemas de interés biológico, resulta bien descrita por la mayoría de los funcionales estudiados, si bien algunos proporcionan tendencias más alejadas de las referencias. Asimismo se demuestra que la técnica de escalado empleada para las energías de xc DFT en IQA proporciona términos energéticos consistentes.

A continuación se ha examinado la eficacia de la metodología combinada IQA-D3 en el estudio de los efectos conformacionales inducidos por el flúor. Para ello se ha partido de cálculos HF-D3, que permiten el tratamiento completamente separado de dispersión e intercambio. Los resultados muestran la idoneidad de dicha metodología para conocer la importancia relativa de efectos electrostáticos, estereoelectrónicos o de dispersión en la determinación de las preferencias conformacionales de los compuestos organofluorados, incluso en presencia de diversos grupos funcionales. La versión de fragmentos de IQA, denominada IQF, agrupa los resultados atómicos en términos moleculares con un significado químico claro, posibilitando así un análisis más sencillo e intuitivo de las energías conformacionales en los sistemas de interés.

En sistemas de interés biológico son especialmente relevantes los efectos del entorno químico sobre la molécula o el fragmento de estudio. Como primer paso hacia un tratamiento más general, se han incorporado por primera vez los efectos electroestáticos del disolvente en la partición IQA. Ello se ha llevado a cabo conforme al modelo de disolvente continuo COSMO, que describe la componente electrostática de la solvatación por medio de un conjunto de cargas puntuales localizadas en la superficie molecular del soluto . Dicho tratamiento permite una descomposición directa en IQA de las energías de interacción de cada átomo del soluto con el disolvente, así como la partición análoga de las energías libres de solvatación en contribuciones efectivas tanto atómicas ΔG_{solv}^{A} como de fragmento ΔG_{solv}^{G}. Los resultados del análisis muestran importantes relaciones entre la carga o entorno químico de los fragmentos y el valor de ΔG_{solv}^{G}, demostrándose así la viabilidad de aplicar la descomposición IQA a los efectos del disolvente, así como la riqueza de la información

que proporciona.

Como colofón a la vertiente bioquímica de esta tesis se encuentra un exhaustivo análisis de las interacciones electrostáticas intermoleculares. En complejos no covalentes, la presencia de átomos polares que inducen una distribución no homogénea de la carga en la molécula otorgan a la electrostática un papel dominante en numerosas situaciones, papel que se ve complementado por la dispersión entre las nubes electrónicas interaccionantes. Es por ello que la formación de estos complejos se ve muy bien descrita por los términos IQF-D3 $E_{elec}^{AB} + D3$. Más aún, la aproximación de orden cero que surge de considerar las densidades moleculares sin relajar se muestra como una buena alternativa al cálculo QM del complejo completo. En modelización biomolecular, sin embargo, las densidades continuas de carga se ven reemplazadas por conjuntos de cargas puntuales o bien multipolos que simplifican los cálculos de las interacciones entre estas. La comparación entre las energías intermoleculares IQF y las obtenidas a partir de cargas puntuales RESP y multipolos AMOEBA demuestra que los términos $E_{elec}^{0,AB}$ siguen tendencias muy similares a las de los campos de fuerzas. Por su parte, los multipolos QTAIM asociados a la partición IQA resultan en energías muy similares a las IQF, tendencia que se mantiene cuando se evalúan los términos atómicos individualmente y que no es, en este caso, reproducida por los campos de fuerzas testeados. Finalmente, la doble partición del espacio real y de las densidades moleculares permite evaluar conforme a las energías IQF la llamada energía de penetración de carga. Dicha energía resulta de la diferencia entre la calculada entre densidades continuas que se interpenetran mutuamente y aquella obtenida por medio de la expansión multipolar. El análisis efectuado permite asignarle principalmente una naturaleza intramolecular, lo que se puede traducir en energías descompensadas si dicha corrección es incorporada en un campo de fuerzas sin tener en cuenta otros factores con los que tiende a cancelarse, como los de *xc* intramoleculares.

La segunda línea de trabajo de esta tesis (Capítulo 3.2) comprende el uso de IQA en combinación con otras metodologías de la topología químico cuántica (QCT) para desentrañar la naturaleza del enlace químico en diversos casos de interés. Uno de estos es el de los enlaces de berilio. Este tipo de NCI se da entre átomos de Be, que actúan como ácidos de Lewis, y especies dadoras de electrones, que actúan como bases. Por ello, la racionalización de este tipo de interacción se ha venido dando fundamentalmente en términos electrostáticos, aunque, poco a poco, los nuevos estudios que han ido surgiendo han apuntado hacia una nada despreciable contribución covalente (o de *xc*). El estudio aquí realizado apunta en la misma dirección: la electrostática es fundamental, si bien la deslocalización electrónica es también muy relevante. Por su parte, el uso de NAdOs para interpretar los canales de dicha deslocalización permite observar efectos no apreciables por otras técnicas, como la retrodonación del BeH_2 hacia la base correspondiente, todo ello amparándose en argumentos que no dependen del conjunto de orbitales empleado para construir la función de onda.

También en términos electrostáticos se interpretan comúnmente los enlaces de halógeno (XB), dados entre un átomo de halógeno y una base de Lewis. Un elemento característico de estas interacciones es la presencia del agujero σ, una zona positiva del potencial electrostático en torno al halógeno en la dirección de interacción con la base de Lewis. Dicho agujero σ, no obstante, no garantiza una preponderancia de la interacción de Coulomb, ni tan siguiera que esta sea favorable. Las energías IQA, por contra, consideran la distribución de carga completa en una región espacial y su análisis en sistemas $[X{-}Y{-}X]^-$ y $X{-}Y{\cdots}X^-$ con distintas combinaciones de Cl, Br y I sugiere que el papel de la electrostática es relativo y la formación de estos sistemas

no se puede explicar sin la contribución asociada a la deslocalización electrónica o covalencia. De hecho, aunque sean sistemas modelo de XB, el enlace queda bien descrito por un modelo clásico de 3c-4e que únicamente descansa en argumentos de interacciones entre orbitales o, en lenguaje del espacio real, de deslocalización electrónica.

Como último caso de enlace controvertido se ha analizado el llamado enlace por desplazamiento de carga (CSB). Desde su caracterización en NOVB, se había argumentado que no se trataba de una nueva subclase de enlace más, sino que era muy diferente a los demás y, por tanto, merecía ser dotado de un estatus similar al de las dos grandes familias de enlace químico: iónico y covalente. La razón principal era que enlaces como el del F_2 no se pueden describir con las estructuras covalente o iónicas de VB, sino que es necesaria su mezcla para alcanzar un estado enlazante, dando lugar a una elevada energía de resonancia entre dichas estructuras. La QTC, sin embargo, que es invariante orbital y, por tanto, no se ve afectada por si el método de cálculo de la estructura electrónica es de VB o de MO, no muestra ninguna característica anómala en dicho enlace, más allá de su longitud inusualmente larga debido a la repulsión de los pares de electrones no enlazantes. Tanto IQA, como EDF o el análisis de $\nabla^2\rho$ no muestran resultados anómalos que no puedan ser explicados en los anteriores términos. Una nueva categoría de enlace químico debería ser visible ante los distintos descriptores que se emplean para caracterizar el enlace químico; sin embargo, en los sistemas estudiados, dichas características propias del CSB no se consiguen encontrar fuera de VB.

Las investigaciones aquí descritas comportan avances en la clarificación de los diversos problemas planteados, pero, de manera más importante, se aúnan dentro de un proyecto mayor: el desarrollo y la popularización del método de átomos cuánticos interaccionantes. Con todo lo aquí expuesto se ha pretendido contribuir a mostrar que esta técnica, una de las tantas que los químicos computacionales tienen a su disposición para descomponer las energías y dotarlas de diversos significados físicos, posee un conjunto de ventajas procedentes de su robustez y sus grandes capacidades explicativas y predictivas.

Chapter 5

Published articles

The full set of articles published in the course of this thesis is:

- Jiménez-Grávalos, F.; Casals-Sainz, J.L.; Francisco, E.; Rocha-Rinza, T; Martín Pendás, Á; Guevara-Vela, J. M. **DFT performance in the IQA energy partition of small water clusters**. *Theor Chem Acc* **2020**, 139:5.

- Díaz, N; Jiménez-Grávalos, F; Suárez, D; Francisco, E.; Martín-Pendás, Á. **Fluorine conformational effects characterized by energy decomposition analysis**. *Phys. Chem. Chem. Phys.* **2019**, 21, 25258–25275.

- Jiménez-Grávalos, F.; Díaz, N.; Francisco, E.; Martín-Pendás, Á; Suárez, D. **Interacting Quantum Atoms Approach and Electrostatic Solvation Energy: Assessing Atomic and Group Solvation Contributions**. *ChemPhysChem* **2018**, 19, 3425.

- Jiménez-Grávalos, F.; Suárez, D. **A Quantum Chemical Topology Picture of Intermolecular Electrostatic Interactions and Charge Penetration Energy**. *J. Chem. Theory Comput.* 2021, 17, 8, 4981-4995.

- Casals-Sainz, J. L.; Jiménez-Grávalos, F.; Costales, A.; Francisco, E.; Martín Pendás, Á. **Beryllium Bonding in the Light of Modern Quantum Chemical Topology Tools** *J. Phys. Chem. A* **2018**, 122, 3, 849-858.

- Jiménez-Grávalos, F.; Gallegos, M.; Martín Pendás, Á; Novikov, A.S. **Challenging the electrostatic σ-hole picture of halogen bonding using minimal models and the interacting quantum atoms approach**. *J Comput Chem.* **2021**, 42, 676–687.

- Casals-Sainz, J. L.; Jiménez-Grávalos, F.; Francisco, E.; Martín Pendás, Á. **Electron-pair bonding in real space. Is the charge-shift family supported?** *Chem. Commun.* **2019**, 55, 5071-5074.

DFT performance in the IQA energy partition of small water clusters

Fernando Jiménez-Grávalos[a], José Luis Casals-Sainz[a], Evelio Francisco[a], Tomás Rocha-Rinza[b],
Ángel Martín Pendás[a], José Manuel Guevara-Vela[a,*]

[a]Departament of Analytical and Physical Chemistry, University of Oviedo, E-33006, Oviedo, Spain.
[b]Institute of Chemistry, National Autonomous University of Mexico, Circuito Exterior, Ciudad Universitaria,
Delegación Coyoacán C.P. 04510, Mexico City, Mexico.

Abstract

This paper addresses an assessment of the performance of a large set of exchange-correlation functionals in the description of hydrogen bonding within the Interacting Quantum Atoms (IQA) energy partition. Specifically, we performed IQA analyses over a series of small water clusters $(H_2O)_n$ with $n \leq 6$. Apart from LDA-like approximations, all the considered families of exchange-correlation functionals (GGA, meta-GGA, and hybrid) reproduce the trends associated with hydrogen-bond non-additive effects computed with reference Møller-Plesset and coupled cluster wave functions. In other words, the IQA energy partition together with most of the functionals addressed herein produce good results concerning the study of non-additivity in hydrogen bonds at a reduced cost as compared with correlated wave functions approximations. This conditions might be further exploited in the examination of larger hydrogen-bonded complexes.

Keywords:
Quantum theory of atoms in molecules, Interacting quantum atoms, Density functional theory

Introduction

Density Functional Theory (DFT) is regarded as one of the most fruitful theories in computational chemistry. [1] Its widespread use arises from its convenient compromise between accuracy and computational cost. [2] DFT has been applied to a wide range of problems, from the study of small chemical systems to the analysis of large macromolecular aggregates. [3] Indeed, the fields of application of DFT increases continuously: it is utilised in the understanding of catalytic processes, the discovery of new drugs, and the development of new materials. [4, 5]

---

*To whom correspondence should be addressed: jmguevarav@gmail.com

Besides these applications, DFT also provides the basis for the understanding and exploitation of concepts such as electronegativity and chemical hardness which are very useful in many branches of chemistry. [6] The development of these concepts in the field of "conceptual DFT" has provied a series of tools for the interpretation of molecular properties, in particular, those concerned with chemical reactivity. [7, 8]

Likewise to conceptual DFT, the topological analysis of the electron density in accordance with the Quantum Theory of Atoms in Molecules (QTAIM) has resulted in new insights about molecules and molecular clusters as well as the processes undergone by these systems. [9, 10] The charge distribution is not the only scalar field that can be examined in this way. Many other functions, e.g. the electron localisation function, the source function or the virial field have been studied with similar topological approaches than that employed in the QTAIM. [11–13] The investigation concerning the properties of these scalar fields and its applications has resulted in the emergence of the field of theoretical chemistry known as Quantum Chemical Topology (QCT). [14]

The Interacting Quantum Atoms (IQA) [15, 16] energy partition is a notable method for wave function analyses. This approach allows for the separation of the total electronic energy of a system into physically-meaningful intra- and inter-atomic contributions. Initially, IQA could only be applied to small systems due to the high computational burden required to perform this type of partition of the electronic energy. Moreover, the set of available electronic structure methods which could be coupled to IQA was limited to HF, CASSCF, CISD, FCI or CCSD. Popelier and coworkers have however recently used IQA with the B3LYP functional, [17] and some of us have extended this energy partition to the rest of the most common exchange-correlation functionals. [18] The coupling of IQA and DFT reduces dramatically the computational cost when an effective treatment of electron correlation is needed and, therefore, it opens new possibilities for the applicability of the IQA analysis.

One of such opportunities is the study of the non-additivity in hydrogen bonding in water clusters. These adducts are prototypical systems for the rationalisation of cooperative and anticooperative effects in hydrogen bonding. For example, we have used IQA coupled with Hartree-Fock and correlated wavefunctions to study cooperativity and anticooperativity in $(H_2O)_n$ [19, 20] clusters with n $\leq$ 6. These investigations lead to a hydrogen bond strength hierarchy within water clusters in terms of the single and double character of the hydrogen bonding donor and acceptor H_2O monomers. We present herein an assessment of the perfor-

mance of sixteen different functionals in the description of the non-additivity of hydrogen bond in the small water clusters shown in Figure (1) with the IQA energy partition. Hereof, we note that DFT describes hydrogen-bonded systems appropriately. On this subject, Michaelides *et al.* [21] concluded that the general DFT description of water is acceptable and that hybrid functionals performed the best in accounting for monomer properties and the non-additivity of hydrogen bonding. However, Mendedev *et al.* showed that for many new hybrid functionals the corresponding densities deviate from the exact ones because they sacrifice physical rigour for the flexibility of empirical fitting. [22]
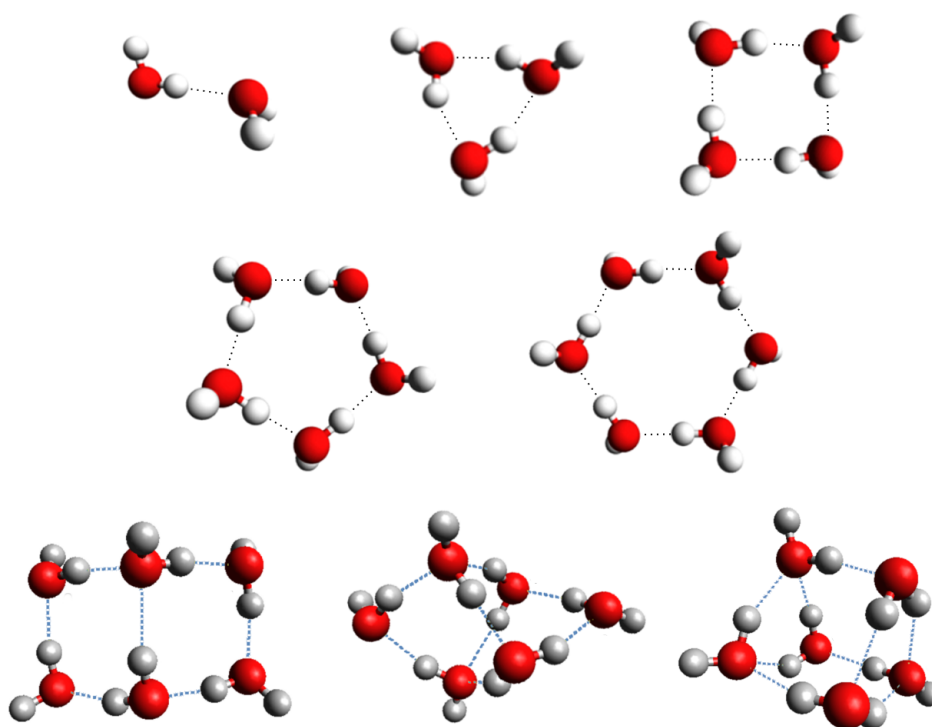
Figure 1: Small water clusters used as models in the present research.

Given this context, the present manuscript is structured as follows. First, we present a brief outline of the IQA partition and an strategy for its application to DFT electronic energies, emphasising the division of the exchange-correlation contribution. Later, we address the computational details employed in this investigation to finally discuss the main results of this work. Overall, we found that IQA/DFT analyses based on most of the functionals considered herein describes suitably the non-additive effects previously described, thus paving the way for future studies on relatively large hydrogen bonded systems.

Interacting Quantum Atoms approach

The Quantum Theory of Atoms in Molecules (QTAIM) formulated by Bader and his coworkers defines a partition of real space from the electron density of the molecular system under study. [23] This division allows in turn to decompose the total electronic energy (E) into intra and inter-atomic components by using the first (1-RDM) and second order (2-RDM) reduced density matrices. Such an energy decomposition constitutes the Interacting Quantum Atoms (IQA) approach: [15, 16]

$$E = \sum_A E_{\text{net}}^A + \sum_A \sum_{A>B} E_{\text{int}}^{AB}, \tag{1}$$

where E_{net}^A denotes the intra-atomic energy of region A and E_{int}^{AB} corresponds to the interaction energy between domains A and B. E_{net}^A is obtained as

$$E_{\text{net}}^A = T^A + V_{\text{ne}}^{AA} + V_{\text{ee}}^{AA}, \tag{2}$$

with T^A, V_{ne}^{AA} and V_{ee}^{AA} being the kinetic energy of atom A, the nucleus-electron interaction and the electron-electron repulsion within atom A, respectively. Similarly, E_{int}^{AB} stands for the sum of the nucleus-nucleus (V_{nn}^{AB}) and the electron-electron (V_{ee}^{AB}) repulsions, together with the attraction between the electrons located in atom A and the nucleus in atom B (V_{ne}^{BA}) and vice versa (V_{ne}^{AB}),

$$E_{\text{int}}^{AB} = V_{\text{nn}}^{AB} + V_{\text{ee}}^{AB} + V_{\text{ne}}^{AB} + V_{\text{ne}}^{BA}. \tag{3}$$

One attractive feature of IQA is the fact that it is based on orbital-invariant scalar functions which enables the separation of E in terms corresponding to atoms and to atomic pairs in an unambiguous and transparent manner. Another important feature of the IQA methodology in the study of intermolecular interactions is the ability to put together atoms to form groups. These groups, or superbasins, can be identified with functional groups within a molecule, or with molecules within a molecular cluster. We can define in this way the energy of a collection of atoms $\mathscr{G}$ as

$$E_{\text{net}}^{\mathscr{G}} = \sum_{A \in \mathscr{G}} E_{\text{net}}^A + \frac{1}{2} \sum_{A \in \mathscr{G}} \sum_{\substack{B \in \mathscr{G} \\ A \neq B}} E_{\text{int}}^{AB}, \tag{4}$$

while the interaction energy between groups $\mathscr{G}$ and $\mathscr{H}$ reads,

$$E_{\text{int}}^{\mathscr{G},\mathscr{H}} = \sum_{A \in \mathscr{G}} \sum_{B \in \mathscr{H}} E_{\text{int}}^{AB}. \tag{5}$$

The change in energy related with the formation of a molecular cluster from molecules $\mathscr{G} \cdots \mathscr{H} \cdots \mathscr{I}$ can be computed using the formula:

$$\Delta E = \sum_{\mathscr{G}} E_{\text{def}}^{\mathscr{G}} + \sum_{\mathscr{G}} \sum_{\mathscr{G} > \mathscr{H}} E_{\text{int}}^{\mathscr{G}\mathscr{H}}, \tag{6}$$

wherein $E_{\text{def}}^{\mathscr{G}}$ is the difference of the energy of group $\mathscr{G}$ within the molecular cluster $\mathscr{G} \cdots \mathscr{H} \cdots \mathscr{I}$ and in its isolated state. [15, 20] It is possible to modify equation (6) by adding fractions of $E_{\text{def}}^{\mathscr{G}}$ and $E_{\text{def}}^{\mathscr{H}}$ to $E_{\text{int}}^{\mathscr{G},\mathscr{H}}$. The resulting expression is a pairwise sum $\mathscr{G} \cdots \mathscr{H}$ of interacting monomers.

$$\begin{aligned} \Delta E &= \sum_{\mathscr{G}} \sum_{\mathscr{G} > \mathscr{H}} \left(E_{\text{int}}^{\mathscr{G}\mathscr{H}} + \left(\frac{E_{\text{int}}^{\mathscr{G}\mathscr{H}}}{\sum_{\mathscr{I} \neq \mathscr{G}} E_{\text{int}}^{\mathscr{I}\mathscr{G}}} \right) E_{\text{def}}^{\mathscr{G}} + \left(\frac{E_{\text{int}}^{\mathscr{G}\mathscr{H}}}{\sum_{\mathscr{I} \neq \mathscr{H}} E_{\text{int}}^{\mathscr{I}\mathscr{H}}} \right) E_{\text{def}}^{\mathscr{H}} \right) \\ &= \sum_{\mathscr{G}} \sum_{\mathscr{G} > \mathscr{H}} E_{\text{int}}^{\mathscr{G}\mathscr{H}\,\prime}. \end{aligned} \tag{7}$$

The electron-electron term can be further divided into Coulombic and exchange-correlation terms. This division enables a rearrangement of the whole interatomic energy as the addition of classical (V_{cl}^{AB}) and exchange-correlation (V_{xc}^{AB}) contributions,

$$E_{\text{int}}^{AB} = V_{\text{cl}}^{AB} + V_{\text{xc}}^{AB}. \tag{8}$$

V_{xc}^{AB} and V_{cl}^{AB} are identified with covalent and ionic components of the interaction between atoms A and B. [16]

IQA implementation of DFT

The IQA energy decomposition implies the calculation of (i) the kinetic energy, which depends on the 1-RDM, (ii) the Coulombic or classical (cl) interaction energy that is determined solely by the electron density $\rho(\mathbf{r})$ of the system and (iii) the exchange-correlation energy that entails the computation of the corresponding density, $\rho_{\text{xc}}(\mathbf{r}_1, \mathbf{r}_2)$, which is related to the pair density $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$, by the following relationship:

$$\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \rho(\mathbf{r}_1)\rho(\mathbf{r}_2) + \rho_{\text{xc}}(\mathbf{r}_1, \mathbf{r}_2) \tag{9}$$

The dependence of the IQA partition on 1-RDM and 2-RDM impedes the direct use of DFT in IQA. This problem had been previously bypassed by assigning the exchange-correlation energy not accounted by the exchange obtained directly from the Kohn-Sham determinant to intra-atomic exchange-correlation terms. This approach leads to the complete recovery of the electronic energy by the IQA partition, although, it overestimates systematically the exchange-correlation component of V_{ee}^{AA} in detriment of E_{int}^{AB}. [17]

Some of us proposed an alternative approach that uses scaling arguments, which results in a complete recovery of the total electronic energy and a correction of the intra- and interatomic exchange-correlation energies. [18] We describe now this coupling of IQA and DFT, which is the one followed in this work. For a non-hybrid exchange-correlation functional $\epsilon(\mathbf{r})$, the DFT exchange-correlation energy, E_{xc}^{DFT}, can be calculated as

$$E_{xc,loc}^{DFT} = \int_{\infty} d\mathbf{r}\rho(\mathbf{r})\epsilon(\mathbf{r}), \tag{10}$$

For hybrid functionals Equation (10) takes the form

$$E_{xc}^{DFT} = E_{xc,loc}^{DFT} + a_0 E_x^{KS}, \tag{11}$$

a_0 is the fraction of the Hartree-Fock exchange used by the functional ϵ_{xc} and E_x^{KS} is the exchange energy calculated using the Kohn-Sham molecular orbitals.

We define now the quantity λ_A as in reference [18]

$$\lambda_A = a_0 + \frac{1}{E_x^{A,KS}} \int_A \rho(\mathbf{r})\epsilon(\mathbf{r}), \tag{12}$$

in which

$$E_{xc}^A = E_{xc}^{AA} + \frac{1}{2}\sum_{B\neq A} E_{xc}^{AB}. \tag{13}$$

with

$$E_{xc}^{AA} = \frac{1}{2}\int_{\Omega A} d\mathbf{r}_1 \int_{\Omega A} d\mathbf{r}_2 r_{12}^{-1}\rho_{xc}(\mathbf{r}_1,\mathbf{r}_2), \tag{14}$$

6

and

$$E_{xc}^{AB} = \int_{\Omega A} d\mathbf{r}_1 \int_{\Omega B} d\mathbf{r}_2 r_{12}^{-1} \rho_{xc}(\mathbf{r}_1, \mathbf{r}_2). \tag{15}$$

Now, we can define intra- (A=B) and inter-atomic (A≠B) IQA/DFT exchange-correlation energies as

$$\tilde{E}_{xc}^{AB} = \frac{1}{2} \left[\lambda_A + \lambda_B \right] V_x^{AB,KS}, \tag{16}$$

from which the total DFT exchange correlation energy (LHS of Eq. (11)) can be completely recovered as

$$E_{xc}^{DFT} = \sum_A \tilde{E}_{xc}^{AA} + \sum_{A>B} \tilde{E}_{xc}^{AB}. \tag{17}$$

Computational Details

We decided to use previously reported CCSD/aug-cc-pVDZ homodromic water cluster structures, comprising from 2 to 6 water molecules, taken from the work of Segarra-Martí *et al.* [24]. Given that the structure of the monomer was not reported in this reference, we optimised the corresponding geometry using the same level of theory. Additionally, we optimised the structures of other four water hexamers using the same approximation (see Figure 1). From these geometries, we procured the electron density of the systems of interest from single point calculations for sixteen different exchange-correlation functionals:

- an LDA functional, the Slater exchange in combination with local VWN correlation (SVWN) [25]

- the generalized gradient approximations (GGA) B97, [26] BLYP, [27, 28] BP86, [27, 29] OLYP, [28, 30] PBE [31] and PW91 [32]

- the hybrid B3LYP, [28, 33] B3P86, [29, 33] and B3PW91 [32, 33]

- the meta-GGA TPSS [34] and

- the heavily parametrised Minnesota functionals M06, [35] M06-2X, [35] M06-L, [36] M06-HF [37] and M11-L [38].

We also considered the Hartree-Fock, MP2 and CCSD wavefunction approximations. All the electronic structure calculations were performed with the GAMESS-US [39] and PySCF [40] packages. Later on, densities obtained from those single points were analysed and their corresponding eletronic energies partitioned under the IQA formalism, using our PROMOLDEN code. [41] The calculation used β-spheres with radii between 0.1 and 0.3a.u. Restricted angular Lebedev quadratures were used. Inside the β-spheres, we considered 451 mapped radial point trapezoidal quadratures and L expansions truncated at $l = 10$. Outside the β-spheres, we increased the number of mapped radial points to 651 and L up to $l = 12$. Finally, we visualised our results with the PYTHON library MATPLOTLIB [42] and the AVOGADRO program [43].

Results

In order to give a comprehensive perspective about the quality of the IQA descriptors using the exchange-correlation functionals addressed in this investigation, we assess the most relevant energetic features of the studied water clusters. Namely, we consider (i) cooperative effects in formation energies, (ii) IQA deformation and interaction energies and (iii) classical, exchange-correlation and total interaction energies of covalent O−H bonds. Later on, we examine the hydrogen bond hierarchy established by considering different structures of $(H_2O)_6$. We consider (iv) the classical and exchange-correlation energies and (v) QTAIM delocalisation indices of different types of hydrogen bond. The DFT functionals employed are tested against HF and MP2 references, which show similar behaviours. Due to the excessive computational resources requested to perform IQA analyses for medium and large-sized systems, comparisons with MP2 results without any approximation (like the one proposed by Müller [44]) cannot be extended to water hexamers. [20] Moreover, CCSD calculations were also performed to serve as a reference in the hydrogen bond cooperativity in homodromic structures.

Hydrogen bond cooperative effects

As discussed in the introduction, cooperativity is a prominent feature of hydrogen bonding homodromic water clusters, (i.e., H_2O rings in which every single molecule forms two hydrogen bonds with two different molecules, being in one case the hydrogen donor and in the other, the hydrogen acceptor) are widely regarded as prototypical systems in the study of hydrogen

8

bond cooperativity. [19] These effects are manifested in the mutual strengthening of hydrogen bonding within these structures.

A way to assess the hydrogen bond cooperativity in cluster formation energies is to compute the difference between the energy of the process $(H_2O)_{n-1} + H_2O \rightleftharpoons (H_2O)_n$ (ΔE_n) and that of the water dimer, (ΔE_2) for which there is only one hydrogen bond,

$$\Delta \Delta E = \Delta E_n - \Delta E_2. \tag{18}$$
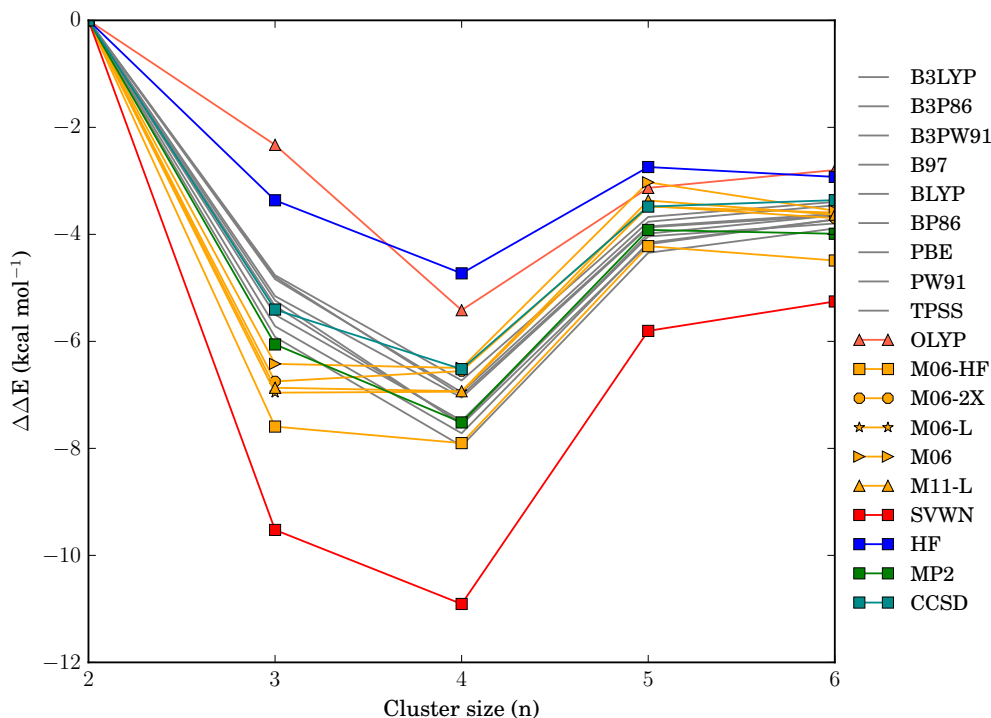
Figure 2: $\Delta\Delta E$ computed using Equation (18) with all the exchange-correlation functionals and wavefunction approximations considered in this investigation. Hydrogen bond cooperativity is indicated by negative values of $\Delta\Delta E$.

Figure 2 shows the computed values of $\Delta\Delta E$ as a function of the cluster size (n = 2 − 6). We note that $\Delta\Delta E$ is negative in all cases, which means that all the considered approximations exhibit cooperativity in the examined homodromic water clusters. Concerning the wavefunction reference methods, HF, MP2 and CCSD, they all behave in a similar manner, resulting in trends and differences among them of a few kcal/mol. By considering CCSD calculations as reference, we can say that MP2 overestimates hydrogen bond cooperativity but to a lesser extent than HF

underestimates it.

The CCSD reference behaviour is reproduced remarkably well by most DFT functionals tested. There are, however, some notable exceptions. First, the LDA-like functional SVWN strongly overestimates hydrogen bond cooperativity. This large deviation is not surprising given the simplicity of the LDA-like functional. Second, every Minnesota functional predict very similar values of hydrogen bond cooperativity for the water trimer and tetramer ($\Delta\Delta E_3 \approx \Delta\Delta E_4$), while for the CCSD reference, we observe that $|\Delta\Delta E_4| > |\Delta\Delta E_3|$. The reason behind this different behaviour may lie in the heavy parameterisation of the Minnesota family of functionals, although this should be further investigated. Finally, OLYP underestimates systematically the amount of cooperativity with respect to the CCSD results, in a very similar fashion to HF. This circumstance may result from the use of a different optimised exchange functional [30] in place of the standard one developed by Becke. [27]

IQA's deformation and interaction energies

We can understand the formation of a molecular cluster as a three-step process within the IQA formalism. In the first step, the interacting monomers rearrange their nuclear positions to those in the complex. This usually leads to a small energy penalty known as preparation or strain energy. The electronic structure of the prepared monomers may then be envisaged to suffer a distortion (that includes polarization, charge transfer, electron delocalisation, etc) that leads to the final distribution as found in the complex. The sum of the energy penalties of these two first steps will be called here deformation energy (E_{def}). [15] In the third step, the rearranged molecules interact to form the molecular cluster in what is denominated as interaction energy (E_{int}). [15] The value of E_{int} for stable compounds is always negative. The formation energy of the molecular cluster is the sum of the deformation and interaction energies ($E_{\text{form}} = E_{\text{def}} + E_{\text{int}}$). Figure 3 shows the deformation, interaction and formation energies per water molecule for the studied clusters. The trends are identical for all the considered approximations: E_{def} and $|E_{\text{int}}|$ increase with the size of the cluster. Additionally, the magnitude of E_{int} rises faster than $|E_{\text{def}}|$ with respecto to the number of interacting monomers (n), which results in an enlargement of E_{form} with the size of the cluster. In all cases, the three quantities exhibit an asymptotic behaviour with n.

The only noticeable deviation from the CCSD reference values of CCSD concerning hydrogen bond cooperative effects in small water clusters is the clear increase of $|E_{\text{form}}|$ by the functional
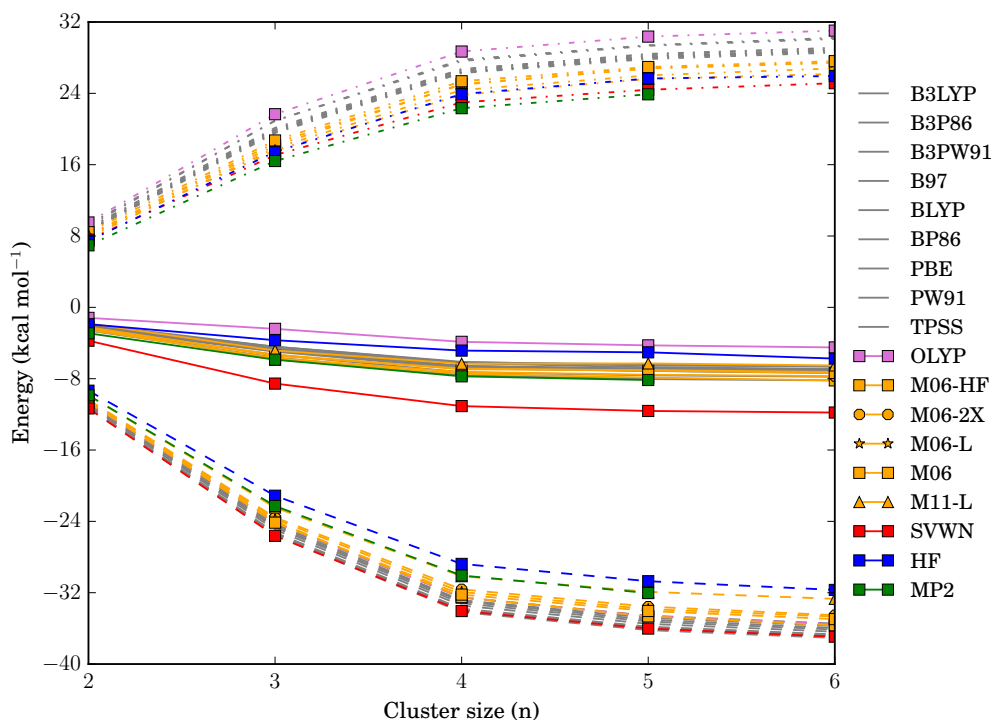
Figure 3: Formation (solid), deformation (dotted) and interaction energies (dashed lines) per water molecule in homodromic water clusters $(H_2O)_n$ (n = 2 − 6).

SVWN. This circumstance results from a combination of an underestimation of the deformation energy, E_{def}, along with an overestimation of the interaction energy, E_{int}. The low values of E_{def} are probably caused by the failure of SVWN to take into account the subtle changes ocurring in the deformation processes of the formation of the molecular clusters under consideration.

Interaction energy of the O−H covalent bond

The effect of including extra water molecules to homodromic water clusters can be observed not only in the IQA interaction and deformation energies, but also in the intramolecular bond energies. In particular, the covalent O−H bonds in which the hydrogen atom is H-bonded to another water molecule are expected to suffer a significant change due to the aforementioned hydrogen bond cooperative effects. Figure 4 shows the differences in the interaction energy, and its exchange-correlation and classical components, for the O−H covalent bond for the homodromic water clusters $(H_2O)_n$ with n = 2 − 6 with respect to the isolated monomer. The covalency of the O−H interaction diminishes with the inclusion of an additional water monomer (dotted lines) with a slight underestimation of ΔV_{xc} by the functional SVWN. We did not find

11

any anomaly among the functionals beyond a moderate underestimation of V_{xc} by the LDA-like functional SVWN.
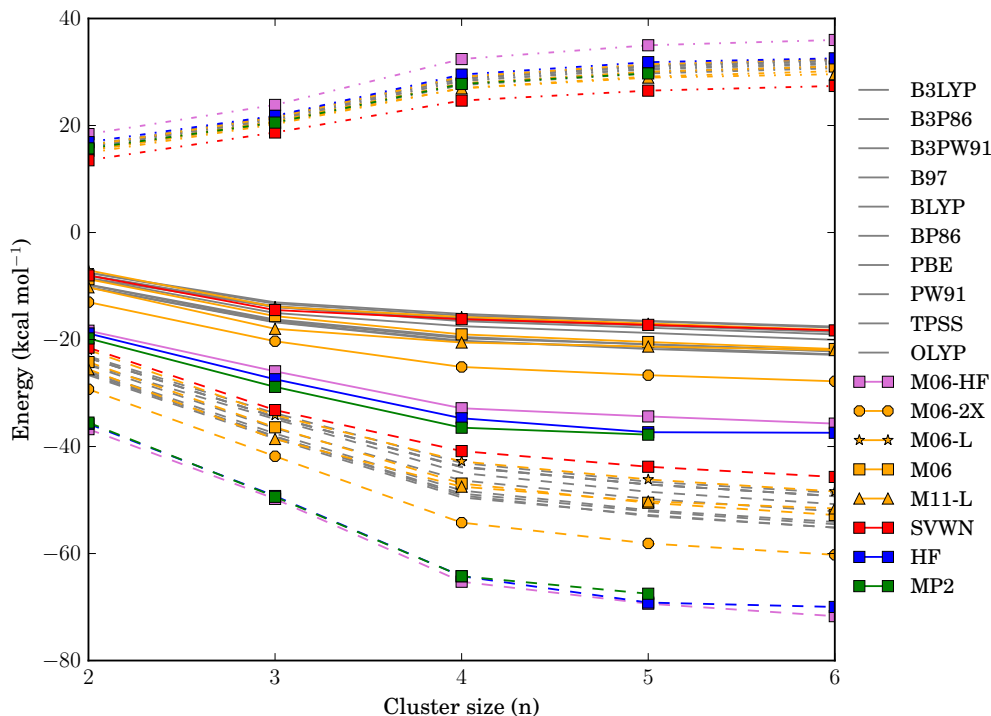
Figure 4: Differences in exchange-correlation (dotted), classical (dashed) and total interaction energies (solid) for O−H bonds in which the hydrogen atom is H-bonded to another water molecule within the homodromic clusters $(H_2O)_n$ (n = 2 –6). The values are given with respect to the monomer average energies.

Regarding the classical component of the interaction energy shown in Figure 4 with dashed lines, the magnitude of the charge of the O and hydrogen-bonded H basins increases when more molecules are included in the system. [19] Accordingly, the addition of water molecules makes the classical component of the O−H interaction energy more stabilising, because of the larger charge of the involved atoms. Although all the approximations follow this trend, the MP2, HF, and the HF-like M06-HF results are clearly separated from the rest. For example, HF overestimates ΔV_{cl}. This overestimation might occur due to the overdelocalised electron density of the uncorrelated HF method, that leads to larger atomic charges than those computed in correlated calculations, and hence to an enlargement of the classical part of the interaction energy. [45] Finally, the magnitude of the overall change in the O−H interaction energy in $(H_2O)_n$ clusters raises with n, because $|\Delta V_{cl}^{O–H}| > |\Delta V_{xc}^{O–H}|$. Once again, the MP2, HF and M06-HF values are separated

Table 1: Scale of hydrogen bond formation energies within water clusters proposed in this study. The hierarchy is presented in an increasing order of magnitude. Taken from reference [20].

Type of HB	Description
(1)	(i) the H atom involved in the hydrogen bond belongs to a double HB donor and (ii) the oxygen that participates in the interaction acts as a double HB acceptor.
(2)	(i) the hydrogen of a double HB donor is bonded to the oxygen of a single HB acceptor or (ii) the oxygen of a double acceptor interacts with a hydrogen of a single donor.
(3)	a hydrogen bond is formed between two double HB donors or two double HB acceptors.
(4)	a hydrogen of a single HB donor is bonded to the oxygen of a single HB acceptor.
(5)	(i) a hydrogen of a double HB acceptor is in contact with the oxygen of a single donor or (ii) the O atom of a double donor interacts with a hydrogen of a single acceptor.
(6)	the oxygen of a double HB donor interacts with a hydrogen of a double HB acceptor

from the rest of approximations considered herein, which describe $\Delta E_{\text{int}}^{\text{O-H}}$ in a very similar way.

Types of hydrogen bonds

The water clusters studied so far are prototypical systems where only hydrogen-bond cooperativity takes place. We can also consider anticooperative effects (i.e., the weakening that the ocurrence of certain hydrogen bonds cause on others) which appear in water hexamers different from the ring structure. We suggested previously a hydrogen bond strength hierarchy based on the coordination number of the hydrogen acceptor and donor involved in the interaction. [20] This hierarchy is shown in Table 1. We assess now this hierarchy with the different functionals considered herein.

Figure 5 depicts the classical and exchange-correlation parts of E_{int}' from Equation (7). In all cases, the dominant term is the exchange-correlation contribution (in solid lines). The classical part of E_{int}' is much smaller in absolute value but still an attractive contribution for all the
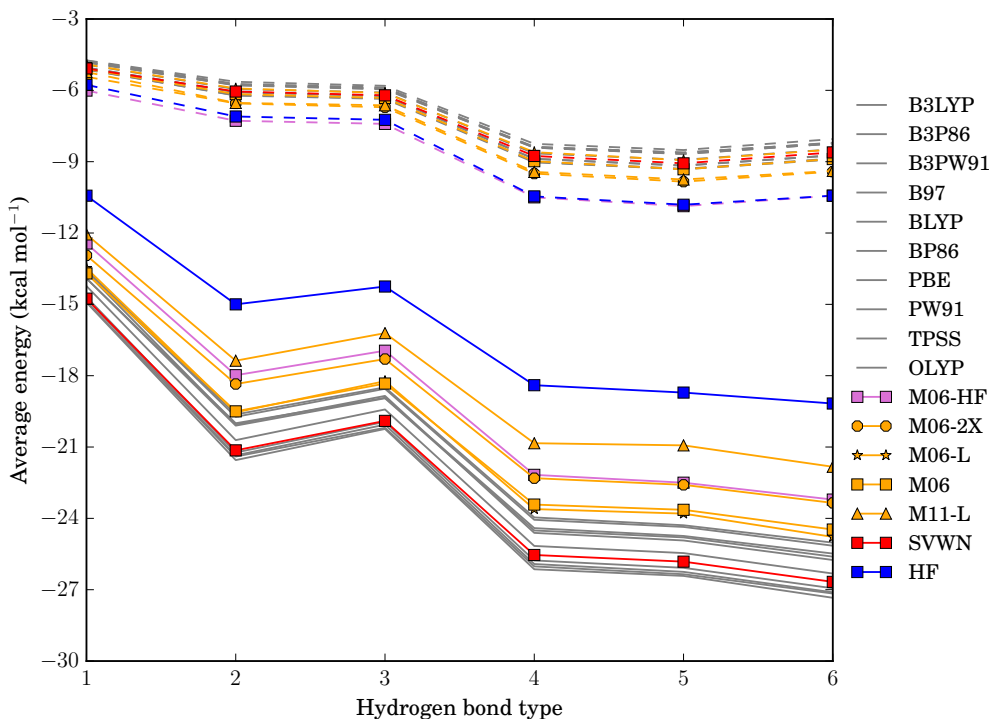
Figure 5: Average values of the classical (dashed) and exchange-correlation (solid lines) contributions to the interaction energy between water molecules computed with the aid of Equation (7).

examined exchange-correlation functionals. The trends in classical and exchange-correlation components match overall those reported in reference [20]. The only appreciable disparity concerns the exchange-correlation energies of the hydrogen bonds of types 2 and 3 (Table 1) whose hydrogen bond strength are in the inverse order with respect to those previously reported in reference [20]. Another noticeable feature of Figure 5 is the appreciable underestimation of the exchange-correlation energy by the HF method. To further assess these results, we computed the delocalisation indexes (DI) between the water molecules within the addressed water hexamers. Figure 6 shows the average value of the delocalisation indices for the different types of hydrogen bond for each functional and for HF. We observe again that hydrogen bonds of type 2 had a larger covalent component. This in contrast with out previous results. [20]

Despite the slight discrepancies observed in the description of hydrogen cooperative and anticooperative effects, most of the exchange-correlation functionals considered herein describe suitably the hydrogen bond non-additivity within water clusters. This results paves the wave for the utilisation of the addressed functionals in the study of larger warter clusters and hydrogen-
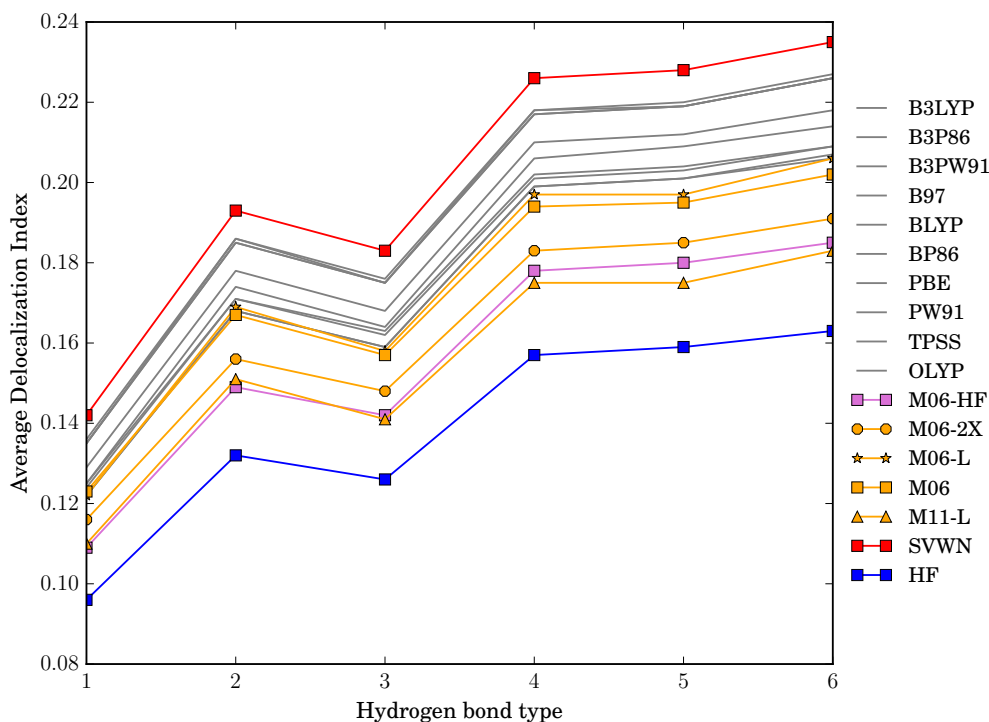
Figure 6: Averages of the delocalisation indices for the type of hydrogen bonds in Table 1.

bonded molecular clusters in general.

Conclusions

We have assessed the ability of a variety of DFT functionals to describe hydrogen bond cooperative and anticooperative effects in different water clusters. Our results indicate that most functionals correctly characterise the tendencies of the IQA partition of the electronic energy computed with ab-initio methodologies. All but the SVWN, OLYP and Minnesota's functionals reproduce the hydrogen bond cooperativity profile of the reference. Only the SVWN overestimates the interaction energy in an appreciable manner, and the interactions $O-H$ and $O{\cdots}H$ are adequately reproduced, being the HF-like M06-HF the closest functional to the HF and MP2 references. When applied to the hydrogen bond hierarchy classification, the same conclusions are found. Regarding the decrease in the quality of the electron densities reported by Medvedev *et al.*, [22] we have not found any appreciable effect in the newest functionals, at least as the description of the non-additive effects in the hydrogen bonds addressed in this investigation is regarded. Overall, we hope that this study represents a basis for the application

of IQA/DFT to a larger vaierty of chemical systems, such as, large water clusters.

Acknowledgements

References

[1] A. D. Becke, *J. Chem. Phys.* **2014**, *140*.

[2] H. S. Yu, S. L. Li, D. G. Truhlar, *J. Chem. Phys.* **2016**, *145*.

[3] R. Galindo-Murillo, M. E. Sandoval-Salinas, J. Barroso-Flores, *J. Chem. Theory Comput.* **2014**, *10*, 825–834.

[4] A. J. Cohen, P. Mori-Sánchez, W. Yang, *Chem. Rev.* **2012**, *112*, 289–320.

[5] D. Lupp, N. J. Christensen, J. R. Dethlefsen, P. Fristrup, *Chem. Eur. J.* **2015**, *21*, 3435–3442.

[6] P. Geerlings, F. De Proft, *Phys. Chem. Chem. Phys.* **2008**, *10*, 3028–3042.

[7] R. A. Miranda-Quintana, P. W. Ayers, *Theor. Chem. Acc.* **2019**, *138*, 44.

[8] M. Franco-Pérez, C. A. Polanco-Ramírez, J. L. Gázquez, P. W. Ayers, *J. Mol. Model.* **2018**, *24*, 285.

[9] F. Cortés-Guzmán, R. Bader, *Coord. Chem. Rev.* **2005**, *249*, 633–662.

[10] E. Romero-Montalvo, J. M. Guevara-Vela, W. E. Vallejo Narváez, A. Costales, A. Martín Pendás, M. Hernández-Rodríguez, T. Rocha-Rinza, *Chem. Commun.* **2017**, *53*, 3516–3519.

[11] R. F. W. Bader, P. M. Beddall, *J. Chem. Phys.* **1972**, *56*, 3320–3329.

[12] A. D. Becke, K. E. Edgecombe, *J. Chem. Phys.* **1990**, *92*, 5397–5403.

[13] C. Gatti, F. Cargnoni, L. Bertini, *J. Comput. Chem.* **2003**, *24*, 422–436.

[14] P. L. A. Popelier, *Applications of Topological Methods in Molecular Chemistry*, Springer International Publishing, **2016**.

[15] M. Á. Blanco, Á. Martín Pendás, E. Francisco, *J. Chem. Theory Comput.* **2005**, *1*, 1096–1109.

[16] E. Francisco, Á. Martín Pendás, M. Á. Blanco, *J. Chem. Theory Comput.* **2006**, *2*, 90–102.

[17] P. Maxwell, Á. Martín Pendás, P. L. A. Popelier, *Phys. Chem. Chem. Phys.* **2016**, *18*, 20986–21000.

[18] E. Francisco, J. L. Casals-Sainz, T. Rocha-Rinza, Á. Martín Pendás, *Theor. Chem. Acc.* **2016**, *135*.

[19] J. Guevara-Vela, R. Chávez-Calvillo, M. García-Revilla, J. Hernández-Trujillo, O. Christiansen, E. Francisco, Á. MartínPendás, T. Rocha-Rinza, *Chem. Eur. J.* **2013**, *19*, 14304–14315.

[20] J. Guevara-Vela, E. Romero-Montalvo, V. Mora Gómez, R. Chávez-Calvillo, M. García-Revilla, E. Francisco, Á. Martín Pendás, T. Rocha-Rinza, *Phys. Chem. Chem. Phys.* **2016**, *18*, 19557–19566.

[21] M. J. Gillan, D. Alfè, A. Michaelides, *J. Chem. Phys.* **2016**, *144*, 130901.

[22] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, K. A. Lyssenko, *Science* **2017**, *355*, 49–52.

[23] R. Bader, *Atoms in molecules : a quantum theory*, Clarendon Press, Oxford New York, **1990**.

[24] J. Segarra-Martí, M. Merchan, D. Roca-Sanjuan, *J. Chem. Phys.* **2012**, *136*, 244306.

[25] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200–1211.

[26] A. D. Becke, *J. Chem. Phys.* **1997**, *107*, 8554–8560.

[27] A. D. Becke, *J. Chem. Phys.* **1988**, *88*, 2547.

[28] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.

[29] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822–8824.

[30] N. A. Handy, A. J. Cohen, *Mol. Phys.* **2001**, *99*, 403–412.

[31] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[32] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, C. Fiolhais, *Phys. Rev. B* **1992**, *46*, 6671–6687.

[33] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.

[34] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, *Phys. Rev. Lett.* **2003**, *91*, 146401.

[35] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215–241.

[36] Y. Zhao, D. G. Truhlar, *J. Chem. Phys.* **2006**, *125*, 194101.

[37] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2006**, *110*, 13126–13130.

[38] R. Peverati, D. G. Truhlar, *J. Phys. Chem. Lett.* **2012**, *3*, 117–124.

[39] M.W.Schmidt, K.K.Baldridge, J.A.Boatz, S.T.Elbert, M.S.Gordon, J.J.Jensen, S.Koseki, N.Matsunaga, K.A.Nguyen, S.Su, T.L.Windus, M.Dupuis, J.A.Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347.

[40] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, G. K.-L. Chan, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1340.

[41] Á. Martín Pendás, E. Francisco, *Promolden. A QTAIM/IQA code (unpublished).*

[42] J. D. Hunter, *Computing In Science & Engineering* **2007**, *9*, 90–95.

[43] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchison, *J. Cheminformatics* **2012**, *4*, 17.

[44] A. Müller, *Phys. Lett. A* **1984**, *105*, 446–452.

[45] J. Thirman, M. Head-Gordon, *J. Phys. Chem. Lett.* **2014**, *5*, 1380–1385.

Fluorine Conformational Effects Characterized by Energy Decomposition Analysis

Natalia Díaz,* Fernando Jiménez-Grávalos, Dimas Suárez, Evelio Francisco, and Ángel Martín-Pendás

Departamento de Química Física y Analítica. Universidad de Oviedo. Avda. Julián Clavería 8, 33006 Oviedo (Asturias). Spain.

diazfnatalia@uniovi.es

Abstract

Electrostatic and stereoelectronic effects associated to fluorine atoms can be exploited as conformational tools for the design of shape-controlled functional molecules. To gain further insight into the nature and strength of these effects, we use the Interacting Quantum Atoms (IQA) method augmented with the semiclassical pairwise dispersion potential to decompose the conformational energies of fluoro-substituted molecules into fragment-based energy contributions, which include deformation/distortion terms and the electrostatic, exchange-correlation and dispersion interactions. The studied molecules comprise various $F\text{-}CH_2\text{-}CH_2\text{-}X$ and $F\text{-}CH_2\text{-}CO\text{-}X$ systems, as well as selected conformers of an α,β-difluoro-γ-amino-acid derivative that is potentially useful for the design of shape-controlled bioactive amino acids and peptides. We identify the most relevant exchange-correlation and/or electrostatic interaction terms contributing to the stability of the various conformers, and we show that IQA can assess the *gauche/anti* or *trans/cis* preferences in molecules with two or more rotatable bonds as well as to study the roles played by other concomitant effects (*e.g.*, CH/OH/NH···F contacts). For the α,β-difluoro-γ-amino acid derivatives, our theoretical analysis points out that the *gauche/anti* and *trans/cis* effects associated to fluorine bonds can be significantly attenuated by other specific intra-molecular contacts.

1

Introduction

In the last decades, fluorine has arisen as a remarkable element in numerous applications. Some of the most successful drugs on the market today include one or more fluorine atoms in their structures. Fluorine-containing molecules have made a significant contribution to the development of modern crop protection agrochemicals (herbicides, insecticides, and fungicides).[2, 3] Fluorinated materials are widely used in liquid crystal displays[4] and in photovoltaic solar cells.[51] Specific fluorine substitution redefines the reactivity of non-fluorinated substrates and catalysts in terms of reaction yields, diastereo- or enantiomeric ratios, and mechanistic pathways.[6, 7] Similarly, fluorination modifies hydrophobicity, ring puckering and secondary structure propensity of amino-acids, and influences peptide/protein folding, stability and intermolecular interactions.[8]

The introduction of fluorine into a molecule can affect different physicochemical properties (*i.e.* hydrophobicity, reactivity, conformation, noncovalent interactions, etc.) due to electrostatic and stereoelectronic effects. Fluorine is a small atom that can replace hydrogen in organic molecules with a minimal steric impact, but with significant electronic consequences. It is the most electronegative element and, accordingly, the C−F bond presents a high polarity (*i.e.* $C^{\delta+}-F^{\delta-}$) and a low polarizability. This results in a strong ionic character and a large dipole moment for the C−F moiety that influence the chemical properties of adjacent groups.[9] For example, pK_a values of acetic acid and its successive α-fluorinated derivatives change from 4.76 (CH_3COOH) to 0.23 (CF_3COOH) upon fluorine substitution.[10] In addition, numerous examples show that stereoselective introduction of fluorine atoms within a molecule results in different conformational properties.[11-15] On one hand, the C−F bond tends to align antiparallel to adjacent C=O bonds to confront the corresponding dipole moments,[11] and to orient itself close to positively charged groups in order to maximize favorable charge/dipole interactions.[16] On the other hand, the high electronegativity of fluorine determines the presence of a low energy σ^*_{C-F} antibonding orbital that can accommodate electron density from stereoelectronically aligned lone pairs or adjacent σ and π bonds. Although this stereoelectronic effect (hyperconjugation) is energetically small, it has been traditionally invoked to explain the stabilization of certain molecular conformations.[9]

Traditionally, 1,2-difluoroethane has been considered the benchmark case for understanding the origin of the *gauche* structural preference (*i.e.* F-C-C-X angle around 60⁰).

2

In previous computational works,[17-19] the natural bond orbital (NBO) method[20] ascribes the *gauche* preference in 1,2-difluoroethane to the $\sigma_{CH} \rightarrow \sigma_{CF}^*$ hyperconjugative interactions arising when adjacent C-H and C-F bonds align antiparallel. However, the ability of NBO to provide a right balance between Lewis (*i.e.* steric and electrostatic) and non-Lewis contributions has been questioned[21] by the Energy Decomposition Analysis (EDA) approach,[22] which similarly dismisses the central role of hyperconjugation in the conformational *gauche* effect. Thus, when EDA characterizes the interaction between two CH_2F radicals in the *gauche/anti* geometries of 1,2-difluoroethane, it turns out that equally-favorable orbital and electrostatic interactions contribute to explain the *gauche* effect.[23] Further support for the role of electrostatic interactions comes from an Interacting Quantum Atoms (IQA) analysis, in which the largest diatomic contribution to *gauche* stability is provided by the electrostatic interaction between carbon and fluorine atoms in positions 1,3.[24] As a result, the authors conclude that the origin of the *gauche* effect in 1,2-difluoroethane is electrostatic rather than hyperconjugation.

Besides 1,2-difluoroethane, the conformational effects associated to the presence of fluorine atoms have been computationally analyzed in related compounds. For instance, DFT energy profiles connecting the *gauche* and *anti* conformers and atomic charges have been evaluated for a series of 16 β-substituted α-fluoroethanes ($F-CH_2CH_2-X$) in order to assess the influence of steric and electrostatic interactions, meanwhile NBO analysis are employed to disclose the role of hyperconjugation.[18] These results point out that both electrostatic and hyperconjugative effects may contribute to conformer stability. In general, the role of hyperconjugation stabilizing the *gauche* structures increases with the electronegativity of the first atom in the X moiety. Favorable electrostatic interactions between fluorine and electropositive atoms in the X substituent also contribute to the *gauche* preference, with a remarkable *gauche* stabilization observed for positively charged groups like in 2-fluoroethylammonium ($F-CH_2CH_2-NH_3^+$).[16, 18, 25] It has been also found that F atoms adjacent to carbonyl groups stabilize the *trans/gauche* F-C-C=O arrangements in the gas-phase, although this intrinsic preference can be reversed to the alternative *cis* form in polar solvents as suggested by DFT calculations coupled with a continuum solvent model.[26-28] On the other hand, the NBO characterization of the 2-fluoroethanol ($F-CH_2CH_2-OH$) conformers has explained the largest stability of a particular *gauche* structure by the presence of a favorable electrostatic O-H$^{\delta+}\cdots^{\delta-}$F interaction, which is termed as a non-classical hydrogen bond.[29] In contrast, a very recent work interprets the $^4J_{HF}$ couplings in α-fluoro amides as a genuine N-

H···F hydrogen bond.[30] Finally, the *anti/gauche* conformers around the F-C-C-S sequence in sulfur-containing systems have been subject to EDA considering the $FH_2C\cdot$ and $\cdot CH_2SX$ radicals as reference fragments.[31] The electrostatic interaction favors preferentially the *gauche* conformer followed by the orbital interaction energy accounting for charge transfer effects. In the same study, [31] NBO indicates that stereoelectronic effects work in favor of *gauche* conformers.

Other more complex systems have been also computationally examined to analyze the conformational effects associated to the introduction of fluorine atoms. DFT geometry optimizations followed by MP2 single point calculations have revealed that the conformational landscape of a series of 25 benzyl alcohols is significantly influenced by the presence of one or two fluorine atoms in the *ortho* position.[32] Intramolecular interactions characterized in terms of Atoms in Molecules (AIM), Noncovalent Interactions (NCI) and NBO analyses, show that OH···F, CH···F and/or CH···O intramolecular contacts contribute to the stabilization of the various conformations.[32] On the other hand, the origin of the drastically different structure adopted by linear perfluoroalkanes and hydrocarbons (*i.e* helical vs zig-zag) has been inspected by means of NBO/NCI analyses using DFT wave functions. The energetic preference for the helical structures increases monotonically with the chain length of the perfluorated alkane, what has been explained in terms of $\sigma_{CC} \rightarrow \sigma_{CF}^{*}$ hyperconjugative interactions, which are absent in the zigzag conformation.[33]

1: 1,2-difluoroethane   **2:** 2-fluoroethyl-acetate   **3:** 3-fluoropropanal   **4:** *N*-(2-fluoroethyl)acetamide

5: 2-(2-fluoroethyl)isoindoline-1,3-dione   **6:** 2-fluoroethan-1-aminium   **7:** 1-(2-fluoroethyl)pyridin-1-ium

8: 1-fluoropropan-2-one   **9:** methyl-2-fluoroacetate   **10:** 2-fluoro-*N*-methylacetamide

11: 1,2,3-trifluoropropane   **12:** 2,3-difluoro-*N*-methyl propanamide   **13:** 2-(2,3-difluoropropyl) isoindoline-1,3-dione

14a/14b

Scheme 1

From our review of the former computational investigations, it turns out that a variety of effects may be behind the conformational preferences observed in fluorinated systems. However, direct comparisons among the results obtained in these studies are largely hampered by the diversity of methodologies employed. Therefore, we pursue in this work to reexamine the conformational effects exerted by fluorine atoms using a common methodology based on the IQA energy partitioning augmented with pairwise dispersion energies. The dispersion-corrected IQA approach facilitates the decomposition of the global conformational energies into fragment contributions that can be further separated into various quantum mechanical and classical electrostatic terms. To better appreciate the conformational impact of these energy contributions, we will consider two classes of organic

compounds. On one hand, we will evaluate intra and inter-atomic contributions to the relative energy of the *gauche*/*anti* conformers of fluoroethylene-derived systems (**1-7** in Scheme 1), the *trans*/*cis* conformers in carbonyl containing systems (**8-10**), and the combined *gauche-gauche* or *gauche-trans* preferences (**11-13**). On the other hand, we will characterize in detail the conformational properties of two diastereomers of an α,β-difluoro-γ-amino-acid derivative (**14a/14b** in Scheme 1) that have been proposed to be particularly relevant for the design of shape-controlled bioactive amino acids and peptides,[14] as they may link to other amino acids resulting in peptides with predictable conformational behavior. According to their crystal structures, the different stereochemistry at the α and β carbons, (*R,R*) for **14a** and (*R,S*) for **14b**, results in a different backbone conformation (extended zig-zag in **14a** and partially bent in **14b**). Both structures exhibit the normally expected *gauche* orientation in the F_{23}-C_{22}-C_{19}-F_{20} and F_{20}-C_{19}-C_{16}-N_{13} bonds and the antiparallel alignment (*trans*-planar) of the F_{23}-C_{22}-C_{25}=O_{26} moiety. In fact, these three effects are supposed to determine the final conformation of the **14a** and **14b** molecules. All in all, we report theoretical results obtained for typical model systems prone to characterize specific conformational effects, but also for actual fluorinated compounds of synthetic and biochemical interest. The IQA methodology will allow us to treat them on the same basis, yielding thus a full energetic description of the various effects influencing their conformational preferences.

Computational methods

IQA energy decomposition

The IQA approach[34] partitions the first- and second-order reduced density matrices into atomic regions such as the attraction basins (Ω_A) that stem from the topological properties of the charge distribution $\rho(r)$ as commonly defined within the framework of the quantum theory of atoms in molecules (QTAIM). Two scalar fields derived from quantum mechanical (QM) wavefunctions are required to accomplish the IQA decomposition: the first order reduced density matrix $\rho_1(\mathbf{r_1},\mathbf{r_1}')$ and the pair density, $\rho_2(\mathbf{r_1},\mathbf{r_2})$. Then IQA decomposes the total energy of a molecular system in the gas-phase as

$$E = \sum_A E_{net}^A + \sum_{A>B} E_{int}^{AB} = \sum_A \left(T^A + V_{ne}^{AA} + V_{ee}^{AA} \right) + \sum_{A>B} \left(V_{nn}^{AB} + V_{ne}^{AB} + V_{ne}^{BA} + V_{ee}^{AB} \right) \tag{1}$$

where $E_{net}^A \equiv E_{net}\left(\Omega_A\right)$ is the net electronic energy of atom A that includes the kinetic energy T^A and the potential energy due to nuclei-electron (*ne*) attractions and electron-electron repulsions (*ee*) within Ω_A. The interaction energy $E_{int}^{AB} = E_{int}\left(\Omega_A, \Omega_B\right)$ between atoms A and B in the molecular system collects various potential energy terms (*nn*, *en*, *ne* and *ee*). We stress that, according to the IQA terminology, an *interaction* energy is a diatomic contribution to the total energy of a molecule. In fact, the interaction energy E_{int}^{AB} can be divided into classical and non-classical contributions, allowing thus to define a purely classical (electrostatic) component of the interaction energy, $E_{int,class}^{AB} = V_{nn}^{AB} + V_{ne}^{AB} + V_{ne}^{BA} + V_{ee,Coul}^{AB}$, along with a quantum (exchange-correlation) contribution such that $E_{int}^{AB} = E_{int,class}^{AB} + E_{int,xc}^{AB}$. In this respect, we note also that the *classical* IQA components are distinguished only in the diatomic interaction terms E_{int}^{AB} , but not within the atomic net energies E_{net}^A .

By summing separately all the atomic and diatomic terms in equation (1), the total energy of a molecule is readily decomposed into net and interaction energy components ($E_{net} = \sum_A E_{net}^A$, $E_{int,class} = \sum_{A<B} \Delta E_{int,class}^{AB}, \dots$). By inserting the corresponding IQA energies for a given pair of *gauche* and *anti* conformers, the energy difference $E_{gauche} - E_{anti}$ can be written as

$$\Delta E = E^{gauche} - E^{anti} = \Delta E_{net} + \Delta E_{int,class} + \Delta E_{int,xc} \qquad (2)$$

Dispersion corrected IQA

In previous work,[35] we have shown that IQA calculations can be complemented with the pairwise formulation of the third-generation dispersion (D3) correction for DFT and HF methods,[36] which is a semiclassical potential inspired on the London formula for the dispersion attraction between two atoms *A* and *B* at large distance and that does not alter the underlying charge density. The correct asymptotic behavior of the *A*-*B* dispersion energy can be reproduced using the Becke-Johnson (BJ)[37] rational damping function, resulting in the D3(BJ) dispersion energy correction (termed simply as D3 in the manuscript) that can be readily combined with the IQA decomposition scheme. To this end, we simply add the dispersion interaction energies $E_{int,disp}^{AB}$ to the rest of the IQA interaction energy terms such that the total D3-corrected IQA decomposition results,

$$E = \sum_A E_{net}^A + \sum_{A>B} \left(E_{int,class}^{AB} + E_{int,xc}^{AB} + E_{int,disp}^{AB} \right) \qquad (3)$$

To keep the computational cost of the IQA calculations withtin reasonable bounds, all the IQA-D3 calculations are done with HF wavefunctions. In fact HF-D3 method provides a reasonable description of molecular structure and energetics,[38, 39] and offers a straightforward physical partitioning because HF entirely lacks dispersion energy. To further assess the utility of the HF-D3 IQA decomposition, we briefly compare in the Supporting Information (Table S12) the HF-D3 IQA results on 1,2-difluoethane with those obtained with correlated methods (B3LYP-D3 and MP2).

Interacting Quantum Fragments (IQF-D3)

In principle, the grouping of the atomic IQA energy components into chemically-meaningful fragment contributions can be done using different notations and protocols depending on the particular problem at hand. For example, to characterize the conformational energy variation upon an internal rotation about a single bond, we will distinguish at least two molecular fragments, P and Q, that are covalently linked (P-Q) through the rotating bond. Such interacting quantum fragments (IQF) partitioning of a molecular system splits the total energy in intra-fragment and inter-fragment energy terms. The intra-fragment net energy (E_{net}^P and E_{net}^Q) collects the atomic net energies and the interaction energies among the atoms placed in the same fragment, whereas the inter-fragment term (E_{int}^{PQ}) sums the classical and quantum interaction energies between atoms in different groups.

Using IQF-D3 quantities, the relative energy ΔE between the *gauche* and *anti* conformers around the P-Q bond can be expressed as:

$$\Delta E = \Delta E_{net}^P + \Delta E_{net}^Q + \Delta E_{int,class}^{PQ} + \Delta E_{int,xc}^{PQ} + \Delta E_{int,disp}^{PQ} \qquad (4)$$

where the relative IQF terms in the right hand are just the corresponding *gauche/anti* differences of the grouped terms ($e.g, \Delta E_{net}^P = E_{net}^{P,gauche} - E_{net}^{P,anti}$) and the ΔE_{int}^{PQ} term is split into the classical interaction energy $\Delta E_{int,class}^{PQ}$ and the purely QM correlation-exchange contribution $\Delta E_{int,xc}^{PQ}$.

To further clarify the meaning of the IQF terms, it may be useful to briefly discuss the similarities and differences between IQF and the energy decomposition analysis (EDA).[40, 41]

Assuming that EDA is applied to analyze the energy difference between two conformers of the P-Q system with reference to separated (unrelaxed) $P\cdot$ and $Q\cdot$ radical fragments, ΔE would be the sum of three terms: a classical electrostatic interaction energy between $P\cdot$ and $Q\cdot$ (V_{elstat}), a Pauli repulsion term (ΔE_{Pauli}) that incorporates the kinetic and potential energy changes due to the antisymmetrization of the fragment wavefunctions, and an orbital relaxation energy (ΔE_{orb}) that arises from inter-fragment charge transfer and polarization effects. In general, the $\Delta E_{net}^{P} + \Delta E_{net}^{Q} + \Delta E_{int,xc}^{PQ}$ sum in IQF (termed as exchange-correlation-repulsion) resembles the $\Delta E_{Pauli} + \Delta E_{orb}$ sum in EDA, while $\Delta E_{int,class}^{PQ}$ would correspond to V_{elstat} although this identification is exact only in the limit of weakly interacting fragments for which ΔE_{orb} tends to zero.[42] Nevertheless, the grouped EDA and IQF terms admit a similar interpretation on the basis of electrostatic and/or QM effects. However, the IQF balance between intra-fragment deformation ($\Delta E_{net}^{P} + \Delta E_{net}^{Q}$) and inter-fragment exchange-correlation effects ($\Delta E_{int,xc}^{PQ}$) provides an alternative description of the QM effects embedded in the $\Delta E_{Pauli} + \Delta E_{orb}$ term without referring to separate fragments.

Charge and Dipole electrostatic interactions

In order to deepen the description of the IQF $\Delta E_{int,class}^{PQ}$ term, the multipole expansion of the electrostatic potential may be useful for assessing the role of the charge distribution in the P and Q regions. In fact the QTAIM and IQA approaches allow us to express the Coulombic energy as a multicenter multipole expansion in a natural way.[43] It is well known that the electrostatic potential exerted by a set of point charges at a sufficiently distant point can be expanded in terms of multipole moments. In our study we focus on the first two contributions to the electrostatic potential, *i.e.*, the total charge of a given basin (q) and the dipole moment (d) emerged from its distribution. Hence, only the charge-charge, charge-dipole and dipole-dipole interactions have been considered. Their respective analytical formulation are given in Eqs. (5)-(7) (in atomic units):

$$E_{qq} = \frac{q^{P} q^{Q}}{R} \tag{5}$$

$$E_{qd} = \frac{q^{P} \mu^{Q} \mathbf{R}}{R^{3}} \tag{6}$$

$$E_{dd} = \frac{R^2 \mathbf{\mu}^P \mathbf{\mu}^Q - 3(\mathbf{\mu}^P \mathbf{R})(\mathbf{\mu}^Q \mathbf{R})}{R^5} \tag{7}$$

where the superscripts P and Q denote the two fragments considered, $\mathbf{R}$ represents the distance between their centers of mass, q^P / q^Q are the total charge in P/Q (*i.e.*, $q^P = \sum_{A \in P} q^A$) and $\mathbf{\mu}^P/\mathbf{\mu}^Q$ are the dipole moment generated by the total charge distribution within P/Q.

Starting from the atomic dipoles arisen upon a particular 3D partitioning of the space, it is not, in general, straightforward to reconstruct fragment dipoles from atomic ones. The simple addition of the corresponding dipole components is not sufficient for the study of the charge-dipole or dipole-dipole fragment interactions. For neutral molecules but not for ions, $\mathbf{\mu}$ is origin independent since any coordinate transformation $\mathbf{r} = \mathbf{r}' + \mathbf{R}_0$ gives

$$\mathbf{\mu} = \int \mathbf{r} \rho(\mathbf{r}) d\mathbf{r} = \int \mathbf{r}' \rho(\mathbf{r}' + \mathbf{R}_0) d\mathbf{r}' + \mathbf{R}_0 \int \rho(\mathbf{r}' + \mathbf{R}_0) d\mathbf{r}' = \int \mathbf{r}' \rho(\mathbf{r}' + \mathbf{R}_0) d\mathbf{r}' + \mathbf{R}_0 q \tag{8}$$

so that the charge-translation $\mathbf{R}_0 q$ term vanishes for neutral molecules. To derive an atomic decomposition of $\mathbf{\mu}$, the charge density ρ can be partitioned into disjoint atomic regions with their own origin of coordinates at positions $\mathbf{R}_0^A$, and performing thus a transformation $\mathbf{r} = \mathbf{r}^A + \mathbf{R}_0^A$ in each basin. The total dipole is then expressed as

$$\mathbf{\mu} = \sum_A \left(\int_{\Omega_A} \mathbf{r}^A \rho(\mathbf{r}^A + \mathbf{R}_0^A) d\mathbf{r}^A + \mathbf{R}_0^A \int_{\Omega_A} \rho(\mathbf{r}^A + \mathbf{R}_0^A) d\mathbf{r}^A \right) = \sum_A \left(\mathbf{\mu}_{int}^A + \mathbf{R}_0^A q^A \right) \tag{9}$$

Eq. (9) shows that $\mathbf{\mu}$ is the summation of atomic dipoles $\mathbf{\mu}^A$ that have two contributions: the *intrinsic* term $\mathbf{\mu}_{int}^A$ that comes from the integration of the dipole density function within the atomic basin plus the corresponding charge-translation term $\mathbf{R}_0^A q^A$. Fragment-based dipoles $\mathbf{\mu}^P$ can be likewise defined by considering both the intrinsic contributions and the net charges q^A of the atomic basins that constitute fragment P:

$$\mathbf{\mu}^P = \sum_{A \in P} \mathbf{\mu}_{int}^A + \sum_{A \in P} \left(\mathbf{R}_0^A - \mathbf{R}_0^P \right) q^A \tag{10}$$

The reconstructed dipoles $\mathbf{\mu}^P$ depend on the origin of coordinates which, in this work, are selected as the center of mass of fragment P ($\mathbf{R}_0^P$). We note that, in the case of neutral molecules, the dependency of the $\mathbf{\mu}^P$ values on the origin of coordinates is small. We also

emphasize that the total electrostatic interaction energy among the various atoms/fragments in a given molecule computed from the multipolar expansion is always origin-independent.

QM calculations

Initial coordinates for the small and medium-sized models were generated using the UCSF Chimera program.[44] Starting from 1,2-difluoroethane in the *gauche* conformation, all the CH_2F-CH_2X systems were built in the *gauche* conformation by replacing the second fluorine atom by the appropriate X group. Then the initial geometries of the corresponding *anti* conformers were obtained by adjusting the F-C-C-X angle to 180°. Similarly, we built the initial *trans/cis* conformers for the CH_2F-COX systems.

All the *gauche/anti* and *trans/cis* structures of the small models were fully optimized in the gas-phase with no symmetry constraints at the HF-D3/cc-pVTZ and the RI-MP2/cc-pVTZ levels of theory. These calculations were done with the ORCA 4.0.1 package.[45] The HF-D3/cc-pVTZ energy minimizations were carried out with the D3 dispersion energy and gradient corrections choosing the Becke-Johnson (BJ) damping function.[36, 37] The resolution-of-the-identity (RI) approximation was activated for the MP2 calculations using the appropriate auxiliary basis set. To further estimate electron correlation effects on the conformational energies, we employed the domain-based Local Pair Natural Orbital (DLPNO) coupled cluster method[45, 46] as implemented in the ORCA 4.0.1 package. More specifically, we performed DLPNO-CCSD(T)/aug-cc-pVTZ single point calculations on the MP2/cc-pVTZ geometries using tight thresholds to control the DLPNO approximations. The DLPNO-CCSD(T) method exhibits near linear scaling at the cost of introducing small deviations from the canonical coupled cluster methods. For tight DLPNO thresholds, the typical errors with respect to canonical results are below 0.25 kcal/mol.[47]

For 1,2,3-trifluoropropane and the related CH_2F-CHF-CH_2X system with an isoindole-derivative as the X substituent, initial coordinates were generated by setting the F-C-C-F and F-C-C-X angles to ±60° and/or 180°, resulting in nine structures accordingly named g^+/g^+, g^+/g^-, $g^+/anti$, etc. In the case of the CH_2F-CHF-CONHCH$_3$ molecule, the F-C-C=O angle takes the values 180° for the *trans* and 0° for the *cis* conformers, which combined with the three options available for the F-C-C-F torsion result in six initial structures termed *g+/trans, g+/cis, g-/trans*, etc. As in the smaller CH_2F-CH_2X systems, initial geometries were fully optimized in the gas-phase at the HF-D3/cc-pVTZ and the RI-MP2/cc-pVTZ levels of theory

followed by single point DLPNO-CCSD(T)/aug-cc-pVTZ calculations on the MP2/cc-pVTZ geometries.

For the **14a** and **14b** compounds, automatic conformational analyses were performed using the multiconformer generator MS-DOCK program.[48] Starting from the corresponding crystal structures,[14] we used the antechamber program to optimize the initial geometries and to assign AM1-BCC atomic charges and SYBYL atom types.[49] Subsequently, 50 different conformers were generated for **14a** and **14b** using the MS-DOCK software and the resulting structures were optimized using the MMFF94 force field. During conformer generation, a filter based on the root mean squared deviation of the Cartesian coordinates and the energy was applied to eliminate similar structures. The **14a** and **14b** different conformers were then optimized at the HF-D3/cc-pVDZ level of theory using the ORCA 4.0.1 program.[45] After inspection of the structure and energy of the minimized structures, only 20 different conformers remained for **14a** and **14b**. These were reoptimized at the HF-D3/cc-pVTZ level of theory using ORCA both in the gas phase and in chloroform using the SMD continuum solvent model.[50] For the five more stable conformers of **14a** and **14b**, selected NMR vicinal proton-fluorine coupling constants $^3J_{H,F}$ were estimated using a seven-parameter Karplus-type relation.[51] The equation correlates a $^3J_{H,F}$ value with the corresponding H-C-C-F torsion angle and it also includes correction terms for substituent electronegativity a well as for H-C-C and F-C-C bonds angles. The $^3J_{H,F}$ values were computed using the HF geometries optimized in chloroform, whereas the electronegativity values for the Cα-Cβ substituents (1.4 for fluorine, 0.9 for carbon and 0.0 for hydrogen) were taken from reference 51.

Promolden calculations

The IQA decomposition of molecular energies at the HF-D3/cc-pVTZ level was performed with a modular version of the PROMOLDEN program[52] that is being developed in our laboratory. As previously noticed, the pairwise dispersion energies computed with the DFTD3 program[53] are combined with the various IQA terms ($E_{\text{net}}\left(\Omega_A\right)$, $E_{\text{int}}\left(\Omega_A,\Omega_B\right)$, …) to formulate the IQA-D3 energy decomposition of the corresponding HF-D3 energies (*i.e.*, three-body dipole–dipole–dipole dispersion energy is not included). The IQA quantities are numerically integrated over the atomic basins Ω_A, which constitute finite and irregular integration domains, using very large angular and radial grids in atomic spherical quadratures. We adopted integration settings that represent a compromise choice between computational cost and accuracy for small and medium-sized molecules. Thus, a β–sphere around each atom

was considered (*i.e.*, a sphere completely contained inside the atomic basin), with a radius equal to 60 % the distance of its nucleus to the closest bond critical point in the electron density. High-quality Lebedev angular grids were used with 5810 and 974 points outside and within the β-spheres of heavy atoms, respectively, (3890 and 590 points for hydrogen atoms). Euler-McLaurin radial quadratures were employed with 512 and 384 radial points outside and inside the β–spheres of heavy atoms, respectively (384 and 256 points for H). The largest value of the radial coordinate in the integrations was 15.0 au for heavy atoms (10.0 au for H atoms). Maximum angular moments, λ_{max}, of 10 and 6 were assigned to the Laplace and bipolar expansions of the $1/r_{12}$ operator outside and within the β-spheres.

Most of the IQA calculations reported in this work were computed with the conventional $\mathcal{O}(N^4)$ algorithm implemented in PROMOLDEN that employs the N occupied canonical molecular orbitals (MO) to expand the first and second-order density matrices. For the **14a/14b** compounds that have 51 atoms, the $\mathcal{O}(N^4)$ algorithm is exceedingly expensive and, therefore, we used a variant that uses localized MOs and employs the multipolar approach[54] for computing selected interatomic exchange-correlation (xc) energies. The LMOs were computed with the Pipek-Mezey algorithm[55] as implemented in the ORCA 4.0.1 package. For each atomic basin Ω_A, a subset of LMOs $\left\{\phi_i^{LMO}\right\}_A$ is then built by requiring that their diagonal contribution to the atomic overlap matrix ($\int_{\Omega_A}\left|\phi_i^{LMO}\right|^2 d\tau$) is greater than 10^{-6} au. The calculation of the IQA $E_{net}\left(\Omega_A\right)$ terms are done using the corresponding subset $\left\{\phi_i^{LMO}\right\}_A$ for each basin. For the calculation of the diatomic $E_{int}\left(\Omega_A,\Omega_B\right)$ terms, the LMO sets of the pair of basins are combined as the union $\left\{\phi_i^{LMO}\right\}_A \cup \left\{\phi_j^{LMO}\right\}_B$ or intersection $\left\{\phi_i^{LMO}\right\}_A \cap \left\{\phi_j^{LMO}\right\}_B$ in order to integrate the Coulombic or exchange-correlation interactions, respectively. The multipolar *xc* approximation at high order of the angular momentum series ($L=10$) is activated for 1-n ($n>4$) intramolecular interactions provided that the interatomic R_{AB} distance is greater than 5.0 au. For R_{AB} >17 au, the $E_{int}^{xc}\left(\Omega_A,\Omega_B\right)$ values are neglected. The goodness of these additional approximations was tested by comparing the results of conventional (MO-based) and LMO-based IQA calculations on the conformers of 2,3-difluoro-*N*-methylpropanamide and 2-(2,3-difluoropropyl)isoindoline-1,3-dione (**12** and **13** in Scheme 1). The IQA error defined as $\left|E-E^{IQA}\right|$, where E is the total energy obtained from the QM calculations and E^{IQA} is the total

energy reconstructed from the IQA terms, had average values of 0.5 kcal/mol (**12**) and 0.9 kcal/mol (**13**) for the MO-based IQA calculations. The corresponding values for the LMO/multipolar-based IQA calculations were similar: 0.5 kcal/mol (**12**) and 1.2 kcal/mol (**13**). Hence, the additional numerical error due to the LMO and multipolar approximations is expected to be small without compromising the conclusions of the IQA analysis.

Results

As above mentioned, 1,2-difluoroethane is usually considered as the reference system to analyze the *gauche*/*anti* conformational preference associated to adjacent fluorine atoms. For this reason, initial models for the *gauche* (F-C-C-F 60º) and *anti* (F-C-C-F 180º) conformers of 1,2-difluoroethane were optimized in the gas phase at the HF-D3/cc-pVTZ and MP2/cc-pVTZ levels of theory. Both geometry optimizations provided very similar structures, with the largest differences in the geometries located in the C-F bond lengths (see Figure 1). Relative energies were further reevaluated by means of DLPNO-CCSD(T)/aug-cc-pVTZ single point calculations performed on the MP2 geometries (see Figure 1), the *gauche* conformer being 0.7 kcal/mol more stable than the *anti* one in agreement with previous results (unless otherwise noticed ΔE values in the text correspond to DLPNO-CCSD(T) data).

Clearly, the magnitude of the *gauche* effect in 1,2-difluoroethane seems too scarce to allow the design of molecules with a frozen conformation. Hence, it is interesting to examine how the *gauche* preference can be modulated in other systems. Actually, replacing one of the fluorine atoms in 1,2-difluoroethane for other functional groups could drive away the energy of the *gauche* and *anti* conformers, increasing the impact of a single C-F bond in selecting a particular conformation. To further clarify this point, we optimized the *gauche* and *anti* conformers for a number of β-substituted α-fluoroethanes (see **1-7** in Scheme1 and Figure 1). Among the neutral molecules in Figure 1, the largest *gauche*/*anti* conformational preference is achieved for acetamide with 1.6 kcal/mol favoring the *gauche* conformation. A more pronounced *gauche* preference is induced by a positive charge adjacent to the fluorine atom. Thus, *gauche* structures with either ammonium or pyridinium groups are 6.6 and 4.2 kcal/mol, respectively, more stable than the corresponding *anti* ones.

Figure 1. Ball-and-stick representation of the gauche/anti conformers optimized for different CH_2F-CH_2-X systems. The C-F and C-C bond distances (Å), the F-C-C-X dihedral angle (°), and selected intramolecular distances (Å) measured in the HF-D3/cc-pVTZ optimized structures (RI-MP2/cc-pVTZ in parentheses) are shown. Energy differences ($E_{gauche}-E_{anti}$) in kcal/mol computed at the HF-D3/cc-pVTZ, RI-MP2/cc-pVTZ (in parentheses) and DLPNO-CCSD(T)/aug-cc-pVTZ [in brackets] levels of theory are also included.

1: 1,2-difluoroethane $\quad\quad\quad \Delta E$= -0.28 (-0.76) [-0.75] **2:** 2-fluoroethyl-acetate $\quad\quad\quad \Delta E$= -0.52 (-0.85) [-0.84]

gauche $\quad\quad\quad\quad\quad\quad$ *anti* $\quad\quad\quad\quad\quad\quad\quad\quad$ gauche

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ anti

3: 3-fluoropropanal $\quad\quad\quad \Delta E$= -0.91 (-0.65) [-0.68] **4:** *N*-(2-fluoroethyl)acetamide $\quad \Delta E$= -1.50 (-1.59) [-1.65]

gauche $\quad\quad\quad\quad\quad\quad\quad\quad\quad$ *anti* $\quad\quad\quad\quad\quad\quad$ gauche

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ anti

5: 2-(2-fluoroethyl)isoindoline-1,3-dione $\quad\quad\quad\quad\quad \Delta E$= -0.28 (-0.46) [-0.42]

gauche

6: 2-fluoroethan-1-ammonium $\quad \Delta E$= -6.53 (-6.63) [-6.56] **7:** 1-(2-fluoroethyl)pyridin-1-ium $\quad \Delta E$= -4.16 (-4.30) [-4.21]

anti

gauche $\quad\quad\quad\quad\quad\quad\quad\quad\quad$ *anti* $\quad\quad\quad\quad\quad\quad$ gauche $\quad\quad\quad\quad\quad\quad\quad\quad$ anti

When a carbonyl group is placed adjacent to a C-F bond as in α-fluoroamides, the *trans*-planar arrangement of the F-C-C=O moiety is energetically favored over the *cis* one. This is observed in Figure 2 for a ketone (**8**) and for an amide (**10**). In contrast, we computed an almost negligible *cis/trans* conformational preference for the ester (**9**).

Figure 2. Ball-and-stick representation of the *cis/trans* conformers optimized for different CH$_2$F-CO-X systems. The C-F and C-C distances (Å), the F-C-C=O dihedral angle (º) and selected intramolecular distances (Å) measured in the HF-D3/cc-pVTZ optimized structures (RI-MP2/cc-pVTZ in parentheses) are shown. Energy differences (E_{trans}–E_{cis}) in kcal/mol computed at the HF-D3/cc-pVTZ, RI-MP2/cc-pVTZ (in parentheses) and DLPNO-CCSD(T)/aug-cc-pVTZ (in squared brackets) levels of theory are also included.

8: 1-fluoropropan-2-one ΔE= -2.79 (-2.06) [-2.09]

cis   *trans*

9: methyl-2-fluoroacetate ΔE= -0.09 (0.04) [0.11]

cis

trans

10: 2-fluoro-*N*-methylacetamide ΔE= 5.80 (5.81) [5.64]

cis

trans

cis   *trans*

To analyze the origin of these energetic preferences, we carried out the IQA partitioning of the conformational energy differences at the HF-D3/cc-pVTZ level. In this respect, we note that relative energies in Figures 1 and 2 confirm that the HF-D3 ΔE values remain reasonably close to the DLPNO-CCSD(T) ones, the mean unsigned difference between them being 0.2 kcal/mol. Hence, we believe that this favorable comparison validates the use of the HF-D3 method in the subsequent IQA analyses.

Atomic and fragment-based IQA analysis of gauche/anti CH₂F-CH₂F

As described in Methods, IQA[56] decomposes molecular energy differences into a collection of atomic and interatomic terms. For 1,2-difluoroethane, the combination of the IQA quantities including the small dispersion contribution, results in an IQA reconstructed $\Delta E = E_{gauche} - E_{anti}$ energy difference of −0.5 kcal/mol, which is close to the HF-D3/cc-pVTZ reference value (−0.3 kcal/mol). This discrepancy is due to the numerical errors accumulated during the calculation of the IQA quantities. As the individual IQA terms are typically one or two orders of magnitude larger than ΔE, their expected relative error would be small.

Inspection of the IQA energies shows that the largest differences between the *gauche* and *anti* conformers arise in the electrostatic interactions (see Table S2). Thus, the *gauche* conformation is largely favored by the $\Delta E_{\text{int},class}$ term corresponding to the $C_1 \cdots F_6$ (-8.0 kcal/mol) and $F_2 \cdots C_5$ (-7.9 kcal/mol) electrostatic contacts. However, this *gauche* stabilization is widely compensated by the repulsive $F_2 \cdots F_6$ (12.1 kcal/mol) and $C_1 \cdots C_5$ (6.9 kcal/mol) electrostatic terms. Moreover, the overall *gauche* preference is also affected by other atomic and diatomic IQA terms, as the $C_1 \cdots C_5$ exchange-correlation interaction (-3.2 kcal/mol), the F_2, F_6, and C_1 electronic distortions (ΔE_{net} = 1.1, 1.0 and 1.0 kcal/mol, respectively), etc. By summing over all the atomic and diatomic contributions (see Table S1), it arises that both the classical electrostatic ($\Delta E_{int,class}$=−2.3 kcal/mol) and the exchange-correlation ($\Delta E_{int,xc}$=−1.9 kcal/mol) interactions favor the *gauche* conformation in 1,2-difluoroethane.

By considering the 1,2-difluoroethane molecule as formed by two interacting quantum fragments (IQF), the *gauche*/*anti* conformational energy difference can be split into intra-fragment deformation terms and electrostatic and quantum-mechanical inter-fragment contributions. Due to the symmetric character of the two CH₂F groups, there is no net charge transfer between them so that ΔE_{net}^{P} and ΔE_{net}^{Q} collect the energetic impact of electronic reorganization within P and Q and of a minimal steric hindrance. The computed net energy differences for each CH₂F group are 0.6 and 0.4 kcal/mol (these distinct values are again due to IQA numerical errors), but the overall deformation (1.0 kcal/mol) favoring the *anti* orientation is clearly smaller than the inter-fragment energy changes (see Table 1). Thus, the largest contribution to the stabilization of the *gauche* conformer comes from the non-classical exchange-correlation interaction (-3.5 kcal/mol) between the CH₂F groups. In contrast, classical electrostatic effects clearly stabilize the *anti* conformer (2.1 kcal/mol), which is usually explained in terms of the repulsion arising from the nearly aligned C-F dipole moments

in the *gauche* orientation (*i.e.,* F-C-C-F dihedral ~60°). This interpretation seems partly correct because, according to our analyses, the dipole-dipole interaction accounts for 55% of the difference in the $CH_2F\cdots CH_2F$ electrostatic interaction between the *gauche*/*anti* conformers so that higher-order multipolar contributions would also be important. Anyway, it is the $\Delta E_{int,xc}^{PQ} + \Delta E_{int,class}^{PQ}$ sum (-1.4 kcal/mol) that mainly controls the stability of the gauche conformer and, therefore, the IQF energy decomposition for 1,2-difluoroethane resembles those in previous proposals about the combined role of hyperconjugation and electrostatics to explain the *gauche* effect.[23] Within the context of IQF, we propose then to measure the *gauche* stabilization in 1,2-difluoroethane and related systems as $\Delta E_{int,xc}^{PQ} + \Delta E_{int,class}^{PQ}$.

Probably, either the IQA or the IQF analysis of the *gauche/anti* 1,2-difluoroethane may be adequate and useful. Nonetheless, chemists usually explain molecular properties in terms of the functional groups or the characteristic moieties that build up a molecule. For large systems, grouping the atomic terms into fragments simplifies the interpretation of the IQA data because the IQF terms smooth out the correlated fluctuations of the atomic quantities within a given group. Therefore, we decided to focus on the IQF results for the rest of the systems studied in this work, although the underlying atomic partitioning will also be used to better characterize some relevant inter- or intra-fragment contributions.

Table 1. IQF energy components at the HF-D3/cc-pVTZ level for the energy difference (kcal/mol) between the two conformers (*gauche/anti* or *cis/trans*) analyzed for a series of small systems related to 1,2-difluoroethane. Two-fragment partitioning scheme (P=CH₂F and Q=CH₂X) is assumed for the different systems.

$\Delta E = E_{gauche} - E_{anti}$	Δq^P	ΔE_{net}^P	ΔE_{net}^Q	$\Delta E_{int,disp}$	$\Delta E_{int,xc}$	$\Delta E_{int,class}$	$\Delta E_{int,class,qq}$	$\Delta E_{int,class,qd}$	$\Delta E_{int,class,dd}$	ΔE_{IQA}
1: 1,2-difluoroethane	0.00	0.6	0.4	-0.0	-3.5	2.1	0.0	0.0	1.1	-0.5
2: 2-fluoroethyl-acetate	0.00	0.7	0.5	-0.0	-3.1	1.4	0.0	0.0 0.0	-0.7	-0.5
3: 3-fluoropropanal	-0.01	0.0	1.4	-0.0	-1.1	-1.5	-0.1	-0.4 0.0	-0.4	-1.2
4: *N*-(2-fluoroethyl)acetamide	-0.02	-0.6	2.1	-0.2	-2.1	-1.4	-0.2	-0.6 0.1	-1.0	-2.1
5: 2-(2-fluoroethyl)isoindoline-1,3-dione	0.00	2.0	1.5	-0.3	-3.6	-0.1	0.0	-0.0 -0.0	0.1	-0.5
6: 2-fluoroethxan-1-aminium	-0.01	1.6	4.7	-0.6	-6.9	-6.0	2.4	1.3 -3.5	-4.6	-7.3
7: 1-(2-fluoroethyl)pyridin-1-ium	-0.01	1.0	3.5	-0.3	-5.5	-3.7	0.9	0.2 -6.1	-0.3	-5.1
$\Delta E = E_{trans} - E_{cis}$	Δq^P	ΔE_{net}^P	ΔE_{net}^Q	$\Delta E_{int,disp}$	$\Delta E_{int,xc}$	$\Delta E_{int,class}$	$\Delta E_{int,class,qq}$	$\Delta E_{int,class,qd}$	$\Delta E_{int,class,dd}$	ΔE_{IQA}
8: 1-fluoropropan-2-one	0.01	0.1	0.6	0.0	1.5	-4.9	0.0	-0.1 0.1	-2.9	-2.8
9: methyl-2-fluoroacetate	0.00	0.6	0.5	0.0	0.3	-1.4	0.0	0.7 0.2	-2.8	-0.1
10: 2-fluoro-*N*-methylacetamide	-0.02	2.1	2.7	-0.2	-4.1	-6.3	0.0	0.2 0.1	-5.2	-5.8

Substituent effects on the gauche/anti CH₂F-CH₂X conformers: Competition between IQF exchange-correlation and electrostatic interactions

For the CH_2F-CH_2X molecules, the *gauche* conformer is energetically favored over the *anti* one (see Figure 1). In all the CH_2F-CH_2X molecules, the charge variation at the CH_2F / CH_2X moieties upon the *anti→gauche* rearrangement is null or minimal (~0.01 e) as computed from the corresponding QTAIM charges.

Regarding 1,2-difluoroethane as the reference compound, we observe that replacing the second fluorine atom by an acetate group (**2** in Scheme 1) results in very small changes in the IQF terms. Again the *gauche* conformer is stabilized by the exchange correlation interaction ($\Delta E_{int,xc}$=−3.1 kcal/mol). Electrostatics favors the *anti* geometry ($\Delta E_{int,class}$=1.4 kcal/mol), but such effect does not arise from favorable μ-μ interactions ($\Delta E_{int,class,dd}$=−0.7 kcal/mol). This is due to the internal geometry of the acetate group, in which the dipole moments along the C=O/C-O bonds combine in a total fragment dipole moment (~1.9 D) that is nearly orthogonal to the CH_2F dipole (~2.2 D) in the *anti* conformer.

Introduction of an aldehyde (**3**) or an amide (**4**) group leads to classical and exchange-correlation inter-fragment interactions that stabilize the *gauche* conformer, being partially compensated by the unfavorable net energy change of the CH_2X fragment (see Table 1). There are, however, some differences between the aldehyde and the amide groups. The largest inter-atomic electrostatic interaction involves the fluorine atom in the two systems, but the F_2···C_8 interaction in the aldehyde **3** stabilizes the *gauche* structure (-15.9 kcal/mol), meanwhile the equivalent F_2···N_6 contact stabilizes the *anti* conformation in **4** (27.9 kcal/mol). If we compare the electrostatic F_2···N_6 interaction in the amide with F_2···O_6 in **2** (20.8 kcal/mol) and F_2···F_6 in **1** (12.1 kcal/mol), it arises an inverse relationship between the electronegativity of the X atom in the F-C-C-X sequence and the penalty for the *gauche* orientation due to the F···X $\Delta E_{int,class}$ electrostatic term.

In contrast with the **1-4** systems, the small *gauche* stabilization computed for the isoindole derivative **5** arises from a large and favorable exchange-correlation term (-3.6 kcal/mol), mainly ascribed to the H_3···O_{19} and F_2···N_{20} interactions (see Table S2). But this favorable $\Delta E_{int,xc}$ term is cancelled by the intra-fragment net energies (2.0 and 1.5 kcal/mol). The change in the inter-fragment electrostatic interaction is almost negligible (-0.1 kcal/mol) due to the cancelation of large inter-atomic electrostatic interactions between fluorine and the heteroatoms in the isoindole moiety.

A cationic group adjacent to a C-F bond significantly stabilizes the *gauche* conformation, what is commonly explained in terms electrostatics. In effect, the $\Delta E_{int,class}$ terms in the *gauche–anti* energy difference in $CH_2F\text{-}CH_2X^+$ become quite important: -6.0 and -3.7 kcal/mol for the ammonium (**6**) and pyridinium (**7**) groups. These two cationic substituents have different size and charge distribution, what is reflected in their μ^Q dipole moments, ~2.5 D and ~0.2/0.5 D for the ammonium and pyridinium CH_2X^+ groups, respectively. The global positive charge is delocalized between the *P* and *Q* fragments (*e.g.*, q^P=0.12 and q^Q= 0.88 in the gauche **6** conformer) and these q^P/ q^Q charges repel more strongly in the *gauche* orientation. Thus, the charge-dipole and dipole-dipole fragment interactions explain the electrostatic preference for the *gauche* location of the ammonium group (see Table 1), while the rather small μ^Q value of pyridinum implies that only the interaction between the bulkier pyridinium charge and the CH_2F dipole makes a relevant contribution to the *gauche* stability (-6.1 kcal/mol). As in the $CH_2F\text{-}CH_2X$ neutral systems, the change in the fragment net energies stabilizes the *anti* orientation in the cationic systems albeit with a more pronounced influence (*e.g.*, ΔE_{net}^Q=4.7 and 3.5 kcal/mol for **6** and **7**). However, the IQF data reveals that the exchange-correlation inter-fragment interactions reinforce the stability of the *gauche* orientation (see Table 1) and compensate the deformation energies. More particularly, in the *gauche* **6** structure, a short F⋯H-N contact rationalizes well the significant $\Delta E_{int,xc}^{PQ}$ value of -6.9 kcal/mol (see Table S2 and Figure 1).

IQF description of trans/cis conformers in CH₂F-(C=O)X: Electrostatic control

When a carbonyl group is placed adjacent to a fluorine atom, electrostatic effects are assumed to stabilize the *trans*-planar orientation of the C=O and C-F groups over the *cis* one (see Figure 2). Our IQF calculations on the α-fluoroketone (**8**), α-fluoroester (**9**), and α-fluoroamide (**10**) molecules confirm that the electrostatic interaction between the CH_2F and (C=O)X moieties is the key element that explains the larger stability of the C-F/C=O *trans* arrangement ($\Delta E_{int,class}$ = -4.9, -1.4 and -6.2 kcal/mol for **8**, **9** and **10**; see Table 1). Furthermore, this electrostatic effect is mainly associated to the dipole-dipole interaction term in the three cases, although contributions from higher multipoles seem also relevant for the α-fluoroketone **8**. The exchange-correlation term also contributes to the stability of the *trans* fluoroamide (-4.1 kcal/mol), whereas it favors the *cis* orientation in the fluoroketone (1.5

kcal/mol) and has little effect on the α-fluoroester. These $\Delta E_{int,\text{xc}}$ contributions mainly result from the $F_2\cdots O_6$ (4.5 kcal/mol)/$F_2\cdots C_7$ (-3.1) interactions in the *cis/trans* α-fluoroketone **8**, and the $F_2\cdots N_7$ (-5.3)/$F_2\cdots O_6$ (4.7)/$F_2\cdots H_8$ (-3.0) contacts in the α-fluoroamide **10**. Nonetheless, the favorable F⋯H-N polar contact in **10** is nearly compensated by significant fragment distortion (see ΔE_{net}^{P} and ΔE_{net}^{Q} in Table 1) so that the electrostatic $\Delta E_{int,class}$ is the major stabilizing energy contribution to the *trans* preference between the vicinal C-F and C=O groups.

−CF−CF−CF− motif: IQA discrimination among gauche effects, 1,3 dipole-dipole and 1,3 CH⋯F contacts

For large molecules bearing various fluorine atoms as well as other functional groups, several fluorine-associated effects may control their conformational preferences. For instance, the preferred conformations in polyfluorinated alkanes have been assumed to arise from maximizing *gauche* interactions between vicinal fluorine atoms and, simultaneously, minimizing unfavorable dipole-dipole interactions between 1,3-difluoro motifs.[9, 13] To further analyze these effects, we optimized the nine conformers of 1,2,3-trifluoropropane (**11**) generated by the *gauche g+* (~ 60°), the gauche *g-* (~ −60°) or the *anti* (~ 180°) orientations of the two consecutive F-C-C-F torsions. All the conformers remain within a narrow energy range of ~3 kcal/mol. In addition, some of the structures correspond to equally-stable conformational enantiomers (pair *g+/g+* and *g-/g-*, pair *g+/anti* and *anti/g-*, and pair *g-/anti* and *anti/g+*) and, accordingly, only the first conformer within each pair was considered for the IQA analysis. Since the relative energies for the **1-7** compounds are given as $\Delta E = E_{gauche} - E_{anti}$, the relative energies for CH_2F-CHF-CH_2F given in the text and in Figure 3 are similarly expressed as $\Delta E = E_{g+/g+} - E_i$ where i stands for any other conformer of **11**. Hence, a negative ΔE value means that conformer i is less stable than the *g+/g+* one.

Figure 3. Ball-and-stick representation of the different conformers obtained for 1,2,3-trifluoropropane (**11** in Scheme 1). F-C-C-F dihedral angles (°) and selected interatomic distances (Å) measured in the HF-D3/cc-pVTZ optimized structures (RI-MP2/cc-pVTZ in parentheses) are shown. Energy differences ($E_{g+/g+} - E_i$) in kcal/mol computed at the HF-D3/cc-pVTZ, RI-MP2/cc-pVTZ (in parentheses) and DLPNO-CCSD(T)/aug-cc-pVTZ [in brackets] levels of theory are also included.

g+/g+
ΔE= 0.00 (0.00) [0.00]

g+/g-
ΔE= -1.18 (-1.23) [-1.17]

g+/anti
ΔE= 0.17 (-0.14) [-0.18]

g-/g+
ΔE= -3.06 (-2.82) [-2.90]

g-/anti
ΔE= 0.74 (0.50) [0.39]

anti/anti
ΔE= -2.07 (-2.53) [-2.59]

Table 2 collects the IQF energy decomposition of the relative energies with respect to the reference *g+/g+* structure considering three molecular fragments (*i.e.*, P=CH$_2$F-, Q=-CHF-, and R=-CH$_2$F). In line with previous proposals, the parallel alignment of the C$_1$-F$_2$ and C$_8$-F$_9$ bonds in the less stable *g-/g+* and *anti/anti* conformers is associated to unfavorable electrostatic interactions between the corresponding CH$_2$F groups ($\Delta E_{int,class}^{P\cdots R}$ of -2.4 and -2.1 kcal/mol, respectively), which are mainly ascribed to dipole-dipole repulsion ($\Delta E_{int,class,dd}^{P\cdots R}$ of -2.0 and -2.1 kcal/mol for *g-/g+* and *anti/anti*). Hence, it seems reasonable to describe them as "1,3-dipole repulsions" in consonance with former suggestions.[9]

Table 2. IQF relative energy components (in kcal/mol) at the HF-D3/cc-pVTZ level for the energetically different conformers of 1,2,3-trifluoropropane (**11**). A three-fragment partitioning scheme (P=CH$_2$F, Q=CHF, and R=CH$_2$F) is assumed for the different structures. Energy differences are computed as $E_{g+/g+} - E_i$, so that a negative value means further stabilization of $g+/g+$.

	ΔE_{net}^P	ΔE_{net}^Q	ΔE_{net}^R	$\Delta E_{int,disp}$			$\Delta E_{int,xc}$			$\Delta E_{int,class}$			ΔE_{IQA}	$\Delta E_{HF\text{-}D3}$
				$P{\cdots}Q$	$P{\cdots}R$	$Q{\cdots}R$	$P{\cdots}Q$	$P{\cdots}R$	$Q{\cdots}R$	$P{\cdots}Q$	$P{\cdots}R$	$Q{\cdots}R$		
$\Delta E = E_{g+/g+} - E_{g+/g\text{-}}$	2.4	-1.9	0.6	-0.0	-0.1	0.0	0.6	-1.4	0.1	-0.3	-0.6	-0.9	-1.5	-1.2
$\Delta E = E_{g+/g+} - E_{g+/anti}$	0.7	-0.7	0.8	-0.0	-0.0	0.0	1.1	-0.2	-2.3	-0.4	-0.1	1.5	0.4	0.2
$\Delta E = E_{g+/g+} - E_{g\text{-}/g+}$	0.6	0.1	-1.4	0.0	-0.3	0.0	0.1	-0.7	-0.2	0.8	-2.4	0.2	-3.2	-3.1
$\Delta E = E_{g+/g+} - E_{g\text{-}/anti}$	0.4	1.3	-1.7	-0.0	0.0	0.0	0.5	1.0	-2.7	0.2	0.1	1.6	0.8	0.7
$\Delta E = E_{g+/g+} - E_{anti/anti}$	1.7	0.3	-0.5	-0.0	-0.3	0.0	-1.8	-1.1	-2.1	2.1	-2.1	1.6	-2.1	-2.1

The *gauche* effect is also expected to play a key role in determining the conformational landscape of 1,2,3-trifluoropropane. However, the missing *gauche* effects between vicinal fluorine atoms do not penalize the *anti/anti* structure as compared to the *g-/g+* one (*i.e.*, *anti/anti* is 1.0 kcal/mol more stable than *g-/g+* at HF-D3/cc-pVTZ). The lack of a net *gauche* effect in the *anti/anti* structure is revealed by IQF because the $\Delta E_{int,xc}$ terms between consecutive fragments favoring the *gauche* arrangement (-1.8 for $P{\cdots}Q$ and -2.1 kcal/mol for $Q{\cdots}R$) are nearly compensated by electrostatic interactions ($\Delta E_{int,class}$= 2.1 and 1.6 kcal/mol for $P{\cdots}Q$ and $Q{\cdots}R$) favoring the *anti* conformation. Other differences between the *g-/g+* and *anti/anti* structures arise in the intra-fragment net energies, that disfavor the *g-/g+* structure (ΔE_{net}^R =-1.4 kcal/mol) though they stabilize the *anti/anti* one (ΔE_{net}^P =1.7 kcal/mol).

The *g+/anti* and *g-/anti* conformers, with only one *gauche* arrangement, are nearly isoenergetic (*g+/anti*) or slightly more stable (*g-/anti*) than the reference *g+/g+* conformer presenting two *gauche* effects (see Figure 3). IQF confirms that the *gauche* effect at F$_6$-C$_5$-C$_8$-F$_9$ favors *g+/g+* over *g+/anti* and *g-/anti,* because the exchange-correlation $\Delta E_{int,xc}^{Q{\cdots}R}$ term (-2.3 and -2.7 kcal/mol for *g+/anti* and *g-/anti*, respectively*)* is greater in absolute value than the electrostatic $\Delta E_{int,class}^{Q{\cdots}R}$ term (1.5 kcal/mol for *g+/anti* and 1.6 kcal/mol for *g-/anti*). The same comparison involving the $\Delta E_{int,xc}^{P{\cdots}Q}$ and $\Delta E_{int,class}^{P{\cdots}Q}$ terms suggest that the *g+/anti* conformer may exhibit a reinforced gauche preference in the F$_2$-C$_1$-C$_5$-F$_6$ angle.

Inspection of the optimized geometries in Figure 3 suggests that C-H$\cdots$F contacts can contribute to the conformational energies. As expected, the formation/loss of the C-H$\cdots$F contacts can be traced back to modifications in either the deformation or the interaction IQA components involving the C-H$\cdots$F atoms (Table S4). The best fingerprint of the C-H$\cdots$F

contact is provided by the exchange-correlation $\Delta E_{\text{int},xc}^{AB}$ energy between the F and H atoms because it scores among the largest $\Delta E_{\text{int},xc}^{AB}$ terms and correlates well with the IQF exchange-correlation energy. For example, the reference $g+/g+$ structure has an $H_4 \cdots F_9$ contact (2.57 Å) which is replaced by a similar $H_3 \cdots F_9$ interaction in the $g+/anti$ conformer (2.53 Å), and this change is linked to significant interatomic $\Delta E_{\text{int},xc}^{AB}$ contributions (-1.6 kcal/mol for $H_4 \cdots F_9$ and 1.9 kcal/mol for $H_3 \cdots F_9$ in Table S4). In the case of $g-/anti$, the $H_4 \cdots F_9$ interaction in $g+/g+$ is preserved and one additional $F_2 \cdots H_{10}$ contact (2.49 Å) is formed that results in a favorable interatomic $F_2 \cdots H_{10}$ $\Delta E_{int,xc}$ contribution (2.3 kcal/mol in Table S4). Furthermore, the lack of 1,3 C-H$\cdots$F contacts in the $g+/g-$ structure (*e.g.* $\Delta E_{int,xc}$ = -1.7 kcal/mol for $H_4 \cdots F_9$) can be invoked to explain its lower stability (1.2 kcal/mol with respect to $g+/g+$) in spite of maintaining two *gauche* effects. Hence, we conclude that the conformational properties of 1,2,3-trifluoropropane are more significantly influenced by through-space exchange-correlation interactions between fluorine and hydrogen atoms than the *gauche* effect between vicinal fluorine atoms, and that the IQA/IQF signature of these contacts is conveniently described in terms of the exchange-correlation interaction energies.

IQF analysis of fragments from an α,β-difluoro-γ-amino-acid

To help ascertain all the effects determining the conformational properties of the α,β-difluoro-γ-amino-acids **14a/14b** containing the -CHF-CHF- motif, we first examined compound **12** formally derived from **11** by replacing one fluorine atom by a polar amide group (see Scheme 1). Either a *gauche* or an *anti* orientation in the adjacent C-F bonds and a *cis* or a *trans* arrangement for the vicinal C-F and C=O bonds can appear in **12** (see Figure 4). Only five conformers were located on the HF and MP2 potential energy surfaces (*g-/cis* turned out to be unstable). Their relative energies given in Figure 4 are expressed as $\Delta E = E_{g+/trans} - E_i$ where i stands for any other conformer of **12**.

The positioning of the –(C=O)NHCH$_3$ group in **12** results in a strong conformational selection in favor of the *g+/trans* or *g-/trans* conformers, which is unequivocally interpreted by the three-fragment (P=CH$_2$F, Q=CHF, and R=CONHCH$_3$) IQF partitioning. As already observed in **10**, the *cis* alignment in the adjacent C$_5$-F$_6$ and C$_8$=O$_9$ bonds in **12** is clearly destabilized by the classical electrostatic interaction between fragments Q and R. The

corresponding $\Delta E_{\text{int,class}}^{Q\cdots R}$ values are -7.5 and -8.4 for the *g+/cis* and *anti/cis* conformers, the exchange-correlation $\Delta E_{\text{int,xc}}^{Q\cdots R}$ being also unfavorable (see Table 3). Some differences appear in the $P\cdots Q$ and $P\cdots R$ interaction terms that contribute to the relative stability of the *g+/cis* and *anti/cis* conformers. Thus, the lack of *gauche* effect between vicinal C-F bonds makes *anti/cis* less stable than *g+/trans* (*e.g.*, $\Delta E_{\text{int,xc}}^{P\cdots Q} + \Delta E_{\text{int,class}}^{P\cdots Q}$ =-2.8 + 1.5= -1.3 kcal/mol). On the other hand, the terminal $P\cdots R$ interaction through a $F_2\cdots H_{11}\text{-}N_{10}$ contact favors *anti/cis* ($\Delta E_{\text{int,xc}}^{P\cdots R}$ =3.0 kcal/mol and $\Delta E_{\text{int,class}}^{P\cdots R}$ =3.4 kcal/mol) over *g+/cis* ($\Delta E_{\text{int,xc}}^{P\cdots R}$ =0.4 kcal/mol and $\Delta E_{\text{int,class}}^{P\cdots R}$ =-2.0 kcal/mol). With respect to the electrostatic terms, the partial alignment of the $C_1\text{-}F_2$ and amide dipole moments contributes to further destabilize *g+/cis* (the $P\cdots Q$ dipole-dipole interaction amounts to -1.7 kcal/mol).

Figure 4. Ball-and-stick representation of the five conformers optimized for 2,3-difluoro-*N*-methylpropanamide (**12** in Scheme 1). F-C-C-F and F-C-C-C=O dihedral angles (º) and selected interatomic distances (Å) measured in the HF-D3/cc-pVTZ optimized structures (RI-MP2/cc-pVTZ in parentheses) are included. Energy differences ($E_{g+/trans} - E_i$) in kcal/mol computed at the HF-D3/cc-pVTZ, RI-MP2/cc-pVTZ (in parentheses) and DLPNO-CCSD(T)/aug-cc-pVTZ [in brackets] levels of theory are also included.

g+/trans ΔE= 0.00 (0.00) [0.00]   **g-/trans** ΔE= -0.25 (-0.18) [-0.26]   **anti/trans** ΔE= -4.18 (-4.04) [-3.95]

g+/cis ΔE= -8.05 (-7.61) [-7.51]   **anti/cis** ΔE= -5.52 (-5.34) [-5.18]

The magnitude of the relative energy of the *anti/trans* conformer of **12** (4.2 kcal/mol) as compared to the *g+/trans* structure is difficult to explain only in terms of the small *gauche* effect estimated as $\Delta E_{\text{int,xc}}^{P\cdots Q} + \Delta E_{\text{int,class}}^{P\cdots Q}$ =-3.4 + 2.9= -0.5 kcal/mol. In contrast, the terminal $P\cdots R$ interactions, $\Delta E_{\text{int,xc}}^{P\cdots R}$ =-1.6 kcal/mol and $\Delta E_{\text{int,class}}^{P\cdots R}$ =-2.6 kcal/mol, make a more significant contribution to the destabilization of *anti/trans*. These terms can be associated with the loss of the $H_3\cdots O_9$ contact ($\Delta E_{\text{int,xc}}^{H_3\cdots O_9}$ =-2.1 kcal/mol in Table S5) when comparing *g+/trans* with *anti/trans*, while the electrostatic contribution can be connected to the partial alignment of the P and R dipole moments in the *anti/trans* structure (the $P\cdots R$ dipole-dipole interaction amounts to -2.6 kcal/mol). Thus, we see again that either 1,3 dipole-dipole interactions or O/F···H contacts can be more determining than the *gauche* effect between adjacent C-F groups.

Table 3. IQF energy components at the HF-D3/cc-pVTZ level for the energy difference (kcal/mol) among the five conformers optimized for 2,3-difluoro-*N*-methylpropanamide (**12**). A three-fragment partitioning scheme (P=CH$_2$F-, Q=-CHF-, and R=-CONHCH$_3$) is assumed for the different conformers. Energy differences are computed as $E_{g+/trans} - E_i$, so that a negative value means further stabilization of *g+/trans*.

	ΔE_{net}^{P}	ΔE_{net}^{Q}	ΔE_{net}^{R}	$\Delta E_{int,disp}$			$\Delta E_{int,xc}$			$\Delta E_{int,class}$			ΔE_{IQA}	ΔE_{HF-D3}
				$P\cdots Q$	$P\cdots R$	$Q\cdots R$	$P\cdots Q$	$P\cdots R$	$Q\cdots R$	$P\cdots Q$	$P\cdots R$	$Q\cdots R$		
$\Delta E = E_{g+/trans} - E_{g-/trans}$	-0.8	-1.7	0.6	0.0,	-0.1,	0.0	-0.1,	-0.4,	1.3	0.0,	0.4,	-0.0	-0.7	-0.2
$\Delta E = E_{g+/trans} - E_{anti/trans}$	0.1	-1.7	0.7	-0.0,	-0.4,	0.0	-3.4,	-1.6,	1.7	2.9,	-2.6,	-0.1	-4.3	-4.2
$\Delta E = E_{g+/trans} - E_{g+/cis}$	1.1	1.7	2.0	-0.0,	0.3,	-0.4	-0.5,	0.4,	-4.8	0.4,	-2.0,	-7.5	-9.3	-8.0
$\Delta E = E_{g+/trans} - E_{anti/cis}$	0.3	0.3	0.1	0.0,	0.6,	-0.4	-2.8,	3.0,	-3.9	1.5,	3.4,	-8.4	-6.3	-5.5

In contrast with the case of compound **12**, the positioning of the bulky isoindole-1,3-dione group –(C=O)NHCH$_3$ group in **13** has a minor conformational influence. For this reason, its analysis is reported in the Supplementary Material (Figure S1 and Tables S6-S7).

*Conformational analysis and QM calculations on the **14a** and **14b** dipeptides*

The QM and IQA calculations on the model compounds **1-13** characterize several conformational effects (*gauche/anti, cis/trans,* through space F···H, dispersion attractions, etc.) than can act simultaneously in the **14a/14b** dipeptides. To better assess these and other effects, we investigated first whether or not additional conformations to those observed in the **14a/14b** crystal structures could be accessible. Thus, we performed an automated conformational search followed by HF-D3/cc-pVTZ geometry optimizations of the resulting

conformers in the gas-phase and in the solvent continuum (see Tables S8-S9 and Figures S2-S3).

Most remarkably, inspection of the optimized geometries and relative energies of the **14a** and **14b** conformers reveals that the X-ray conformations (labelled as **14a-conf1** and **14b-conf1**) are indeed among the most stable ones in the gas-phase and in the chloroform solution. In addition, we found that the structures presenting the expected F_{23}-C_{22}-C_{19}-F_{20} and F_{20}-C_{19}-C_{16}-N_{13} *gauche* orientation and the *trans* F_{23}-C_{22}-C_{25}=O_{26} arrangement (see Figure 5 for atom numbering) are, in general, energetically favored as expected by common assumptions. There are, however, low energy conformers, especially for **14a**, that do not present the expected *gauche* orientation around the fluorine atoms (see for instance **14a-conf4** and **14a-conf5**). Moreover, it turns out that the conformer **14a-conf6** is not drastically penalized in the gas-phase despite presenting a *cis* F_{23}-C_{22}-C_{25}=O_{26} alignment and an almost *anti* orientation for the F_{23}-C_{22}-C_{19}-F_{20} and F_{20}-C_{19}-C_{16}-N_{13} bonds.

The results of our conformational search can be addressed with NMR experimental data. Thus, the NMR spectra of **14a** and **14b** have been recorded at 300 K in deuterated chloroform solution and their coupling constants determined by simulation/iteration sequences.[14] We also estimated $^3J_{HF}$ coupling constants around the FC_{22}-C_{19}F bond using a Karplus-type relation[51] for the five most stable conformers of **14a** and **14b**, and the corresponding values were Boltzmann-averaged at 300 K according to the HF-D3/cc-pVTZ relative energies in chloroform (Table S10). For **14a**, the favorable comparison between the computed and experimental values validates the results of our conformational analysis (the computational averages for $^3J_{FH}$ ($F_{23}H_{21}$) and $^3J_{HF}$ ($H_{24}F_{20}$) are 27.4 and 28.2 *Hz* compared with the experimental ones of 27.0 and 28.9 *Hz*). It is interesting to note that some conformational variability of **14a** in chloroform solution can be expected because the ΔE difference between the two most stable conformers is not large (1.0 kcal/mol). For **14b**, the agreement between the estimated coupling constants and the experimental ones is only moderate (the calculated averages for $^3J_{FH}$ and $^3J_{HF}$ are 26.1 and 13.0 *Hz* while the experimental values are 23.0 and 21.4 *Hz*). However, the computed $^3J_{FH}$ and $^3J_{HF}$ values for **14b-conf1** (30.8 and 8.1) and **14b-conf2** (6.6 and 31.7) would approach to the experimentally-derived ones (23.0 and 21.4) upon averaging, provided that these two conformers become nearly isoenergetic. Although the calculated ΔE value between **14b-conf1** and **14b-conf2** is already small (0.7 kcal/mol), it is likely that method/basis-set

improvements in the QM level of theory and/or a better description of thermal and solvation effects would be required to increase the agreement with experimental data.

*IQF assessment of the conformational effects acting upon **14a***

Our QM calculations suggest that intramolecular factors largely determine the conformational preferences of **14a/14b** as their low-energy conformers in the gas-phase are quite close to their crystallographic structures. Moreover, the intrinsic stability of the two diastereomers is similar and we expect that the same local effects influence their conformational energies. Taking also into account the large computational cost of the IQA calculations on these systems, we focused on the IQA analysis of the relative stability of selected **14a** conformers. Thus, we performed IQA calculations on **14a-conf1** and **14a-conf2**, which are two different *gauche* forms (*g+* and *g-*, respectively) that interconvert into each other through a ~120° rotation about the $F_{23}C_{22}$–$C_{19}F_{20}$ bond. We also selected conformer **14a-conf5**, which shows an *anti* conformation at F_{23}–C_{22}–C_{19}–F_{20}, and conformer **14a-conf6**, which is a helical-type compact structure presenting all *anti* (F_{23}-C_{22}-C_{19}-F_{20} and F_{20}-C_{19}-C_{16}-N_{13}) and *cis* (F_{23}-C_{22}-C_{25}=O_{26}) arrangements (see Figure 6). Table 5 summarizes the IQF decomposition of the HF-D3/cc-pVTZ relative energies involving a four-fragment O-P-Q-R partitioning in which fragments P and Q correspond to the central $C_{22}HF$ and $C_{19}HF$ units, respectively, fragment O comprises the Ace-CH(Ph)-NHCO– residue and fragment R contains the isoindole-1,3-dione moiety. The relative energies are now expressed as $\Delta E = E_{14a\text{-}conf1} - E_{14a\text{-}confi}$ where i=2, 5, or 6.

Figure 5. Ball-and-stick representation of selected conformers optimized for **14a** at the HF-D3/cc-pVTZ level of theory. F-C-C-F, F-C-C-N, and F-C-C-C=O dihedral angles (°) and selected interatomic distances (Å) measured in the HF-D3/cc-pVTZ gas phase optimized structures are included. Energy differences ($E_{14a\text{-}conf1} - E_{14a\text{-}confi}$) in kcal/mol computed at the HF-D3/cc-pVTZ level of theory in the gas-phase and in chloroform (in Italics) are also included.

14a-conf1 ΔE= 0.00 *0.00*   **14a-conf2** ΔE= 0.74 *-1.05*

14a-conf5 ΔE= -1.97 *-2.14*   **14a-conf6** ΔE= -2.64 *-4.13*

We first analyze the variation of the F_{23}-C_{22}-C_{19}-F_{20} dihedral from 67.5° in conformer **14a-conf1** to -58.4° in **14a-conf2**, affecting the relative positioning of the O and R fragments (see Figure 5). As the two conformers are presumably stabilized by a similar *gauche* effect and the placement of the terminal isoindole-1,3-dione group in R with reference to the central P-Q fragments has little influence, their energy difference in the gas-phase is quite small, 0.7 kcal/mol at the HF-D3/cc-pVTZ level favoring **14a-conf2**. Nonetheless, examination of the IQF terms collected in Table 4 reveals some fine details concerning this energy difference. On one hand, there are some indications about a more stabilizing *gauche* effect in **14a-conf2** given that $\Delta E_{\text{int,xc}}^{P\cdots Q} + \Delta E_{\text{int,class}}^{P\cdots Q}$ =1.5 - 0.9= 0.6 kcal/mol. On the other one, IQF reveals substantial energy compensation occurring upon the formation of new $O\cdots R$ contacts with simultaneous

loss of $P\cdots R$ and $Q\cdots R$ ones. Thus, the exchange-correlation interaction between fragments O and R largely stabilizes **14a-conf2** (17.6 kcal/mol, comprising specific $H_{17}\cdots O_{26}=5.0$ kcal/mol, $C_{16}\cdots O_{26}=3.1$, $O_{12}\cdots H_{42}=3.0$ and $O_{12}\cdots H_{44}=2.3$ pair interactions; see Figure 5 and Table S11). This attractive component is partly neutralized by the equivalent terms accounting for the $O\cdots Q$ and $P\cdots R$ fragment interactions, which are -7.6 and -4.4 kcal/mol, respectively, reflecting the loss of the $F_{20}\cdots H_{44}$ and $H_{21}\cdots O_{26}$ contacts (-3.2 and -2.2 kcal/mol) and the $F_{23}\cdots H_{18}$ and $F_{23}\cdots C_{16}$ ones (-2.0 and -1.8 kcal/mol). In addition, the shortening of the $O\cdots R$ distance in **14a-conf2** implies not only stabilizing inter-fragment energy contributions, but also a large intra-fragment distortion as in the case of intermolecular complexes. For example, the ΔE_{net}^{R} term destabilizes **14a-conf2** by 9.7 kcal/mol. Altogether, these IQA components allow us to energetically weigh the formation/loss of intramolecular contacts upon the **14a-conf1→14a-conf2** transition, which nearly compensate each other.

Table 4. IQF energy components at the HF-D3/cc-pVTZ level for the energy difference (kcal/mol) among selected conformers optimized for compound **14a**. A four-fragment partitioning scheme (O = $CH_3OCOCH(CH_2Ph)NHCO$-, P = -CHF-, Q = -CHF-, and R = -CH$_2$-isoindol-1,3-dione) is assumed for the different conformers. Relative energies are given as $E_{14a\text{-}conf1} - E_{14a\text{-}confi}$ so that a negative value means further stabilization of **14a-conf1**.

	ΔE_{net}^{O}	ΔE_{net}^{P}	ΔE_{net}^{Q}	ΔE_{net}^{R}	$\Delta E_{int,disp}$			$\Delta E_{int,xc}$			$\Delta E_{int,class}$			ΔE_{IQA}	$\Delta E_{HF\text{-}D3}$
					$P\cdots Q$	$P\cdots R$	$O\cdots P$	$P\cdots Q$	$P\cdots R$	$O\cdots P$	$P\cdots Q$	$P\cdots R$	$O\cdots P$		
					$Q\cdots R$	$O\cdots Q$	$O\cdots R$	$Q\cdots R$	$O\cdots Q$	$O\cdots R$	$Q\cdots R$	$O\cdots Q$	$O\cdots R$		
$\Delta E = E_{14a\text{-}conf1} - E_{14a\text{-}conf2}$	-5.5	-1.2	2.8	-9.7	0.3, -0.8, 0.1			1.5, -4.4, 1.6			-0.9, -1.1, -0.8			0.3	0.7
					0.2, -0.9, 4.5			0.2, -7.6, 17.6			0.8, -1.0, 4.7				
$\Delta E = E_{14a\text{-}conf1} - E_{14a\text{-}conf5}$	-17.5	-1.0	4.5	-19.8	0.3, -0.9, 0.1			-2.5, -4.5, 2.2			2.8, -1.2, 0.5			-2.2	-2.0
					-0.4, -1.8, 12.5			-0.5, -9.6, 35.2			-0.2, -3.2, 2.8				
$\Delta E = E_{14a\text{-}conf1} - E_{14a\text{-}conf6}$	-16.9	-3.4	3.2	-21.7	0.3, 0.7, -0.4			-1.5, 1.1, -2.4			2.5, -0.1, -6.7			-2.1	-2.6
					-0.7, -0.8, 10.8			-4.6, -4.9, 33.7			2.0, 1.1, 6.5				

More interesting conformational effects act upon **14a-conf5**, whose overall shape differs significantly with respect to those of **14a-conf1/14a-conf2**. There is an *anti* F_{23}-C_{22}-C_{19}-F_{20} arrangement in **14a-conf5,** which is 2.0 kcal/mol less stable than **14a-conf1** at HF-D3/cc-pVTZ. However, this energy difference cannot be assigned to the loss of the *gauche* F_{23}-C_{22}-C_{19}-F_{20} effect, because the IQF exchange-correlation term favoring the *gauche* orientation ($\Delta E_{int,xc}^{P\cdots Q}$=-2.5 kcal/mol) is completely canceled out by the electrostatic term ($\Delta E_{int,class}^{P\cdots Q}$=2.7 kcal/mol) favoring the *anti* one. Likewise the *g+/g+* and *anti/anti* pair in **11**, the exchange-correlation stabilization of the gauche F-C-C-F arrangement in **14a-conf1** is reversed by the electrostatic term favoring the *anti* orientation in **14a-conf5**, what is in contrast with the results obtained for the small models **1-5**. Hence, the destabilization of **14a-**

conf5 with reference to **14a-conf1** is better rationalized in terms of the forming/breaking of intramolecular contacts. Thus, the intermediate distance (5.1 Å) between the center of mass of the aromatic rings in fragments O and R with an almost parallel arrangement of the ring planes (12.1º) may allow some π–π favorable interaction in **14a-conf5**, whereas the loss of the CH···F contacts characteristic of **14a-conf1** would disfavor it. The energetic impact of these structural changes can be assessed through the IQF components. One the one hand, the approaching of the terminal O···R groups gives large and stabilizing values for the $\Delta E_{\text{int,disp}}^{O\cdots R}$ (12.5 kcal/mol) and $\Delta E_{\text{int,xc}}^{O\cdots R}$ (35.2 kcal/mol) terms. The $\Delta E_{\text{int,xc}}^{O\cdots R}$ value stems from many inter-atomic contributions involving the –CH_2-N moiety in fragment R and the amide and phenyl groups in fragment O (*e.g.*, H_{18}···O_{26} 3.6 kcal/mol, N_{13}···C_{39} 2.6 kcal/mol, C_{16}···O_{26} 2.2 kcal/mol, H_{17}···C_{35} 2.1 kcal/mol, etc.). The closer O···R contacts are accompanied by opposite intra-fragment distortion effects for the O and R moieties (-17.5 and -19.8 kcal/mol) so that the sum $\Delta E_{net}^{O} + \Delta E_{net}^{R} + \Delta E_{\text{int,disp}}^{O\cdots R} + \Delta E_{\text{int},xc}^{O\cdots R}$ amounts to 10.4 kcal/mol favoring **14a-conf5**. On the other hand, this contribution is overcompensated by the penalty associated with the loss of the contacts exhibited by **14a-conf1** (C_{16}-H_{18}···F_{23}, F_{20}···H_{44}, and H_{21}···O_{26}, see Figure 6) that is mainly accounted for by $\Delta E_{\text{int,xc}}^{P\cdots R}$ =-4.5 and $\Delta E_{\text{int,xc}}^{O\cdots Q}$ =-9.6 kcal/mol, determining thus the lower stability of **14a-conf5**.

As previously noticed, the *cis* F-C-C=O arrangement in **10** or **12** implies a large destabilization of more than 5 kcal/mol with respect to the *trans* orientation, basically due to the dipole-dipole interaction between the F-C and C=O groups. We see in Figure 6 that **14a-conf6** has a *cis* F_{23}-C_{22}-C_{25}=O_{26} orientation and two *anti* F_{23}-C_{22}-C_{19}-F_{20} and F_{20}-C_{19}-C_{16}-N_{13} groups, but it is only 2.6 kcal/mol less stable than **14a-conf1**. Nevertheless, the impact of the *cis* F_{23}-C_{22}-C_{25}=O_{26} dihedral in the electrostatic $\Delta E_{\text{int,class}}^{O\cdots P}$ value is -6.7 kcal/mol, which is similar to those observed in **10** or **12**. The *gauche* effect is again reversed in the **14a-conf1**/**14a-conf6** pair as the sum $\Delta E_{\text{int,xc}}^{P\cdots Q} + \Delta E_{\text{int,class}}^{P\cdots Q}$ equals to 1.0 kcal/mol. Further stabilization of **14a-conf6** comes from the various O···R interactions between the bulky terminal fragments, partially compensating the unfavorable *cis* F-C-C=O orientation. These interactions can be measured in terms of the sum $\Delta E_{net}^{O} + \Delta E_{net}^{R} + \Delta E_{\text{int,disp}}^{O\cdots R} + \Delta E_{\text{int},xc}^{O\cdots R}$ = 5.9 kcal/mol. In addition, only one of the two CH···F contacts is lost upon the **14a-conf1**→**14a-conf6** transition, determining an energy change $\Delta E_{\text{int,xc}}^{P\cdots R} + \Delta E_{\text{int,xc}}^{O\cdots Q}$ = -3.8 kcal/mol, which does not neutralize the global effect of the O···R contacts. Therefore, IQF shows how the *a priori* strong preference for the *trans*

F-C-C=O conformation in the **14a-conf1/14a-conf6** pair is significantly reduced by the unexpected reversal of the *gauche* effect and the corresponding balance of interactions between the terminal groups.

Discussion and Conclusions

Among the major conformational effects associated with C-F bonds, the *gauche* preference in F-C-C-X moieties (X=F, O, N, or C), the repulsion between C-F bonds aligned parallel in 1,3-positions, and the favored *trans*-planar arrangement in F-C-C=O moieties, are usually invoked to rationalize or foresee conformations in fluorine containing systems. However, either X-ray structures or NMR measurements have revealed molecular conformations that differ from those expected according to the usual fluorine effects.[15] Moreover, two possible *gauche* orientations (+60º/-60º) are accessible to F-C-C-X moieties, which in most cases result in different molecular conformations. Hence, additional intramolecular interactions and environmental effects may play a significant role in order to explain the conformational preferences of fluorinated molecules. To gain further understanding of such effects, IQA, as a reference-free energy decomposition method, can be useful to express the conformational energies predicted by QM calculations into unambiguously contributions, which include electrostatic and exchange-correlation interatomic interactions.

Concerning the prototypical *gauche* effect in 1,2-difluoroethane, the IQA method has been formerly employed[24] to highlight specifically the *gauche* stabilization due to the 1,3 Coulombic attraction between C and F atoms. However, we note that there are strong F···F and C···C repulsions as characterized by IQA that can be seen to compensate the C···F attractive energies. Moreover, the larger C-C-F angles and C···F distances obtained in the *gauche* form as compared to the *anti* one (110.3/108.1º and 2.34/2.32 Å) seem not compatible with explaining the *gauche* effect as the consequence of the 1,3 electrostatic attraction between C and F atoms. In addition, we also consider that other terms like the C···C and C···F exchange-correlation interactions contribute to the *gauche* stability. Thus, when the IQA atomic terms are grouped by defining two interacting CH_2F fragments separated by the rotatable bond, it emerges that the *gauche* structure is stabilized by the exchange-correlation interaction and destabilized by the electrostatic $CH_2F···CH_2F$ tem, involving only a minor distortion at each fragment. In this fragment-based IQF analysis, the electrostatic interaction can be further decomposed to show how the *gauche* destabilization is partially due to dipole-dipole repulsion, the dipoles being mainly associated to the C-F

bonds. We believe that the IQF interpretation of the *gauche* effect in 1,2-difluoroethane is more chemically appealing and, accordingly, we apply it to the rest of molecules considered in this work. We also note that, regardless of atomic contributions being grouped or not, IQA clearly shows that both electrostatic and exchange-correlation interactions should be considered to explain the *gauche* effect in 1,2-difluorethane.

Examination of the 1,2-difluoroethane related systems **2-7** bearing different polar/charged substituents points out that the quantum exchange-correlation inter-fragment interaction systematically stabilizes the *gauche* conformation. In contrast, the electrostatic interaction term is system-dependent because the sign and magnitude of $\Delta E_{\text{int,class}}^{P\cdots Q}$ is mainly determined by the particular interaction(s) between the fluorine atom in fragment P and the closest atoms in fragment Q. For the positively charged systems, the exchange-correlation and the classical electrostatic inter-fragment interactions are large and comparable, which contrasts with previous assumptions explaining their *gauche* preference only in terms of electrostatic effects. IQF also emphasizes that the magnitude and nature of the electrostatic interactions depends on the particular characteristics of the charged moiety.

As previously proposed,[9] dipole-dipole interaction arises as the main contribution to the stability of the *trans*-planar F-C-C=O arrangement in the CH_2F-COX molecules (**8-10**). However, the exchange-correlation term is also relevant for the ketone (**8**) and the amide (**10**) derivatives. More particularly, the interatomic F···HN interaction in **10** makes a large contribution to the $\Delta E_{\text{int,xc}}^{P\cdots Q}$ term. This relatively large exchange-correlation interaction between fluorine and nearby hydrogen atoms, also observed in 3-fluoropropanal (**3**) and 2-fluoroethan-1-aminium (**6**), can be considered as the signature for a (weak) hydrogen bond. Fluorine is considered a poor hydrogen bonding acceptor although there are examples of crystal structures showing short intramolecular contacts between organic fluorine and HO-/HN-moieties[9, 57] that have been assumed to result almost exclusively from dipole-dipole electrostatic interactions, excepting for F···HC contacts where it is thought that dispersion also plays a role. Nonetheless, our calculations help clarify the nature and impact of these F···HN and F···HC contacts that particularly affect the quantum mechanical interaction between the H-bonded groups, in line with recent $^4J_{HF}$ measurements and NBO calculations performed for α-fluoro amides.[30] Thus, we conclude that the largest exchange-correlation IQA energies help identify the relevant F···HC/HN interactions.

To analyze the various fluorine effects in more complex settings, we examined all the possible conformations in 1,2,3-trifluoropropane (**11**) and two other compounds (**12** and **13**) related to **14a/14b**. The two less stable conformers obtained for 1,2,3-trifluoropropane present the unfavorable parallel alignment of 1,3 C-F bonds previously described in polyfluorinated alkanes.[9] IQF confirms the appearance of an unfavorable electrostatic interaction between the 1,3 CH_2F groups and ascribes it to dipole-dipole repulsion. But unexpectedly, the most stable conformer obtained for 1,2,3-trifluoropropane does not result from maximizing the number of *gauche* effects and minimizing the 1,3 dipole-dipole repulsions. According to our analyses, the presence of 1,3 F···H-C interactions also contributes to explain the relative stability of the conformers. These F···H-C contacts are among the largest exchange-correlation interatomic interactions in 1,2,3-trifluoropropane and their stabilizing effect is larger than the *gauche* effect. The partial alignment of the amide dipole with a C-F bond and the presence of different F···H contacts also explain the conformational landscape obtained for 2,3-difluoro-*N*-methylpropanamide (**12**).

The various analyses carried out on the model systems may help us to better ascertain the various effects determining the conformational properties of the difluorinated dipeptide **14a** and **14b** diastereomers. According to X-ray crystallography, the (*R,R*) or (*R,S*) stereochemistry for the fluorination at the central C_{22} and C_{19} atoms leads to a different backbone conformation for the whole molecule.[14] In the two crystal structures, the vicinal C_{22}-F_{23}/C_{19}-F_{20} and C_{19}-F_{20}/C_{16}-N_{13} bonds are in *gauche* while the C_{22}-F_{23} bond aligns antiparallel to the adjacent amide carbonyl, which seems in consonance with expectations. The QM-refined conformational search performed for **14a/14b** predicts that the lowest energy conformer either in the gas-phase or in the solvent-continuum is structurally close to the crystallographic structure. This seems to indicate that crystal packing and solvent effects would play only a minor role in the conformational preferences of **14a/14b**, the intramolecular factors being dominant. However, the conformational search illustrates that the observed crystal structures are not the only conformers compatible with the effects usually assigned to fluorine atoms. For instance, two alternative *gauche* forms (*g+* or *g-*) could be accessible for the vicinal C_{22}-F_{23}/C_{19}-F_{20} and C_{19}-F_{20}/C_{16}-N_{13} bonds in the gas-phase or in solution. This seems confirmed by our estimations of the $^3J_{HF}$ values in chloroform obtained for the vicinal fluorine and hydrogen atoms bound to C_{22}/C_{19} in **14a/14b**, which are in reasonable agreement with experimental data. Moreover, the energetic penalty associated to the presence of unfavorable *anti* or *cis* arrangements around the fluorine atoms is not dramatic (*e.g.*, 2.5

kcal/mol), pointing thus towards the compensating roles of other stabilizing intra-molecular interactions.

The IQA energy decomposition can treat relatively large molecules like the selected conformers of **14a**. Three of them differ in the *g+*, *g-*, or *anti* arrangement of the F_{23}-C_{22}-C_{19}-F_{20} moiety, whereas the fourth one presents the less favorable F_{23}-C_{22}-C_{19}-F_{20} and F_{20}-C_{19}-C_{16}-N_{13} *anti* and F_{23}-C_{22}-C_{25}=O_{26} *cis* orientations. In all the cases, the careful examination of the IQF descriptors allows us to assess the energetic impact of the *gauche/anti* and *cis/trans* effects. In comparison with the smaller model systems **8-10** and **12**, IQF indicates that the important electrostatic F-C-C=O *trans* preference (~6-8 kcal/mol) is appreciably transferable. Interestingly, the exchange-correlation and electrostatic balance favoring the *gauche* F-C-C-F orientation can be altered in some of the **14a** conformers (also in **11** conformers) resulting in a small inverted *gauche* effect (<1.0 kcal/mol) that benefits the *anti* F-C-C-F arrangement. Furthermore, our four-fragment IQA partitioning points out that the formation/rupture of other intramolecular contacts (*e.g.*, π···π, F···H-C, etc.) can modulate and/or attenuate the F-C-C=O *trans* and the F-C-C-F *gauche* effects. Therefore, our results stress that the *gauche* effect can be both system and conformation dependent, what is in consonance with previous studies[19, 26-28] showing that environmental (solvent) effects can modulate or alter the *gauche* conformational preference.

In summary, our QM calculations complemented with the HF-D3 IQF energy decompositions are useful to analyze in a systematic and consistent manner the energetic preferences of small fluorinated compounds having either *gauche/anti* or *cis/trans* conformations. Following the IQF approach, we find a clear correspondence between specific exchange-correlation and/or electrostatic fragment-interaction energies with the appearance of the *gauche/anti* or *cis/trans* effects, providing also insight into their magnitude and nature. The same IQF approach can be extended to assess those *gauche/anti* or *cis/trans* effects in molecules with two or more rotatable bonds as well as to study the roles played by other concomitant effects (*e.g.*, specific CH/OH/NH···F contacts, 1-3 electrostatic interactions, etc.). For the relatively large α,β-difluoro-γ-amino acid derivatives, our conformational search followed by QM and selected IQF calculations complement well their crystallographic and NMR characterization. The conformational preferences of these compounds as detailed by the theoretical analysis point out that the *gauche/anti* and *cis/trans* effects associated to fluorine bonds may be attenuated in large molecules, where the most preferred conformations may be dictated by other non-fluorine specific intra-molecular interactions.

Finally, we note that similar computational protocols including solvent effects could help in the rational design of fluorinated molecules having a nearly frozen conformation tailored for a particular application. This is still a challenging task as most of the fluorinated compounds have emerged from broad chemical screening programs and the actual influence of fluorine is considered retrospectively.[9] However, the QM and IQF assessments of the conformational effects associated to the presence of fluorine atoms (*gauche* effect, the 1,3 C-F repulsion and other electrostatic interactions, the hydrogen bond acceptor capability of organic fluorine, etc.) could result in new guidelines to predict the conformation of structurally-complex fluorinated molecules.

Electronic Supplementary Information

Table S1 with the sum of the different IQA atomic contribution for the small (**1-10**) models. Tables S2-S5 and S11 with the largest changes in the IQA atomic contributions for the different systems. Figure S1 with a representation of the different conformers obtained for **13** and their relative energies. Table S6 and S7 with IQF and the largest changes in the IQA atomic contributions and results obtained for **13**. Tables S8 and S9 with energies and geometrical parameters obtained for the different conformers optimized for **14a** and **14b**. Figures S2 and S3 with a ball-and-stick representation of the different conformers optimized for **14a** and **14b**. Table S10 with the $^3J_{H,F}$ coupling constants estimated for some of the **14a** and **14b** conformers. Table S12 with the equivalence between the filenames with the Cartesian coordinates (COORD.zip) and the HF/MP2 optimized compounds. Table S13 comparing the HF-D3 IQF and correlated IQF results for 1,2-difluoroethane.

References

1. Y. Zhou, J. Wang, Z. Gu, S. Wang, W. Zhu, J. L. Aceña, V. A. Soloshonok, K. Izawa and H. Liu, *Chem. Rev.*, 2016, 116, 422-518.
2. P. Jeschke, *ChemBioChem*, 2004, 5, 570-589.
3. T. Fujiwara and D. O'Hagan, *J. Fluor. Chem.*, 2014, 167, 16-29.
4. P. Kirsch, *J. Fluor. Chem.*, 2015, 177, 29-36.
5. J. H. Yun, S. Park, J. H. Heo, H.-S. Lee, S. Yoon, J. Kang, S. H. Im, H. Kim, W. Lee, B. Kim, M. J. Ko, D. S. Chung and H. J. Son, *Chem. Sci.*, 2016, 7, 6649-6661.
6. L. E. Zimmer, C. Sparr and R. Gilmour, *Angew. Chem. Int. Ed.*, 2011, 50, 11860-11871.
7. D. Cahard and V. Bizet, *Chem. Soc. Rev.*, 2014, 43, 135-147.
8. A. A. Berger, J.-S. Völler, N. Budisa and B. Koksch, *Acc. Chem. Res.*, 2017, 50, 2093-2103.
9. D. O'Hagan, *Chem. Soc. Rev.*, 2008, 37, 308-319.
10. H.-J. Böhm, D. Banner, S. Bendels, M. Kansy, B. Kuhn, K. Müller, U. Obst-Sander and M. Stahl, *ChemBioChem*, 2004, 5, 637-643.
11. J. W. Banks, A. S. Batsanov, J. A. K. Howard, D. O'Hagan, H. S. Rzepa and S. Martin-Santamaria, *J. Chem. Soc., Perkin Trans. 2*, 1999, 2409-2411.
12. M. Schüler, D. O'Hagan and A. M. Z. Slawin, *Chem. Commun.*, 2005, 4324-4326.
13. L. Hunter, P. Kirsch, A. M. Z. Slawin and D. O'Hagan, *Angew. Chem.*, 2009, 121, 5565-5568.
14. L. Hunter, K. A. Jolliffe, M. J. T. Jordan, P. Jensen and R. B. Macquart, *Chem. Eur. J.*, 2011, 17, 2340-2343.
15. F. Scheidt, P. Selter, N. Santschi, M. C. Holland, D. V. Dudenko, C. Daniliuc, C. Meck-Lichtenfeld, M. R. Hansen and R. Gilmour, *Chem. Eur. J.*, 2017, 23, 6142-6149.
16. N. E. J. Gooseman, D. O'Hagan, M. J. G. Peach, A. M. Z. Slawin, D. J. Tozer and R. J. Young, *Angew. Chem.*, 2007, 119, 6008-6012.
17. L. Goodman, H. Gu and V. Pophristic, *J. Phys. Chem. A*, 2005, 109, 1223-1229.
18. D. Y. Buissonneaud, T. van Mourik and D. O'Hagan, *Tetrahedron*, 2010, 66, 2196-2202.
19. F. A. Martins and M. P. Freitas, *Eur. J. Org. Chem.*, 2019, 6401-6406.
20. F. Weinhold, C. R. Landis and E. D. Glendening, *Int. Rev. Phys. Chem.*, 2016, 35, 399-440.

21. F. M. Bickelhaupt and E. J. Baerends, *Angew. Chem. Int. Ed.*, 2003, 42, 4183-4188.

22. P. Su and H. Li, *J. Chem. Phys.*, 2009, 131, 014102.

23. M. Baranac-Stojanović, *RSC Adv.*, 2014, 4, 43834-43838.

24. J. C. R. Thacker and P. L. A. Popelier, *J. Phys. Chem. A*, 2018, 122, 1439-1450.

25. C. R. S. Briggs, M. J. Allen, D. O'Hagan, D. J. Tozer, A. M. Z. Slawin, A. E. Goeta and J. A. K. Howard, *Org. Biomol. Chem.*, 2004, 2, 732-740.

26. C. Tormena, M. P. Freitas, R. Rittner and R. J. Abraham, *Phys. Chem. Chem. Phys.*, 2004, 6, 1152-1156.

27. B. C. Fiorin, E. A. Basso, C. Tormena, R. Rittner and R. J. Abraham, *J. Phys. Chem. A*, 2009, 113, 2906-2913.

28. G. Pattison, *Beilstein J. Org. Chem.*, 2017, 13, 2915-2921.

29. C. Trindle, P. Crum and K. Douglass, *J. Phys. Chem. A*, 2003, 107, 6236-6242.

30. E. Cosimi, N. Trapp, M.-C. Ebert and H. Wennemers, *Chem. Commun.*, 2019, 55, 2253-2256.

31. J. Aleksić, M. Stojanović and M. Baranac-Stojanović, *J. Org. Chem.*, 2015, 80, 10197-10207.

32. E. Bogdan, G. Compain, L. Mtashobya, J.-Y. Le Questel, F. Besseau, N. Galland, B. Linclau and J. Graton, *Chem. Eur. J.*, 2015, 21, 11462-11474.

33. R. A. Cormanich, D. O´Hagan and M. Bühl, *Angew. Chem. Int. Ed.*, 2017, 56, 7867-7870.

34. M. A. Blanco, A. Martín Pendás and E. Francisco, *J. Chem. Theory Comput.*, 2005, 1, 1096-1109.

35. D. Suárez, N. Díaz, E. Francisco and A. Martín Pendás, *ChemPhysChem*, 2018, 19, 973-987.

36. S. Grimme, *WIREs Comput Mol Sci*, 2011, 1, 211-228.

37. A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2005, 122, 154104.

38. L. Goerigk and J. R. Reimers, *J. Chem. Theory Comput.*, 2013, 9, 3240-3251.

39. L. Goerigk, C. A. Collyer and J. R. Reimers, *J. Phys. Chem. B*, 2014, 118, 14612-14626.

40. T. Ziegler and A. Rauk, *Theoret. Chim. Acta*, 1977, 46, 1-10.

41. F. M. Bickelhaupt and E. J. Baerends, in *Rev. Comput. Chem.*, John Wiley & Sons, Inc., 2007, DOI: 10.1002/9780470125922.ch1, pp. 1-86.

42. E. Francisco and A. Martín Pendás, in *Non-Covalent Interactions in Quantum Chemistry and Physics*, eds. A. Otero de la Roza and G. A. DiLabio, Elsevier, 2017, DOI: https://doi.org/10.1016/B978-0-12-809835-6.00003-7, pp. 27-64.

43. C. Outeiral, M. A. Vincent, Á. Martín Pendás and P. L. A. Popelier, *Chemical Science*, 2018, 9, 5517-5529.

44. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem.*, 2004, 25, 1605-1612.

45. F. Neese, *WIREs Comput Mol Sci*, 2012, 2, 73-78.

46. S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Chem. Rev.*, 2016, 116, 5105-5154.

47. D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin and F. Neese, *J. Chem. Theory Comput.*, 2015, 11, 1525-1539.

48. N. Sauton, D. Lagorce, B. O. Villoutreix and M. A. Miteva, *BMC Bioinformatics*, 2008, 9, 184.

49. J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J Mol Graph Model*, 2006, 25, 247-260.

50. A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, 113, 6378-6396.

51. C. Thibaudeau, J. Plavec and J. Chattopadhyaya, *J. Org. Chem.*, 1998, 63, 4967-4984.

52. A. M. Pendas and E. Francisco, Unpublished, 2015.

53. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, 132, 154104.

54. E. Francisco, D. Menéndez Crespo, A. Costales and Á. Martín Pendás, *J. Comput. Chem.*, 2017, 38, 816-829.

55. J. Pipek and P. G. Mezey, *J. Chem. Phys.*, 1989, 90, 4916-4926.

56. A. Martín Pendás, M. A. Blanco and E. Francisco, *J Comput Chem*, 2009, 30, 98-109.

57. I. Hyla-Kryspin, G. Haufe and S. Grimme, *Chem. Eur. J.*, 2004, 10, 3411-3422.

Interacting Quantum Atoms Approach and Electrostatic Solvation Energy: Assessing Atomic and Group Solvation Contributions

Fernando Jiménez-Grávalos, Natalia Díaz, Evelio Francisco, Ángel Martín-Pendás and Dimas Suárez*

Departamento de Química Física y Analítica. Universidad de Oviedo.

Julián Clavería 8. 33006 Oviedo (Asturias) Spain.

Abstract: The interacting quantum atoms (IQA) method decomposes the total energy of a molecular system in terms of one- and two-center (atomic) contributions within the context of the quantum theory of atoms in molecules. Here we incorporate electrostatic continuum solvent effects into the IQA energy decomposition. To this end, the interaction between the solute electrostatic potential and the solvent screening charges as defined within the COSMO solvation model is now included in a new version of the PROMOLDEN code, allowing thus to apply IQA in combination with COSMO-quantum chemical methods as well as to partition the electrostatic solvation energy into effective atomic and group contributions. To test the robustness of this approach, we carry out COSMO-HF/aug-cc-pVTZ calculations followed by IQA calculations on more than 400 neutral and ionic solutes extracted from the MNSol database. The computational results reveal a detailed atomic mapping of the electrostatic solvation energy that is useful to assess to what extent the solvation energy can be decomposed into atomic and group contributions of various parts of a solute molecule, as generally assumed by empirical methodologies that estimate solvation energy and/or *logP* values.

Introduction

The Quantum Chemical Topology (QCT) methods[1] take advantage of the topological properties of scalar fields (charge density and others) in order to gain new chemical information about bonding and molecular properties. Among them, the interacting quantum atoms (IQA) approach,[2-3] employs the first- and second-order reduced density matrices to partition the expectation values into atomic regions such as the attraction basins (Ω_A) of the gradient field of the electron density. Thus, IQA provides self-atomic energies, $E(\Omega_A)$, which tend to the free atomic energies at the limit of non-interacting atoms, and diatomic $E(\Omega_A, \Omega_B)$ energies that unambiguously discriminate between classical electrostatic and exchange-correlation energy terms. Using DFT (and HF) charge densities, IQA can be augmented with the Grimme's D3 potential,[4-5] which yields pairwise dispersion energies E_{disp}(A-B) that complement the diatomic $E(\Omega_A, \Omega_B)$ IQA terms, constituting thus an effective D3-IQA decomposition scheme[6] applicable to medium-sized and large systems. Thus, the IQA or D3-IQA decomposition has been successfully applied to quantify many different aspects of chemical bonds and intermolecular forces. Among the various topics that have been recently addressed using IQA, we find the nature and cooperativity of H-bond interactions,[7-10] halogen bonding patterns,[11] interactions within transition metal complexes,[12-13] description of short-range repulsions,[14] fine-tuning effects of electron correlation within covalent and non-bonded interactions,[15] the categorization of non-covalent bonding and the atomic decomposition of intermolecular binding energies,[6] etc..

Up to date all the IQA calculations have been performed considering molecular species (single molecules, dimers, clusters) in the gas-phase. However, it is clear that solvent plays a major role in determining the stability and molecular properties of organic molecules and biomolecules in solution. Moreover, since we are interested in pursuing the application of D3-IQA to quantify atomic and group energy contributions in biomolecular systems with many functional groups, the treatment of solvent effects within the IQA framework is therefore a prerequisite. To this end, the combination of implicit solvent models and quantum mechanical (QM) methods constitutes probably the best methodological

choice given that the continuum treatment of solvent focuses on the degrees of freedom of the solute(s) while it provides an accurate description of the strong, long-range electrostatic forces that dominate solvation energies in high dielectric solvents.[16] In these hybrid approaches, the solute-solvent electrostatic interaction is usually described in terms of the reaction (electric) field exerted on the QM charge density of the solute by the solvent that in turn is polarized. Other empirical and semiempirical methods, which generally do not affect the charge density of the solute, have been proposed to estimate the non-electrostatic contributions to solvation energy, which are significant, especially in non-polar solvents.[17]

The decomposition of the QM energy in solution by IQA would render atomic and group contributions to the solute solvation energy. Nonetheless, the actual significance of this partitioning should be carefully considered. Thus, in classical Statistical Mechanics,[18] the free energy of solvation of a rigid solute can be expressed as:

$$\Delta G_{solv} = -RT \ln \left\langle \exp\left(-V/RT\right)\right\rangle_{NPT}$$

where V is the solute-solvent interaction potential and $\langle \ \rangle$ represents the ensemble average over all possible configurations of the solvent molecules in the system. Although the total interaction energy V may be split into group/atomic contributions of the different solute atoms, the ensemble average in the above expression cannot be factorized into a product of two or more average quantities. Physically, this means that the solvation shells around the solute atoms/groups are correlated at varying degrees and, therefore, it is not feasible to achieve an exact additivity of ΔG_{solv}. Nevertheless, computational models have been developed for the fast prediction of hydration energies or partition coefficients (such as the logarithm of the octanol-water partition coefficient, which can be calculated as $\log P = \left(\Delta G_{solv}^{octanol} - \Delta G_{solv}^{water}\right)\big/2.303RT$) that rely on the assumption of atomic/fragment additivity for ΔG_{solv} and derive atomic/fragment parameters using different optimization strategies.[19-23]

In a previous work,[24] QM calculations have been used to heuristically determine group contributions to the free energy of solvation. However, this study is limited to a family of closely-related heterocyclic compounds, although it is concluded that group contributions are slightly affected by the chemical environment. Therefore, the IQA decomposition of the solvation free energy into *effective* atomic terms, which ultimately arise from the topological partitioning of the solute charge density, constitutes an opportunity to further assess the additivity assumption.

In the rest of the paper, we will briefly describe the theoretical details of the IQA extension to accomplish the decomposition of the QM energy of solute molecules embedded within a solvent continuum. In doing so, we will focus on the electrostatic solute-solvent interaction accounted for by the *co*nductor-like *s*creening *mo*del (COSMO). The IQA-COSMO protocol will be applied to a large set of organic molecules retrieved from the Minnesota Solvation database (MNSol),[25-26] which collects experimental free energies of solvation for hundreds of solutes and QM optimized geometries for the corresponding solutes. For a subset of neutral and ionic solutes comprising 412 molecules, we perform geometry optimizations both in the gas-phase and in solution using the HF method with a triple-ζ basis set followed by full IQA calculations. Then we will assess the accuracy of the calculated solvation energies and the numerical errors in the IQA-reconstructed energies. Subsequently, we will characterize statistically the fragment-based IQA contributions to the electrostatic solvation energies. The chemical fragments comprise united atom types and functional groups that are selected using similar prescriptions to those of Meylan and Howard.[23] Finally, we will assess the goodness of the solvation energy additivity approximation by comparing between COSMO-HF energies and additive energies for an additional set of MNSol structures not considered in the IQA calculations.

Theory

IQA in the gas-phase

Starting with the atomic basins (Ω_A), which stem from the topological properties of the charge distribution $\rho(\mathbf{r})$, the IQA approach[2, 27] needs two scalar fields derived from the QM wavefunction, the

first order reduced density matrix $\rho_1(\mathbf{r_1},\mathbf{r_1}')$ and the pair density, $\rho_2(\mathbf{r_1},\mathbf{r_2})$. Then IQA decomposes the total energy of a molecular system in the gas-phase (E^{gas}) as

$$
\begin{aligned}
E^{gas} &= \sum_A E_{net}^A + \sum_{A>B} E_{int}^{AB} = \sum_A \left(T^A + V_{ne}^A + V_{ee}^A \right) \\
&+ \sum_{A>B} \left(V_{nn}^{AB} + V_{ne}^{AB} + V_{ne}^{BA} + V_{ee}^{AB} \right)
\end{aligned}
\tag{1}
$$

where $E_{net}^A \equiv E_{net}(\Omega_A)$ is the net electronic energy of atom A that includes the kinetic energy T^A and the potential energy due to nucleus-electron (*ne*) attractions and electron-electron repulsions (*ee*) within Ω_A. The interaction energy $E_{int}^{AB} = E_{int}(\Omega_A, \Omega_B)$ between atoms A and B in the molecular system collects various potential energy terms (*nn*, *en*, *ne* and *ee*). Note that in the IQA terminology, *interaction* energies are the diatomic contributions to the *absolute* energy of a molecule. By grouping half the interaction energy terms involving atom Ω_A and its net energy, we define its additive energy,

$$
E_{add}^A \equiv E_{add}(\Omega_A) = E_{net}(\Omega_A) + \frac{1}{2} \sum_{B \neq A} E_{int}(\Omega_A, \Omega_B), \quad (2)
$$

so that the sum of all the E_{add}^A terms reproduces the total energy E^{gas}.

Implicit solvent methods: COSMO

Several excellent reviews have been published[16-17, 28-29] that examine the various approximations underlying the implicit solvent methods. Herein, we briefly review the most basic concepts and some details of the COSMO method that are required to understand the IQA-COSMO protocol. Thus, the definition of the molecular cavity and the description of electrostatic solute-solvent interaction are the basic elements of a continuum solvent model.[28] The shape and size of the cavity are typically defined by a solvent excluded surface (SES), which encloses the volume in which the solvent molecules cannot penetrate. Thus, the SES, which is the boundary of the molecular cavities, can be computed using different sets of atomic radii and numerical algorithms depending on the continuum model. Among the various techniques for solving the electrostatic problem, the apparent surface charges (ASC) method[28] allows a direct implementation of continuum solvent effects within IQA. In this approach, the reaction field potential generated by the polarization of the dielectric medium is expressed in terms

of a set of point charges q_k assigned to small surface segments (tesserae) located at positions $\mathbf{s}_k$. The values of q_k are determined by imposing the proper boundary conditions on the SES that, in the COSMO model,[30-31] correspond to the vanishing potential on and within a grounded conductor. The COSMO model mimics solvents with finite dielectric constant by scaling down the q_k values by a factor $f(\varepsilon)=(\varepsilon-1)/(\varepsilon+x)$ with $x=0.5$ and 0.0 for neutral and ionic molecules, respectively. The so-called outlying charge, which arises from the tail of the solute electron density that lies outside the molecular cavity, can affect negatively the results of the continuum models. In the COSMO methodology, the outlying charge correction (*occ*) is an heuristic approximation that corrects both the apparent surface charges q_k and the solute electrostatic potential $\Phi(\mathbf{s}_k)$.[32] However, we found that the COSMO hydration energies with and without the *occ* term correlate with experimental data very similarly (see below) and, therefore, we decided not to include the *occ* term in the IQA decomposition.

Knowing the values of the apparent surface charges q_k, the solute-solvent electrostatic energy (V_{solv}) is

$$V_{solv} = V_{solv,n} + V_{solv,e} = \sum_A \sum_k \frac{Z_A q_k}{|\mathbf{R}_A - \mathbf{s}_k|} + \sum_k \int \frac{\rho(\mathbf{r}) q_k}{|\mathbf{r} - \mathbf{s}_k|} d\mathbf{r} \qquad (3)$$

where $\rho(\mathbf{r})$ is the electron charge density of the solute and Z_A is the nuclear charge of the atom A located at $\mathbf{R_A}$. This expression can be rewritten as

$$V_{solv} = \sum_k \Phi(\mathbf{s}_k) q_k \qquad\qquad\qquad (4)$$

where $\Phi(\mathbf{s}_k)$ is the total electrostatic potential created by the solute acting on the tesserae $\mathbf{s}_k$. This solute-solvent interaction energy accounts for the electric work (free energy) needed to transfer the unperturbed solute from the gas-phase to the solvent cavity in the presence of the q_k charges located at the $\mathbf{s}_k$ positions, but does not describe the polarization of the solvent continuum. Assuming a linear response regime, it can be shown that the free energy due to the building of the solvent polarization is equal to $-V_{solv}/2$.[29] Therefore, the total free energy gain associated to the solvation process would be $V_{solv}/2$.

To solve the QM problem of the solute embedded in the continuum, the implicit solvent methods construct effective Hamiltonians that include both the solute-solvent interaction and the solvent polarization. Mutual solute-solvent polarization effects are considered through a self-consistent reaction field (SCRF) iterative process. For HF/DFT methods, the wavefunction/charge density in solution is iteratively obtained so that the molecular/Kohn-Sham orbitals and the reaction field potential (*i.e.*, q_k values in COSMO method) are updated at each self-consistent-field cycle. After convergence, the QM energy of a molecular system and the dielectric continuum is obtained by adding the $V_{solv}/2$ term to the rest of kinetic and potential energy terms associated to the electronic and nuclear degrees of freedom of the solute.

IQA partitioning of the solute-solvent interaction energy

The IQA partitioning of the total QM energy in solution derived from the COSMO method relies on the monoelectronic character of the solute electrostatic potential, Φ. Its decomposition into atomic contributions is straightforward,

$$\Phi(\mathbf{s}_k) = \sum_A \Phi^A(\mathbf{s}_k) \tag{5}$$

so that $\Phi^A(\mathbf{s}_k)$ is the electrostatic potential created by the nuclear charge and electron density confined within the atomic basin Ω_A. This quantity is readily computable within the IQA framework, yielding thus the atomic contribution to V_{solv},

$$V_{solv}^A = \sum_k \Phi^A(\mathbf{s}_k) q_k \tag{6}$$

Once that the SCRF process is converged and the corresponding density matrices become available, IQA decomposes the QM energy in solution as,

$$E^{sol} = \sum_A \left(T^A + V_{ne}^A + V_{ee}^A + \frac{1}{2} V_{solv}^A \right)$$
$$+ \sum_{A>B} \left(V_{nn}^{AB} + V_{ne}^{AB} + V_{ne}^{BA} + V_{ee}^{AB} \right) \tag{7}$$

Solvation free energy

Ignoring thermal corrections to the free energy associated with the solute degrees of freedom, it turns out [29] that the electrostatic solvation free energy ΔG_{solv} is the difference between the QM energies in solution and in the gas-phase

$$\Delta G_{solv} = E^{sol} - E^{gas} \qquad (8)$$

Two separate contributions to ΔG_{solv} can be defined: the Coulomb term ΔG_{solv}^{Coul} due to the electrostatic interaction between the unperturbed (*i.e.*, not polarized) solute and the solvent continuum and the polarization or induction ΔG_{solv}^{pol} energy gained upon mutual polarization. By computing E^{sol} through a single SCRF cycle with the unperturbed gas-phase charge density, ΔG_{solv}^{Coul} can be known and thereby $\Delta G_{solv}^{pol} = \Delta G_{solv} - \Delta G_{solv}^{Coul}$. Inserting then the corresponding IQA additive energies, ΔG_{solv}, ΔG_{solv}^{Coul} and ΔG_{solv}^{pol} can decomposed into effective atomic solvation energies. For example,

$$\Delta G_{solv} = E^{sol} - E^{gas} = \sum_A \left(E_{add}^{sol,A} - E_{add}^{gas,A} \right) = \sum_A \Delta G_{solv}^A \qquad (9)$$

Let us stress that this is indeed an *effective* partitioning given that each ΔG_{solv}^A term collects the mutual solute-solvent polarization effects due to the charge density $\rho(\mathbf{r})$ of the solute within the basin Ω_A and *all* the apparent surface charges distributed over the molecular surface.

Results and Discussion

Solvation energy calculations

The MNSol database contains the Cartesian coordinates of 533 molecules solvated by pure water and of other 106 molecules in mixed aqueous organic solvent (*e.g.*, water-octanol). From these MNSol data, we selected 412 molecules to carry out the QM and IQA calculations on the basis of appropriate molecular size. Thus, small molecules containing one or two heavy atoms (*e.g.*, water, ammonia, acetylene, hydroxide anion, etc.) were not considered as the emphasis is placed on the analysis of functional group contributions. Relatively big molecules containing more than 25 atoms were not selected neither in order to keep the computational cost of the expensive IQA calculations within reasonable bounds. In the final set, 57 anionic and 49 cationic species were included. The molecular

geometries of all the molecules were fully optimized at the HF/aug-cc-pVTZ level both in the gas-phase and in the solvent continuum.

Figure 1. Comparison between the COSMO-HF/aug-cc-pVTZ calculated hydration energies (ΔGcalc in kcal/mol) and the corresponding experimental values (ΔGexp) of the selected MNSol structures in aqueous solvent. The determination coefficient (R2), the Spearman correlation coefficient (ρ) and the root mean square (RMS) error in kcal/mol are also indicated for the whole data set (in black) and for the neutral molecules (in red), anionic (in blue) and cationic (in magenta) categories. The blue dashed line is the least squared fit line between the calculated and the reference data.
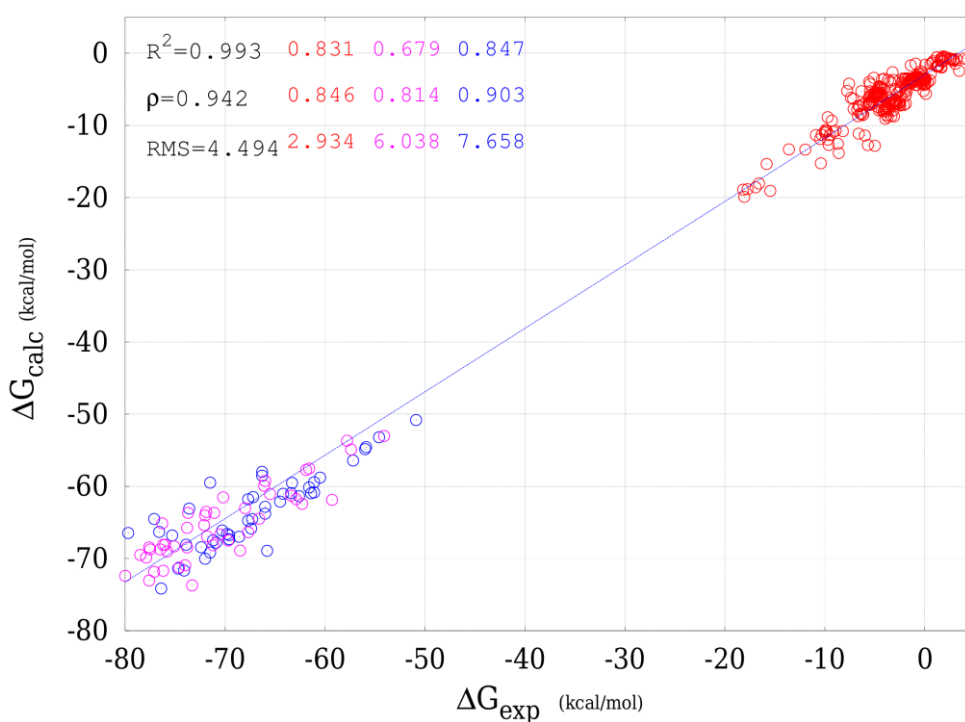
In Figure 1 the experimental hydration energies of 364 molecules are compared with the COSMO-HF/aug-cc-pVTZ ΔG_{solv} values. The global correlation between calculated and experimental data is strong as the determination coefficient R^2 amounts to 0.993. The root mean squared (RMS) error is significant, 4.5 kcal/mol, which is not entirely unexpected due to the lack of electron correlation effects, non-polar solvation contributions and conformational sampling. The performance of the COSMO-HF/aug-cc-pVTZ level depends on the charge state of the solute molecules. Thus, the

solvation energies for neutral and anionic molecules show a good correlation with experimental values (R^2=0.83 and 0.85, respectively) whereas the prediction capacity for the cationic species turns out to be lower (R^2=0.68). We also note in passing that the COSMO-HF/aug-cc-pVTZ ΔG_{solv} energies including the outlying charge correction have very similar statistical metrics (R^2= 0.992, 0.846, 0.844 and 0.688 for the full set, neutral, anionic and cationic solutes, respectively).

The reliability of the COSMO-HF ΔG_{solv} values is not far from that of more sophisticated QM solvent models like the embedded cluster reference interaction site model (EC-RISM) integral equation theory coupled with the MP2/6-311+G(d,p) ab initio method.[33] This optimized EC-RISM protocol, which incorporates conformational sampling, yields a global R^2=0.99 and RMS error of 2.4 kcal/mol for the MNSol structures in water. The statistical measurements of the EC-RISM data are also less satisfactory if neutral and charged species are analyzed separately (R^2=0.89, 0.88 and 0.85 for neutrals, anions and cations, respectively). Hence, we conclude that the HF/aug-cc-pVTZ level, which is particularly suitable for carrying out the IQA calculations, captures reasonably well the trends exhibited by the hydration energies of neutral and anionic molecules, the case of cations being somewhat less satisfactory.

Figure 2. Atomic distribution of the IQA $\Delta G_{solv}{}^{A}$ energies (in kcal/mol) and Bader atomic charges (in parentheses) for some neutral MNSol molecules: *E*-2-pentene, *m*-xylene, 1,1,2-trichloro-1,2,2-trifluoroethane, pentanol, butanal, 2-methylpyridine, nitroethane, and 3-chlorophenylurea. Total ΔG_{solv} values in kcal/mol are also indicated (in red).
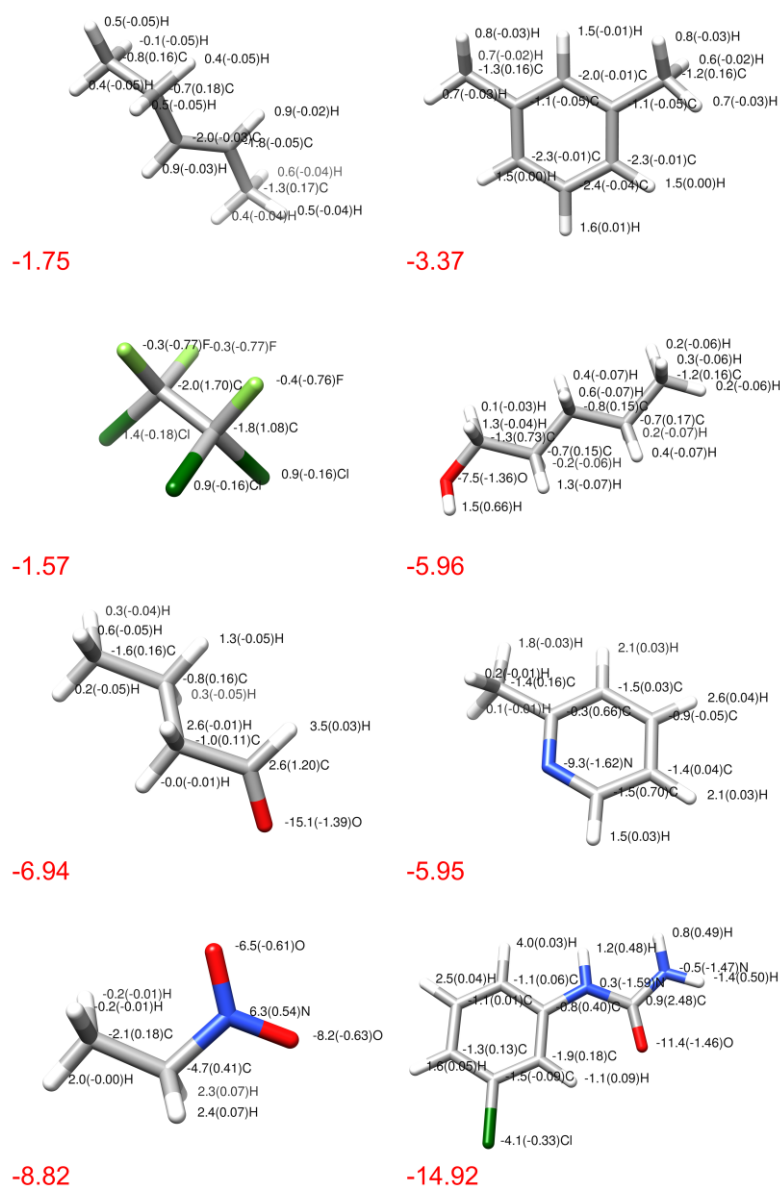
IQA decomposition of solvation energy

The calculation of the IQA energy terms involves six-dimensional numerical integration over the atomic basins, which is computationally expensive and introduces some numerical error.[34-35] To estimate this error, we compared the solvation energies ΔG_{solv} derived from the gas-phase and COSMO HF calculations with their counterpart values obtained from the IQA-reconstructed energies (ΔG_{solv}^{IQA}). The absolute differences $|\Delta G_{solv} - \Delta G_{solv}^{IQA}|$ can be considered as a measure of the "IQA numerical error" in the decomposition of solvation energies. Its mean value is 0.76 ± 1.36 kcal/mol, which corresponds to a average error per atom of 0.06 ± 0.10 kcal/mol. The magnitudes of these error estimates are similar to those previously found in the IQA decomposition of formation energies for non-covalent complexes.[6] We note again that the actual interest of the IQA energy partitioning resides in the atomic and/or fragment-based IQA components and they have values ranging from ~0.5 to tens of kcal/mol in absolute value (see below) that are well above the mean numerical error per atom.

Figure 2 displays the stick models of the COSMO HF/aug-cc-pVTZ optimized structures for selected neutral solutes together with the ΔG_{solv}^{A} energies (in kcal/mol) and the charge corresponding to each atomic basin (Figure S1 in the supporting information shows the same data for the whole set of structures). For the non-polar hydrocarbon molecules, the effective atomic contributions to the electrostatic solvation energy are mainly negative (favorable) for the C atoms and positive for Hs, the absolute values of ΔG_{solv}^{A} being small (~0.5-2.0 kcal/mol). The atomic charges throughout these hydrophobic molecules are also quite small (0.01-0.05 in absolute value) and they are uncorrelated with the ΔG_{solv}^{A} values. Indeed another non-polar molecule, the halogen-substituted ethane, exhibits relatively large atomic charges (e.g., q_A= -0.8, +1.7), but $|\Delta G_{solv}^{A}|$ values < 1.6 kcal/mol (see Figure 2).

Figure 3. Atomic distribution of the IQA $\Delta G_{\text{solv}}^{\text{A}}$ energies (in kcal/mol) and Bader atomic charges (in parentheses) for some ionic MNSol molecules: *N*-ethylethanamine (+), acetonitrile (+), ethanol (+), acrylic acid (-), phenylmethanol (-), 3-chloroaniline (+).Total ΔG_{solv} values in kcal/mol are also indicated (in red).
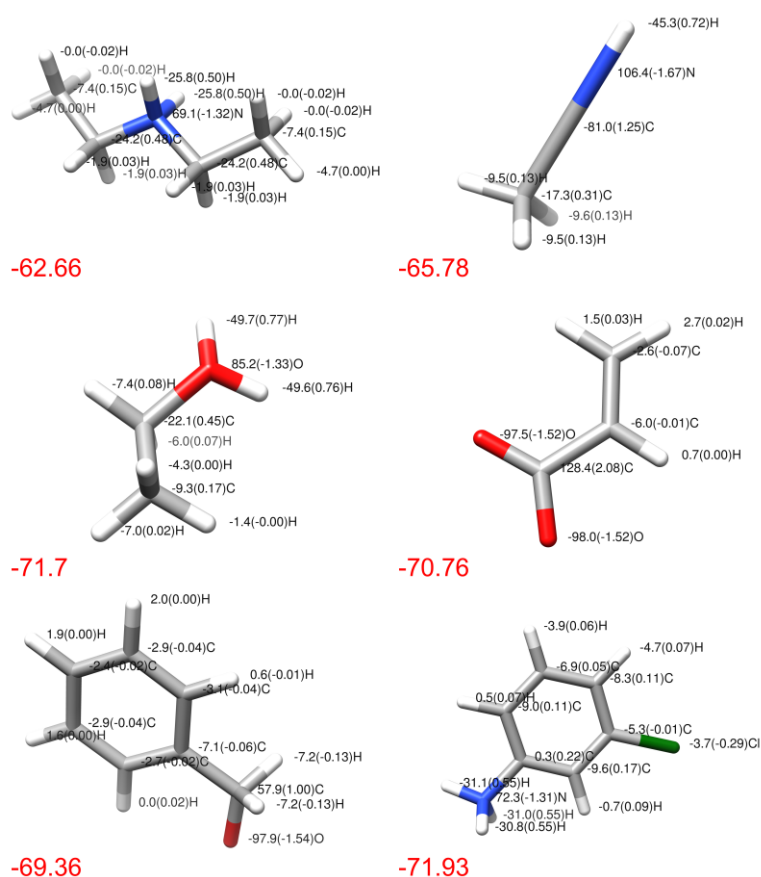
The stronger hydration induced by the polar groups (alcohol, aldehyde, amino, nitro, etc.) into the neutral solutes, whose total ΔG_{solv} values are around -6, -8 kcal/mol, arises from relatively large $\Delta G_{\text{solv}}^{\text{A}}$ contributions associated to the atomic basins of the polar functional groups. For example, the three atoms of the aldehyde HC=O group in butanal result in a $\Delta G_{\text{solv}}^{\text{CHO}}$ term of -9.0 kcal/mol, the molecular ΔG_{solv} value being -6.9 kcal/mol. The non-polar hydrocarbon moieties in the mono-substituted compounds yield atomic $\Delta G_{\text{solv}}^{\text{A}}$ values (~0.5-2.0 kcal/mol in absolute value) that are similar to those in the non-polar hydrocarbon molecules. The $\Delta G_{\text{solv}}^{\text{A}}$ and q^{A} values of polar sites are more widely distributed than in non-polar molecules and exhibit R^2 values ranging between 0.4 and 0.8 depending

13

on the solute molecule. Thus, it seems that atomic charges significantly determine the polar $\Delta G_{\text{solv}}^{A}$ contributions.

Figure 4. Correlation plot between the Bader's charges of atomic basins (q^A, derived from COSMO-HF/aug-cc-pVTZ density) and the corresponding IQA $\Delta G_{\text{solv}}^{A}$ energies for ionic molecules: cationic (in blue) and anionic (in red). The dashed lines are the least squared fit lines. The determination coefficient (R^2) and the Spearman correlation coefficient (ρ) are also indicated.
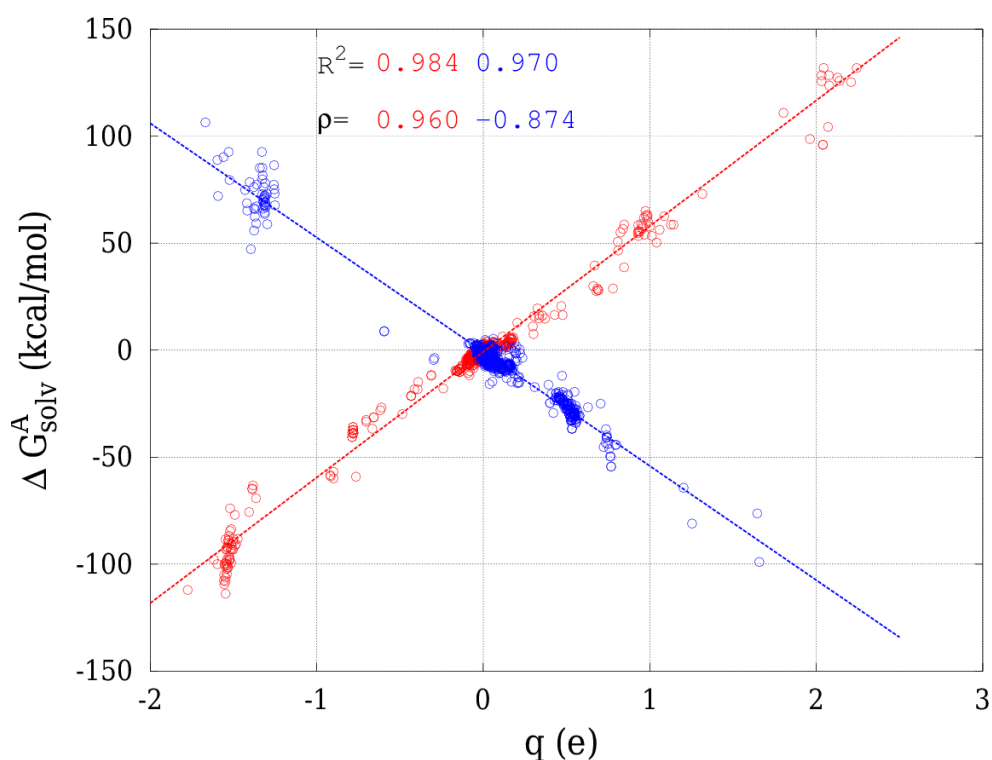
Ionic solutes are characterized by large solvation energies of tens or even hundreds of kcal/mol in water. For the cationic and anionic solutes shown in Figure 3, the IQA COSMO-HF calculations indicate again that the ionic functionalities ($-OH_2^+$, $-COO^-$, $-NH_3^+$, …) concentrate the solute-solvent interaction and present the largest $\Delta G_{\text{solv}}^{A}$ energies. For example, the carboxylate group in the acrylic acid gives a $\Delta G_{\text{solv}}^{COO^-}$ term of -67.1 kcal/mol while the calculated hydration energy is -71.7 kcal/mol. However, the global charge of the ionic species also affects the atomic $\Delta G_{\text{solv}}^{A}$ contributions of non-polar sites that tend to have values from ±5 to ±20 kca/mol, the greater contributions occurring at the vicinal positions with respect to the charged groups. With respect to the neutral solutes, the $\Delta G_{\text{solv}}^{A}$

energies and the Bader charges q^A are distributed over much wider ranges so that a stronger dependency can be expected. As a matter of fact, linear regression in the (ΔG_{solv}^A, q^A) data set derived from *all* the anionic or cationic solutes results in overall R^2 coefficients of 0.98 and 0.97 for cations and anions, respectively (see Figure 4). The COSMO-HF q^A values were used in this correlation analysis, but nearly identical statistical parameters are obtained if the gas-phase HF charges are adopted instead. Therefore, we conclude that the ΔG_{solv}^A / q^A relationship would be transferable for other ions and that the atomic charge distribution closely determines the hydration of the ionic solutes. We also note that, for the slightly polar or non-polar solutes, the electrostatic ΔG_{solv}^A values would probably be controlled by other multipolar terms (dipole, quadrupole, …) associated to the solute charge density.

Assessment of atomic/fragment contributions to solvation energy

Inspection of the ΔG_{solv}^A energies in Figures 2-3 shows that the solvation contributions of two covalently bonded atoms have opposite signs as they usually have opposite q^A charges too. In the case of *A-B* polar bonds that imply a significant charge separation, the *A* and *B* contributions can be comparable to the total ΔG_{solv} and even larger for ionized groups. For example, the ammonium group -NH_3^+ in 3-chloroaniline (see Figure 3) has a ΔG_{solv}^{NH3+} of -20.7 kcal/mol arising from the sum of the N (+72.3 kcal/mol) and H terms (-31.1, -31.1 and -30.8). Therefore, we believe that the best strategy for standardizing and analyzing the IQA decomposition of the solvation energy would consist of adopting a *united atom* approach in such a way that H contributions in -XH_n fragments are merged with that of the heavy atom X. Subsequently, the resulting united atom X' groups can be useful to define a set of atom types and functional groups into which a given organic molecule can be formally decomposed.

It may be interesting to note that the *united atom* approach suggested by the IQA analysis has been used in empirical solvation and/or *logP* methods like the atom/fragment contribution method of Meyland and Howard,[23] which defines 130 fragments to estimate the *logP* of organic compounds. This

model employs multiple linear regression of thousands of compounds to estimate the *logP* in terms of fragment contributions (f_G) using the following equation,

$$\log P = \sum_{G} f_G n_G + \sum_{I} c_I n_I + b \qquad (10)$$

where b is a regression parameter, n_G is the number of times that a group occurs in the structure and c_I are specific correction terms that apply only for a subset of fragment combinations involving aromatic ring substituents, ring strain, electronic conjugation, etc. Thus, in this scenario, the f_G terms could be somehow related to the IQA $\Delta G_{\text{solv}}{}^{G}$ terms.

Table 1 lists the 51 atom types/functional groups (G) that were selected for analyzing the fragment-based IQA contributions to the electrostatic solvation energy. The mean values and standard deviations of the corresponding $\Delta G_{\text{solv}}{}^{G}$ energies are also collected in Table 1. We decided to analyze the fragment contributions of groups that appear at least 5 times in the set of MNSol structures. For this reason, several groups appearing in the solute molecules (*e.g.*, thiol –SH, phosphate –PO$_4$, etc.) are not included in Table 1. Therefore, our analysis is not exhaustive and is not oriented to derive a working additive model of hydration energies for organic molecules, but to find out whether or not we can extract useful information from the IQA decomposition.

Table 1. Selected atom/functional groups (*G*) for the analysis of electrostatic solvation. The number of ΔG_{solv}^{G} values (*n*), their mean value (μ in kcal/mol), standard deviation (σ), skewness (skw) and excess kurtosis (kurt) are given. The symbol of groups belonging to anionic/cationic compounds is augmented by (-)/(+), respectively. Charged groups are denoted by placing their sing into square brackets ([+]/[-]).

Atom/ Group	Description	*n*	μ	σ	Skw	Kurt
-Br	bromine	29	-4.7	3.0	0.0	-1.1
=C<	sp^2 C	10	0.9	1.4	0.2	-1.7
>C<	sp^3 C	39	-2.2	1.5	0.0	-1.2
C(Ar)	aromatic C	150	-0.8	1.0	0.0	0.4
C(Ar)(-)	aromatic C	25	25.5	27.9	-0.1	-1.8
C(Ar)(+)	aromatic C	24	-3.5	4.3	-1.4	1.7
=CH-	CH (sp2 C)	25	0.7	2.0	1.5	1.8
>CH-	CH (sp3 C)	28	0.2	1.2	0.1	-0.3
=CH$_2$	terminal CH$_2$ (sp2 C)	13	-0.7	1.6	1.1	0.2
-CH$_2$-	methylene	316	0.6	1.4	0.8	0.9
=CH$_2$(-)	terminal CH$_2$ (sp2 C)	6	-8.8	9.4	-0.3	-2.0
-CH$_2$-(-)	methylene	33	22.1	20.9	-0.2	-1.7
-CH$_2$-(+)	methylene	53	-22.8	11.8	0.3	-0.9
-CH$_3$	methyl	283	0.4	1.1	1.1	1.0
-CH$_3$(-)	methyl	35	-4.0	4.1	0.4	-0.2
-CH$_3$(+)	methyl	52	-24.0	15.2	-0.6	-1.0
CH(Ar)	aromatic CH	379	0.2	1.5	0.4	1.5
CH(Ar)(-)	aromatic CH	66	-2.6	3.4	0.5	-0.8
CH(Ar)(+)	aromatic CH	73	-8.9	5.0	-0.6	1.1
-CHO	aldehyde	7	-9.4	2.7	-0.4	-0.9
>CH-(-)	CH (sp3 c)	7	24.9	31.7	-0.6	-1.4
-Cl	chlorine	89	-2.1	2.4	-0.4	-0.7
-Cl(-)	chlorine	8	-17.4	4.1	0.3	-1.8
-CN	cyanide	8	-12.5	1.5	-0.3	-1.0
-CO-	carbonyl	44	-11.3	2.0	0.0	0.4
-CONH-	amide	5	-12.0	1.9	0.5	-1.9

-CONH2	terminal amide	7	-12.3	2.6	0.1	-1.7
-COO-	ester	9	-9.3	0.3	0.2	-1.4
-COO[-]	carboxylate	8	-62.2	8.4	0.2	-1.9
-COOH	carboxyl	8	-10.1	1.2	0.6	-0.7
>C<(-)	sp3 c	7	95.0	5.5	0.3	-1.4
-F	fluorine	92	-0.3	1.6	-1.1	0.5
-F(-)	fluorine	21	-36.8	2.2	0.0	-1.1
>N-	tertiary amine	7	-2.3	1.3	1.0	-0.7
N(Ar)	aromatic n	30	-6.8	4.7	1.0	0.0
>NH-[+]	tertiary ammonium	5	46.6	4.1	0.7	-1.4
-NH-	secondary amine	11	-3.4	0.9	0.8	-0.1
-NH$_2$	primary amine	17	-3.5	0.7	0.4	-0.7
-NH$_2$-[+]	secondary ammonium	12	17.5	1.1	-0.5	-0.1
-NH$_3$[+]	primary ammonium	14	-20.5	1.3	-1.1	0.5
NH(Ar)	aromatic NH	25	3.3	2.5	-0.1	-1.2
-NHCONH$_2$	carbamide	5	-11.0	1.9	-0.3	-2.2
-NO$_2$	nitro	13	-8.1	1.4	-0.3	-0.4
-O-	ether	40	-2.1	2.9	0.6	-0.7
-O[-]	alkoxide	27	-96.4	9.1	0.0	-0.9
-OH	alcohol	26	-4.3	1.9	0.5	-0.2
-O(H)[-]	H-bonded alkoxide	10	-89.5	5.2	0.6	-0.4
-ONO$_2$	nitrate	5	-3.8	2.2	-0.2	-2.0
-S-	sulfide	5	-6.8	2.1	-0.1	-2.2
-SS-	disulfide	11	1.7	7.1	-0.5	-1.4

The selection and notation of the groups listed in Table 1 is partially based on the fragments defined by Meyland and Howard[23] in their empirical *logP* method. For example, the symbol ">C<" stands for a tetra-substituted *sp³* C atom, -CH$_2$- for a methylene group, CH(Ar) for an aromatic CH fragment, -NH$_2$ for a neutral primary amine group, -NH$_2$-[+] for a protonated secondary amine, and so on. A particular feature of our fragment selection is that, for the majority of neutral fragments *G*, we also distinguish two ionic versions labelled as *G*(-) and *G*(+) depending on the global charge of the molecule

bearing the corresponding fragment. For instance, -CH$_2$-(+) and -CH$_2$-(-) stand for the methylene groups in cationic and anionic molecules, respectively. This distinction is due to the significant variation in the $\Delta G_{\text{solv}}{}^G$ values due to the global charge of the molecule and the relative positioning of the G group with respect to the ionic groups (see below).

Besides the mean values (μ) of $\Delta G_{\text{solv}}{}^G$ and their standard deviations (σ), Table 1 contains the skewness (*skw*) and excess kutorsis indexes that measure, respectively, the asymmetry and the shape of the peak and tails of the underlying distribution with respect to the normal distribution (skw=0 and kurt=0). The $\Delta G_{\text{solv}}{}^G$ data corresponding to fragments in neutral molecules result in more or less narrow distributions (σ= ~2.0 kcal/mol) that are moderately asymmetrical, albeit with varying shapes as the kurtosis coefficient can be positive or negative. In the case of ionic solutes, both the ionic functionalities and the neutral fragments have $\Delta G_{\text{solv}}{}^G$ distributions that are quite wide (e.g., σ= ~9 kcal/mol and above in anionic systems) and predominantly flat (*i.e.*, negative excess kurtosis).

To better characterize the width and shape of the $\Delta G_{\text{solv}}{}^G$ distributions, Figure 5 displays the histogram plots of a few selected groups. Closer inspection of the IQA data can reveal how the features of the various distributions are related to structural patterns. For example, electrostatic solvation of methylene (-CH$_2$-) groups is, on average, not favorable and the distribution is quite concentrated around its mean value (μ=0.6, σ=1.4). The tallest distribution peak corresponds to the solvation of – CH$_2$- fragments attached to zero or one polar groups while the shoulder at $\Delta G_{\text{solv}}{}^G$ =+3.0 kcal/mol is due to di-substituted X-CH$_2$-Y groups with X, Y=polar.

Figure 5. Histogram of selected fragment-based IQA ΔG_{solv}^G values.
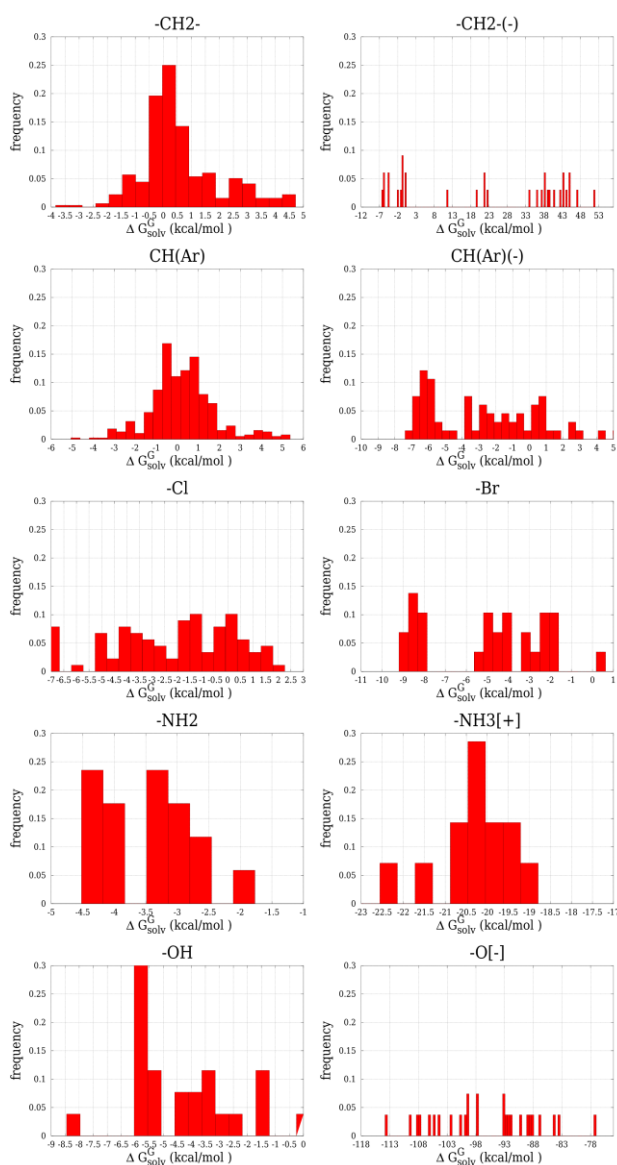
The presence of a global negative charge dramatically changes the ΔG_{solv}^G distribution, which becomes quite wide and flat over a 40 kcal/mol interval (see –CH$_2$-(-) in Figure 5). Thus, all the –CH$_2$- fragments attached to ionic groups (e.g., -NH$_3^+$) have negative contributions to ΔG_{solv} (from -40 to -20 kcal/mol). As a matter of fact, the solvation stabilization of tetrahedral ammonium groups stems from the positive ΔG_{solv}^A of the N atom and the negative (stabilizing) terms from the four surrounding groups including H atoms and -CH$_2$- fragments (see N-ethylethanamine in Figure 3). Other methylene groups located at the β or γ positions with respect to the positively charged group have also negative ΔG_{solv}^G

values (between -12 and -3 kcal/mol), which are below the values for –CH$_2$- in neutral molecules. The CH(Ar) fragments in neutral solutes give a hydrophobic contribution (μ=0.2, σ=1.5) similar to that of the –CH$_2$- groups. Most of the ΔG_{solv}^G data come from benzene ring groups although some CH(Ar) groups in heterocycles appear at the positive tail. The alkoxide –O$^-$ and carboxylate –COO$^-$ substituents at phenyl rings modify the hydration of the CH(Ar) groups, that become hydrophilic sites with negative ΔG_{solv}^G values. However, the aromatic CH groups are less perturbed by the ionic substituents than the aliphatic –CH$_2$- fragments. The largest effect is at CH sites located at *ortho-* positions with $\Delta G_{\text{solv}}^G \sim$ -7/-8 kcal/mol while those at the *meta-* and *para-* have values of $\sim$-2/-3 kcal/mol.

Other examples of fragment hydration energy distributions in Figure 5 are those of the halogen substituents (Cl and Br), which exhibit quite wide and flat histograms (μ=-2.1, σ=2.4 for Cl and μ=-4.7, σ=3.0 for Br). In fact a detailed examination shows that the more negative ΔG_{solv}^G values occur at monosubstituted aliphatic R-X compounds, the least favorable terms are those of polisubstituted halogenated alkanes and the values in between correspond to aromatic molecules. Neutral polar groups like –NH$_2$, -OH as well as the cationic –NH$_3^+$ tend to give relatively concentrated ΔG_{solv}^G distributions so that their corresponding average values μ may be reliable estimators. In sharp contrast, anionic groups like the alkoxide –O$^-$ result in scattered data over 40 kcal/mol interval (see Figure 5). In this case the more negative ΔG_{solv}^G values for –O$^-$ are due to phenolate groups ($\sim$ -110 kcal/mol).

Additivity of fragment contributions

The analysis of the histograms in Figure 5 and other data in the Supporting Information confirms that the IQA decomposition of the COSMO-HF solvation energies can provide a detailed assessment of the fragment contributions to solvation. Although the amount of data gathered for some of the examined functionalities is limited, several structure-activity trends can be outlined regarding the constancy/dispersion of fragment contributions and their relationship with structural and substituent effects. In particular, the additivity of the mean values of the fragment ΔG_{solv}^G energies to estimate the

total solvation energy is not unreasonable for neutral solute molecules given that the dispersion (σ) values collected in Table 1 tend to be moderate (< 2 kcal/mol).

To find out to what extent the mean $\Delta G_{solv}{}^G$ values are additive, we calculated the COSMO HF/aug-cc-pVTZ solvation energy for a set of 32 MNSol molecules not considered in the former IQA calculations. All of these molecules are neutral and possess the functional groups listed in Table 1 (see Figure S2 in the Supporting Information). In Figure 6 the calculated ΔG_{solv} and the fragment-based estimations $\Delta G_{estimated} = \sum_G \left\langle \Delta G_{solv}^G \right\rangle$ are compared. The computed and estimated values show a moderate correlation ($R^2 = 0.70$) and the RMS error is 5.9 kcal/mol. The largest discrepancies arise in compounds that have large aliphatic or aromatic moieties (*e.g.*, octanol $\Delta G_{calc} = -6.0$, $\Delta G_{estimated} = 0.2$) whereas for molecules with 2 or more polar groups the simple additive model tends to work better (*e.g.*, 3-methylthiophenylurea $\Delta G_{calc} = -17.0$, $\Delta G_{estimated} = -18.0$).

Some empirical models developed for *logP* or hydration energy estimations exhibit a better performance ($R^2 \sim 0.8$-0.9).[22-23, 36] They include, not only atomic solvation parameters, but also exposure factors and/or correcting terms that modulate the sum of atomic terms. The results shown in Figure 6 suggest that a fragment-based method including weighing parameters to be fitted against a large set of hydration energies, could be also a reasonable approach. Further improvements in the additivity of the IQA-based solvation energies could be gained by defining new atom types/functional groups as suggested by the detailed analysis of histogram data. This could be the case of the halogen atom types that may be categorized as halogen attached to either aliphatic- or aromatic-C atom. These results also indicate that the electrostatic contributions of aliphatic and aromatic sites (and eventually their non-polar terms too) should receive a special attention. Thus, correction factors could be derived to take into account the influence of ionic/polar groups on the contributions of the nearby aliphatic/aromatic fragments. Nevertheless, these and other possible alternatives are beyond the scope

of the present work, which is focused on the description of the IQA-decomposition of solvation energy rather than in the development of a fragment solvation method.

Figure 6. Comparison between the COSMO-HF/aug-cc-pVTZ calculated hydration energies (ΔG_{calc} in kcal/mol) and the estimated ones assuming the additivity of fragment contributions ($\Delta G_{estimated}$) of the selected MNSol structures in aqueous solvent. The determination coefficient (R^2), the Spearman correlation coefficient (ρ) and the root mean square (RMS) error in kcal/mol are also indicated. The dashed line is the least squared fit line between the calculated and the estimated data.
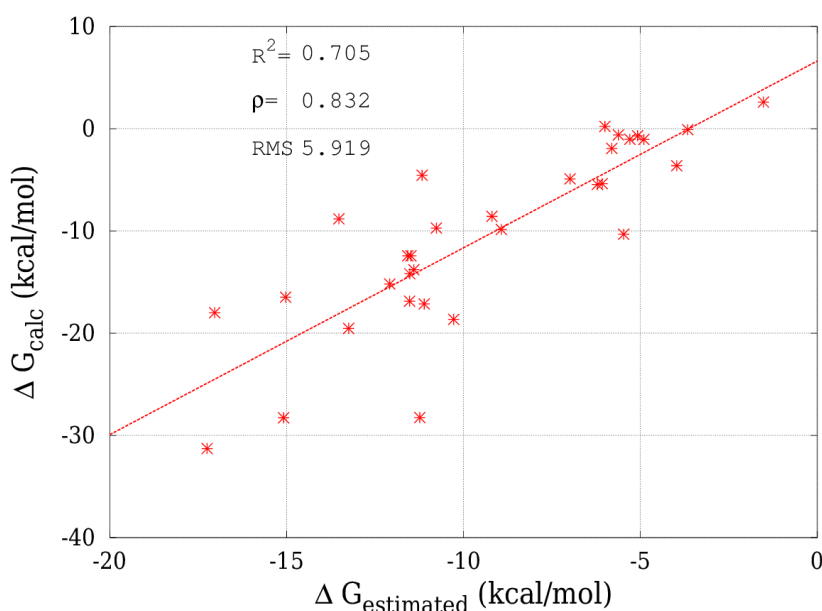
Coulomb and Polarization Effects

The Coulomb contribution ΔG_{solv}^{Coul} to the hydration energies of the MNSOL structures examined in this work was computed by means of single-point HF/aug-cc-pVTZ COSMO calculations on the gas-phase geometries and using the unpolarized (gas-phase) wavefunction. The polarization term is then obtained by substraction, $\Delta G_{solv}^{pol} = \Delta G_{solv} - \Delta G_{solv}^{Coul}$. Herein, we briefly comment on the results (see also Supporting Information).

For the neutral solutes, the ΔG_{solv}^{Coul} values correlate with experimental data (R^2=0.83 and RMS error 2.3 kcal/mol) similarly as the ΔG_{solv} energies do. The cationic hydration energies are also similar. In

consonance with expectation, the lack of polarization in the ΔG_{solv}^{Coul} values of the anionic solutes results in less correlation (R^2=0.78) with experimental data and larger absolute errors (~12 kcal/mol). Concerning the stability gained by solute-solvent mutual polarization, most of the calculated ΔG_{solv}^{pol} energies have values around -1, -2 kcal/mol. Only in those molecules containing the most polarizable groups (Cl, Br, anions, …), ΔG_{solv}^{pol} have values between -2 and -7 kcal/mol.

We also examined the distribution of group contributions, $\Delta G_{solv}^{Coul,\,G}$, using the same group definitions and statistical indexes as in Table 1 and Figure 5. In general, the dispersion (σ) of the $\Delta G_{solv}^{Coul,\,G}$ data for a given group is only slightly lower than that of ΔG_{solv}^{G}. Thus, the Coulomb $\Delta G_{solv}^{Coul,\,G}$ contributions depend on the chemical environment much to the same extent as ΔG_{solv}^{G}. More significant variations are observed in the shape of the ΔG_{solv}^{G} and $\Delta G_{solv}^{Coul,\,G}$ distributions characterized in terms of the skewness and excess kurtosis. In particular, several non-polar groups (-CH$_2$-, -CH$_3$, CH(Ar)) in neutral molecules, which have very small group contributions to ΔG_{solv}, exhibit sharp-peaked distributions of their $\Delta G_{solv}^{Coul,\,G}$ values having also a large excess kurtosis. As a consequence, the corresponding polarization group components $\Delta G_{solv}^{pol,\,G}$ are slightly positive (+0.5,+0.8, +0.7), what is more a statistical artefact than a physical effect. The rest of functional groups show negative stabilizing $\Delta G_{solv}^{pol,\,G}$ values. Finally, we also examined the additivity of the mean $\Delta G_{solv}^{Coul,\,G}$ values, finding a worse correlation with the calculated Coulomb solvation energies than in the case of the full solvation energies. Overall, we conclude that the separate treatment of the Coulomb and polarization electrostatic effects does not lead to an improved description of the fragment contributions to solvation energies.

About the extension of IQA to other implicit solvent models

The present COSMO-IQA calculations indicate that QM energies in implicit solvent are prone to be decomposed within the IQA method. COSMO belongs to the family of ASC methods that express the electrostatic solute-solvent energy as a single sum, $V_{solv} = \sum_k \Phi(\mathbf{s}_k) q_k$, involving the solute electrostatic potential and the ASCs. As mentioned above, this feature enables IQA to absorb solvent effects into

the net atomic energies. Hence, it can be reasonably expected that other ASC methods including the original Polarizable Continuum Model (PCM) [28]and the integral-equation-formalism PCM (IEFPCM)[37] could be similarly coupled with IQA provided that q_k and Φ data are available.

IQA extensions to other QM SCRF methods as the generalized Born theory (GB) and the multipolar expansion methods would be more problematic. On one hand, the GB approach[38] uses a modified and further parameterized Coulombic potential and evaluates the solvation energy as the total Coulomb-like interaction over the atom pairs in the solute molecule and, therefore, it is not evident how to decompose it into meaningful atomic contributions. On the other, the multipolar expansion of the solute charge density can be extended to atom-centered expansions although there is an infinite number of manners of weighting the multipoles. [39-40] However, the resulting solute-solvent interaction energy is written in terms of diatomic reaction potential terms that do not admit an evident atomic partitioning either.

In this work, we focus on the partitioning of the electrostatic solvation energy. However, the consideration of non-polar solvation effects within the IQA-like analysis could be feasible by means of empirical approaches. For example, cavitation free energy, solute-solvent dispersion and solvent-structural effects can be accounted for by means of the G_{CDS} empirical potential implemented in the SMD solvation method[41]:

$$G_{CDS} = \sum_A \left(\gamma_A + \gamma^M \right) \sigma_A$$

where γ_A and γ^M are molecular surface tension parameters and σ_A is the solvent-accessible surface of atom A. Hence, the atomic contributions to G_{CDS} could be combined with the electrostatic $\Delta G_{solv}{}^A$ terms to yield an atomic mapping of the total solvation energy.

Conclusions

The computational results presented in this work demonstrate that it is feasible to incorporate electrostatic solute-solvent effects into the IQA energy decomposition method. In this way the usual IQA analysis of energy differences can be carried out including continuum solvent effects, extending thus its applicability. Basing on the extensive solvation energy calculations followed by the IQA decomposition of the electrostatic solvation energy, we have also shown that IQA yields a detailed atomic mapping of solvation energies and suggests a united atom approach for considering fragment contributions. A tentative selection of fragments has been made and their solvation energies have been characterized statistically, finding that the distributions of fragment solvation energies, which may have relatively large deviation for some groups, depend on structural and substituent effects. For neutral molecules, the simple additivity assumption, commonly adopted in empirical solvation methods, leads to approximate solvation energies that exhibit only moderate correlation with reference values and have significant errors of several kcal/mol. More specific fragment-types and extra-parameters would be required to derive improved fragment solvation methods from QM SCRF and IQA calculations on a larger database of solute structures.

Computational Section

QM calculations

Cartesian coordinates and reference hydration energies for all the solute molecules were retrieved from the MNSol database. The general ab initio quantum chemistry program GAMESS[42-43] was used to perform all the QM calculations. First, we relaxed all the structures by means of unconstrained energy minimizations that were started from the corresponding MNSol geometries. These calculations were carried out first in the gas-phase combining the Hartree-Fock (HF) method with the aug-cc-pVTZ basis set.[44-45] The solute geometries were also optimized at the HF/aug-cc-pVTZ level in combination with the COSMO solvation model.[30] A dielectric $\varepsilon=80$ was selected for mimicking water as solvent while a multiplicative factor of 1.2 was applied to the standard van der Waals radii for cavity

construction. Each atomic sphere that contributes to build the molecular cavity is divided into 92 tesserae. The cosprt module in GAMESS was locally modified to print out the charge and Cartesian coordinates of the apparent surface charges of the optimized HF-COSMO structures. The Chimera visualization system[46] was used to draw the models of the solute molecules.

IQA calculations

The IQA decomposition of the total energies was performed with a modular version of the PROMOLDEN program [47] that is being developed in our laboratory. In this version, the program reads the apparent surface charge data generated by the COSMO implementation in GAMESS in order to compute the solute-solvent interaction term V_{solv} using the same integration algorithm that is employed for computing the electron-nucleus interaction terms V_{en}.[2]

The IQA quantities are numerically integrated by PROMOLDEN over finite and irregular integration domains (i.e.,atomic basins Ω_A) using angular and radial grids in atomic spherical quadratures that are much finer than those typically used by other QM software. [2, 34] We employed similar integration settings to those used in previous work[6] and that represent a compromise choice between computational cost and accuracy for small and medium-sized molecules. Thus, a β−sphere around each atom was considered (i.e., a sphere completely contained inside the atomic basin), with a radius equal to 60 % the distance of its nucleus to the closest bond critical point in the electron density. High-quality Lebedev angular grids were used with 5810 and 974 points outside and within the β-spheres of heavy atoms, respectively, (3890 and 590 points for hydrogen atoms). Euler-McLaurin radial quadratures were employed with 512 and 384 radial points outside and inside the β−spheres of heavy atoms, respectively (384 and 256 points for H). The largest value of the radial coordinate in the integrations was 15.0 au for heavy atoms (10.0 au for H atoms). Maximum angular moments, λ_{max}, of 10 and 6 were assigned to the Laplace and bipolar expansions of $1/r_{12}$ outside and within the β-spheres.

References

1.   Popelier, P. L. A., Quantum Chemical Topology. In *The Chemical Bond Ii: 100 Years Old and Getting Stronger*, Mingos, D. M. P., Ed. Springer International Publishing: Cham, 2016; pp 71-117.

2.   Blanco, M. A.; Martín-Pendás, A.; Francisco, E., Interacting Quantum Atoms: A Correlated Energy Decomposition Scheme Based on the Quantum Theory of Atoms in Molecules. *J. Chem. Theory Comput.* **2005,** *1*, 1096-1109.

3.   Francisco, E.; Martín-Pendás, A.; Blanco, M. A., A Molecular Energy Decomposition Scheme for Atoms in Molecules. *J. Chem. Theory Comput.* **2006,** *2*, 90-102.

4.   Grimme, S., Density Functional Theory with London Dispersion Corrections. *WIREs Comput Mol Sci* **2011,** *1*, 211-228.

5.   Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C., Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016,** *116*, 5105-5154.

6.   Suárez, D.; Díaz, N.; Francisco, E.; Martín-Pendás, A., Application of the Interacting Quantum Atoms Approach to the S66 and Ionic-Hydrogen-Bond Datasets for Noncovalent Interactions. *ChemPhysChem* **2018,** *19*, 973-987.

7.   Martín-Pendás, A.; Blanco, M. A.; Francisco, E., The Nature of the Hydrogen Bond: A Synthesis from the Interacting Quantum Atoms Picture. *J. Chem. Phys.* **2006,** *125*.

8.   Guevara-Vela, J. M.; Chavez-Calvillo, R.; Garcia-Revilla, M.; Hernandez-Trujillo, J.; Christiansen, O.; Francisco, E.; Martín-Pendás, A.; Rocha-Rinza, T., Hydrogen-Bond Cooperative Effects in Small Cyclic Water Clusters as Revealed by the Interacting Quantum Atoms Approach. *Chemistry-a European Journal* **2013,** *19*, 14304-14315.

9.   Guevara-Vela, J. M.; Romero-Montalvo, E.; Mora Gomez, V. A.; Chavez-Calvillo, R.; Garcia-Revilla, M.; Francisco, E.; Martín-Pendás, A.; Rocha-Rinza, T., Hydrogen Bond Cooperativity and Anticooperativity within the Water Hexamer. *Phys. Chem. Chem. Phys.* **2016,** *18*, 19557-19566.

10.   Romero-Montalvo, E.; Guevara-Vela, J. M.; Costales, A.; Martín-Pendás, Á.; Rocha-Rinza, T., Cooperative and Anticooperative Effects in Resonance Assisted Hydrogen Bonds in Merged Structures of Malondialdehyde. *Phys. Chem. Chem. Phys.* **2017,** *19*, 97-107.

11.   Bartashevich, E.; Troitskaya, E.; Martín-Pendás, A.; Tsirelson, V., Understanding the Bifurcated Halogen Bonding N Center Dot Center Dot Center Dot Hal Center Dot Center Dot Center Dot N in Bidentate Diazaheterocyclic Compounds. *Computational and Theoretical Chemistry* **2015,** *1053*, 229-237.

12.   Tiana, D.; Francisco, E.; Macchi, P.; Sironi, A.; Martín-Pendás, A., An Interacting Quantum Atoms Analysis of the Metal-Metal Bond in M-2(Co)(8) (N) Systems. *J. Phys. Chem. A* **2015,** *119*, 2153-2160.

13.   Cukrowski, I.; de Lange, J. H.; Mitoraj, M., Physical Nature of Interactions in Zn [Ii] Complexes with 2,2′-Bipyridyl: Quantum Theory of Atoms in Molecules (Qtaim), Interacting Quantum Atoms (Iqa), Noncovalent Interactions (Nci), and Extended Transition State Coupled with Natural Orbitals for Chemical Valence (Ets-Nocv) Comparative Studies. *J. Phys. Chem. A* **2014,** *118*, 623-637.

14.   Wilson, A. L.; Popelier, P. L., Exponential Relationships Capturing Atomistic Short-Range Repulsion from the Interacting Quantum Atoms (Iqa) Method. *J Phys Chem A* **2016,** *120*, 9647-9659.

15. McDonagh, J. L.; Silva, A. F.; Vincent, M. A.; Popelier, P. L. A., Quantifying Electron Correlation of the Chemical Bond. **2017,** *8*, 1937-1942.

16. Cramer, C. J.; Truhlar, D. G., Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999,** *99*, 2161-2200.

17. Tomasi, J., Thirty Years of Continuum Solvation Chemistry: A Review, and Prospects for the near Future. *Theoretical Chemistry Accounts* **2004,** *112*, 184-203.

18. Ben-Naim, A., *Statistical Thermodynamics for Chemists and Biochemists*. Springer Science Business Media: New York, 1992.

19. Park, H., Extended Solvent-Contact Model Approach to Sampl4 Blind Prediction Challenge for Hydration Free Energies. *J. Comput. Aided Mol. Des.* **2014,** *28*, 175-186.

20. Bernazzani, L.; Duce, C.; Micheli, A.; Mollica, V.; Tiné, M. R., Quantitative Structure−Property Relationship (Qspr) Prediction of Solvation Gibbs Energy of Bifunctional Compounds by Recursive Neural Networks. *Journal of Chemical & Engineering Data* **2010,** *55*, 5425-5428.

21. Kang, H.; Choi, H.; Park, H., Prediction of Molecular Solvation Free Energy Based on the Optimization of Atomic Solvation Parameters with Genetic Algorithm. *J. Chem. Inf. Model.* **2007,** *47*, 509-514.

22. Jianfeng, P.; Qi, W.; Jiaju, Z.; Luhua, L., Estimating Protein–Ligand Binding Free Energy: Atomic Solvation Parameters for Partition Coefficient and Solvation Free Energy Calculation. *Proteins* **2004,** *57*, 651-664.

23. Meylan, W. M.; Howard, P. H., Atom/Fragment Contribution Method for Estimating Octanol-Water Partition Coefficients. *J Pharm Sci* **1995,** *84*, 83-92.

24.   Soteras, I.; Morreale, A.; López, J. M.; Orozco, M.; Luque, F. J., Group Contributions to the Solvation Free Energy

from Mst Continuum Calculations. *Brazilian Journal of Physics* **2004,** *34*, 48-57.

25.   Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G., Minnesota Solvation Database-Version 2012. Minnesota, U. o., Ed. Minneapolis, 2012.

26.   Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G., Predicting Aqueous Free Energies of Solvation as Functions of Temperature. *J. Phys. Chem. B* **2006,** *110*, 5665-5675.

27.   Francisco, E.; Martín-Pendás, A., Energy Partition Analyses: Symmetry-Adapted Perturbation Theory and Other Techniques. In *Non-Covalent Interactions in Quantum Chemistry and Physics. Theory and Applications*, Otero-de-la-Roza, A.; DiLabio, G., Eds. Elsevier: 2017; pp 27-62.

28.   Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005,** *105*, 2999-3094.

29.   Tomasi, J.; Persico, M., Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994,** *94*, 2027-2094.

30.   Baldridge, K.; Klamt, A., First Principles Implementation of Solvent Effects without Outlying Charge Error. *J. Chem. Phys.* **1997,** *106*, 6622-6633.

31.   Klamt, A., *Cosmo-Rs from Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design* Elsevier: Amsterdam, 2005.

32.   Klamt, A.; Jonas, V., Treatment of the Outlying Charge in Continuum Solvation Models. *J. Chem. Phys.* **1996,** *105*, 9972-9981.

33.   Tielker, N.; Tomazic, D.; Heil, J.; Kloss, T.; Ehrhart, S.; Güssregen, S.; Schmidt, K. F.; Kast, S. M., The Sampl5 Challenge for Embedded-Cluster Integral Equation Theory: Solvation Free Energies, Aqueous Pk a, and Cyclohexane–Water Log D. *J. Comput. Aided Mol. Des.* **2016,** *30*, 1035-1044.

34.   Martín-Pendás, A.; Blanco, M. A.; Francisco, E., Two-Electron Integrations in the Quantum Theory of Atoms in Molecules. *J. Chem. Phys.* **2004,** *120*, 4581-4592.

35.   Martín-Pendás, A.; Francisco, E.; Blanco, M. A., Two-Electron Integrations in the Quantum Theory of Atoms in Molecules with Correlated Wave Functions. *J. Comput. Chem.* **2005,** *26*, 344-351.

36.   Chung, K.-C.; Park, H., Accuracy Enhancement in the Estimation of Molecular Hydration Free Energies by Implementing the Intramolecular Hydrogen Bond Effects. *Journal of Cheminformatics* **2015,** *7*, 57.

37.   Cancès, E.; Mennucci, B.; Tomasi, J., A New Integral Equation Formalism for the Polarizable Continuum Model: Theoretical Background and Applications to Isotropic and Anisotropic Dielectrics. *J. Chem. Phys.* **1997,** *107*, 3032-3041.

38.   Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T., Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *JACS* **1990,** *112*, 6127-6129.

39.   Rinaldi, D.; Bouchy, A.; Rivail, J.-L.; Dillet, V., A Self-Consistent Reaction Field Model of Solvation Using Distributed Multipoles. I. Energy and Energy Derivatives. *J. Chem. Phys.* **2004,** *120*, 2343-2350.

40.   Rinaldi, D.; Bouchy, A.; Rivail, J.-L., A Self-Consistent Reaction Field Model of Solvation Using Distributed Multipoles. Ii: Second Energy Derivatives and Application to Vibrational Spectra. *Theoretical Chemistry Accounts* **2006,** *116*, 664-669.

41.   Cramer, C. J.; Truhlar, D. G., A Universal Approach to Solvation Modeling. *Acc. Chem. Res.* **2008,** *41*, 760-768.

42.   Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993,** *14*, 1347-1363.

43.   Gordon, M. S.; Schmidt, M. W., Chapter 41 - Advances in Electronic Structure Theory: Gamess a Decade Later A2 - Dykstra, Clifford E. In *Theory and Applications of Computational Chemistry*, Frenking, G.; Kim, K. S.; Scuseria, G. E., Eds. Elsevier: Amsterdam, 2005; pp 1167-1189.

44.   Jr., T. H. D., Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989,** *90*, 1007-1023.

45.   Kendall, R. A.; Dunning Jr., T. H.; Harrison, R. J., *J. Chem. Phys.* **1992,** *96*, 6796-6802.

46.   Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., Ucsf Chimera--a Visualization System for Exploratory Research and Analysis. *J Comput Chem* **2004,** *25*, 1605-1612.

47.   Martín-Pendás, A.; Francisco, E. *Promolden: A Qtaim /Iqa Code*, Unpublished: 2015.

A Quantum Chemical Topology Picture of Intermolecular Electrostatic Interactions and Charge Penetration Energy

Fernando Jiménez-Grávalos* and Dimas Suárez*

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Based on the Interacting Quantum Atoms approach, we present herein a conceptual and theoretical framework of short-range electrostatic interactions, whose accurate description is still a challenging problem in molecular modeling. For all the noncovalent complexes in the S66 database, the fragment-based and atomic decomposition of the electrostatic binding energies is performed using both the charge density of the dimers and the unrelaxed densities of the monomers. This energy decomposition together with dispersion corrections gives rise to a pairwise approximation to the total binding energy. It also provides energetic descriptors at varying distance that directly address the atomic and molecular electrostatic interactions as described by point-charge or multipole-based potentials. Additionally, we propose a consistent definition of the charge penetration energy within quantum chemical topology, which is mainly characterized in terms of the intramolecular electrostatic energy. Finally, we discuss some practical implications of our results for the design and validation of electrostatic potentials.
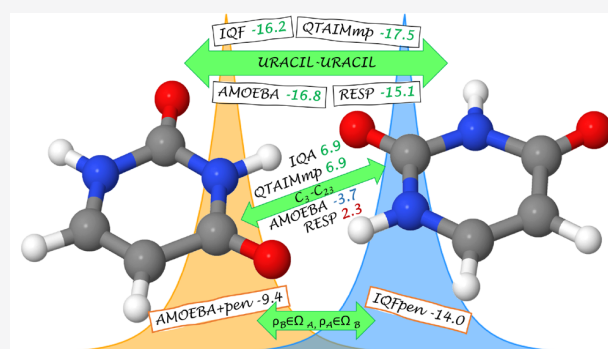
1. INTRODUCTION

Electrostatic interactions are central to molecular modeling because of their slow decay and strength. Especially when polar atoms or charged species are involved, they largely determine the stability and activity of biomolecules such as proteins, nucleic acids, or lipids, among others.[1,2] As such, a reliable description in molecular mechanics (MM) potentials is essential both in the short- and in the long-range.

Within the framework of MM methods, interactions comprising nonbonded atoms are usually represented by pairwise potentials such as the Lennard-Jones and the Coulomb ones. In the latter case, the use of point charges or higher order multipoles to avoid the integration of interacting charge densities has resulted in accurate electrostatics at long-range, with significant improvements to speed up and facilitate convergence such as the Ewald summation and its variants to perform, for example, molecular simulations in solution under periodic boundary conditions.[3–8] At short-range, however, the approximations taken for long distances become less accurate or invalid,[9] and a correct electrostatic description in this regime stills poses a challenge. Hence, there is a growing interest in improving short-range electrostatics (e.g., for troublesome hydrogen bonds), mainly focused on capturing the effects associated with the non-negligible interpenetration of densities, leading to the so-called charge penetration (CP) energy, which is typically defined as the difference between the electrostatic energy computed from continuous charge density distributions and that from multipolar approximations.[10] Thus,

several investigations have been devoted in the last years to incorporate the charge penetration energy into the MM electrostatic potentials.[10–14]

The separation of various energy terms as implemented in the MM potentials is somehow paralleled by the energy decomposition analysis (EDA) methods.[15] A major goal of any EDA approach is to ascertain the nature and type of the interactions among molecules as well as to rationalize their stabilizing or destabilizing roles, which may have implications for the design, parametrization, and validation of MM potentials such as the electrostatic ones. However, there is no unique recipe to decompose the energy, and thus many EDAs have been developed rooted in different approaches. Hence, symmetry-adapted perturbation theory (SAPT) makes use of a perturbative approach to differentiate the distinct nature of the intermolecular interactions,[16,17] while orbital-based EDAs exploit a stepped scheme to calculate the different energies according to some reference electronic states,[18–20] and the interacting quantum atoms (IQA) method relies on a real space partition of the quantum mechanical (QM) density

matrices,[21,22] being thus classified as a quantum chemical topology (QCT) method.

According to recent studies, in spite of their crude approximations, it may be feasible to improve the classical MM potentials by utilizing the information provided by EDAs.[23] More specifically, it has been shown that the SAPT energy components (electrostatics, induction, exchange-repulsion, and dispersion) can be modeled with relatively simple MM functions.[24,25] In particular, it has been demonstrated that the combination of empirical damping functions accounting for the CP energy with point multipoles results in electrostatic energies at short-range that are quite close to the SAPT ones. Actually, the SAPT electrostatic energy provides the required reference to parametrize and validate the CP-augmented potentials. However, different interpretations of short-range energetic effects involving the overlap of the electron densities of two or more fragments may be possible depending on the particular EDA of choice.[15] As such, other schemes such as the absolutely localized molecular orbital (ALMO) EDA, that relies on a different non-perturbative decomposition of energy terms, have also been proposed.[26] In this work, we reexamine the nature of electrostatic interactions under the prism of an orbital-invariant, reference-free technique. The IQA approach fulfills these requirements as it is a QCT, real-space energy decomposition resorting to the partition of the reduced density matrices (RDMs). IQA distinguishes not only between electrostatic or exchange-correlation components of the interaction energy but also between intra- or interatomic (or fragment) contributions. Moreover, since IQA splits the total energy of a system and not only the interaction between selected fragments, it is capable of reconstructing (or dissecting) the energy ascribed to both covalent and noncovalent binding, allowing thus covalent bond energies to be characterized[27] as well as the accuracy of the energy components handled by QM fragment methods to be investigated.[28]

Herein, we study in detail the electrostatic interactions involved in noncovalent complexes with a twofold goal. On the one hand, we aim to compare in a consistent and systematic manner the atomic and fragment contributions to the electrostatic energy as evaluated throughout a hierarchy of QM and MM approximations and at varying intermolecular distances. In this way, we seek to identify the best correspondence between the IQA and the MM electrostatic terms. On the other hand, we critically examine the CP concept and propose a novel definition relying on a joint orbital and real-space decomposition scheme, which can give new insight into the CP energy. To help fulfill these goals, the rest of the manuscript is structured as follows. First, we present and describe the theoretical scaffold that holds our work, paying particular attention to the IQA—and its IQF variant—energy decomposition, followed by subsections concerning the zeroth-order approximation, the electrostatic MM potentials, and a final assessment of the CP energy and the alternative definition proposed in this work. Subsequently, we describe some computational settings and the results of our test calculations, which were carried out on the S66 and S66x8 data sets.[29,30] The various levels of description of the electrostatic interactions are then discussed based on their statistical correlation with benchmark data, their dependence with the intermolecular separation, etc. The QM and IQA calculations yield further information, not only about the magnitude of the CP energy, but more importantly, about its different role in the IQA descriptors. Finally, we conclude that the aim of improving the electrostatic description is essentially fulfilled at the expense of accounting for intramolecular effects.

2. THEORY AND METHODS

2.1. IQA Decomposition of QM Energies. The interacting quantum atoms method is a robust and physically sound approach to decompose the total QM energy of a system into chemically meaningful components.[21,22] It is based on partitioning the first- and second-order RDMs, as can be done with the real space partition proposed by Bader and co-workers within their Quantum Theory of Atoms in Molecules (QTAIM).[31] Thus, the three-dimensional space is decomposed into atomic regions (Ω_I) as the attraction basins of the gradient field of the electron density.

Given a global energy E of a system, IQA permits its decomposition into atomic components and pair interaction energies according to

$$E = \sum_I E_{net}^I + \sum_{I<J} E_{int}^{IJ} \tag{1}$$

where E_{net}^I is called the net atomic energy and, under the Born–Oppenheimer approximation, represents the energy due to the kinetic energy of electrons plus all the interactions involved (i.e., electron–electron and electron–nucleus) inside the atomic basin of each atom I. Similarly, each E_{int}^{IJ} term comprises the interaction energy between the electrons (e) and nucleus (n) located in atom I with those ascribed to other atoms J, which can be separated into n–e, e–e, and n–n contributions.

In order to compute the potential energy, the pair density $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$ is required. This object can be split according to $\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \rho(\mathbf{r}_1)\rho(\mathbf{r}_2) + \rho_{xc}(\mathbf{r}_1, \mathbf{r}_2)$ in two contributions. On the one hand, $\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)$ represents a noncorrelated product of densities, whereas electron correlation is accounted for by the exchange-correlation (xc) density $\rho_{xc}(\mathbf{r}_1, \mathbf{r}_2)$. Accordingly, the total interaction energy between two atoms can be decomposed into a Coulomb or electrostatic term E_{ele}^{IJ} and a quantum mechanical exchange-correlation one E_{xc}^{IJ}:

$$E_{int}^{IJ} = E_{ele}^{IJ} + E_{xc}^{IJ} \tag{2}$$

the latter term comprising all the associated QM effects that other (e.g., perturbative) approaches identify separately as dispersion, charge-transfer, polarization, etc. However, such a decomposition of E_{int}^{IJ} into two terms is particularly relevant when assessing the nature of a given bond or interaction, since the electrostatic term is associated with ionicity and the exchange-correlation contribution with covalency.[22]

IQA admits the grouping of atomic terms into fragment contributions (e.g., functional groups, molecules). Thus, a fragment decomposition similar to eq 1 of a molecular aggregate constituted by two moieties A and B involves

$$E_{net}^A = \sum_{\substack{I\in A}} E_{net}^I + \sum_{\substack{J>I \\ I,J\in A}} E_{int}^{IJ} \tag{3}$$

$$E_{int}^{AB} = \sum_{\substack{I\in A \\ J\in B}} E_{int}^{IJ} \tag{4}$$

where E_{net}^B can be calculated analogously. For practical purposes, we use the IQA acronym to refer to the atomic analysis, whereas for its fragment version the term interacting quantum fragments (IQF) is preferred.

In a previous work,[32] it was shown that IQF may be useful to dissect the formation energy of noncovalent complexes. Moreover, the IQA/IQF terms can be augmented with Grimme's D3 dispersion correction[33] as combined with the Becke-Johnson damping function[34] to complement the DFT and HF descriptions. Using the IQF-D3 protocol, the formation energy of a two-fragment system AB given by the process $A + B \rightarrow A \cdots B$ is split as

$$\Delta E_{form} = \Delta E_{net}^A + \Delta E_{net}^B + E_{ele}^{AB} + E_{xc}^{AB} + E_{D3-disp}^{AB}$$
$$= E_{def}^A + E_{def}^B + E_{int}^{AB} \tag{5}$$

The deformation term E_{def}^A (E_{def}^B) in the above equation corresponds to the net energy variation ΔE_{net}^A (ΔE_{net}^B) of fragment A (B), whereas the interfragment interaction energy E_{int}^{AB} comprises the electrostatic (E_{ele}^{AB}), exchange-correlation (E_{xc}^{AB}), and dispersion $(E_{D3-disp}^{AB})$ energies between the two fragments, the latter being thus separated from the whole exchange-correlation one. Overall, the contribution of electrostatics and exchange-correlation to ΔE_{form} is split between the intrafragment deformation and the interfragment interactions.

2.2. Electrostatic Energy from Continuous Charge Densities. The purely electrostatic energy for a given system with total charge density $\rho(\mathbf{r})$ $(\rho(\mathbf{r}) \equiv \rho_{tot}(\mathbf{r}) = \sum_I Z_I \delta(\mathbf{r} - \mathbf{R}_I) - \rho_e(\mathbf{r}))$, including both the electron density $\rho_e(\mathbf{r})$ and the nuclear charges Z_I at positions $\mathbf{R}_I$ is readily computed using the Coulomb law,

$$E_{ele} = \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} \, d\mathbf{r}_1 \, d\mathbf{r}_2 \tag{6}$$

where, for the sake of simplicity, the electrostatic potential in this and the rest of the equations is expressed in atomic units. Interestingly, the QTAIM real space partition derived from $\rho_e(\mathbf{r})$ allows us to decompose the electrostatic energy at the atomic level,

$$E_{ele} = \sum_I \frac{1}{2} \int_{\Omega_I} d\mathbf{r}_1 \int_{\Omega_I} d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}}$$
$$+ \sum_{I<J} \int_{\Omega_I} d\mathbf{r}_1 \int_{\Omega_J} d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}}$$
$$= \sum_I E_{ele}^I + \sum_{I<J} E_{ele}^{IJ} \tag{7}$$

Similarly, the fragment-based decomposition can be readily accomplished in an analogous way, allowing thus the specific assessment of the electrostatic component of the formation energy ΔE_{ele} of a two-fragment system AB as

$$\Delta E_{ele} = \Delta E_{ele}^A + \Delta E_{ele}^B + E_{ele}^{AB} \tag{8}$$

where ΔE_{ele} is expressed in terms of two contributions, namely, the intrafragment variations of electrostatic energy in the formation process, ΔE_{ele}^A and ΔE_{ele}^B, and the interfragment electrostatic interaction, E_{ele}^{AB}. At this point, we note that although ΔE_{ele} is commonly termed as a classical electrostatic interaction energy, we will refer to it as the electrostatic

contribution to the formation energy of the A⋯B complex in order to help avoid confusions with the IQA/IQF interaction energy terms. When the charge density is constructed from the unrelaxed fragment densities as $\rho^0(\mathbf{r}) = \rho_A^0(\mathbf{r}) + \rho_B^0(\mathbf{r})$, the electrostatic contribution to the formation energy, which is named here as the zeroth-order energy ΔE_{ele}^0, equals the Coulomb interaction between the unrelaxed densities:

$$\Delta E_{ele}^0 = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho_A^0(\mathbf{r}_1)\rho_B^0(\mathbf{r}_2)}{r_{12}} \, d\mathbf{r}_1 \, d\mathbf{r}_2 \tag{9}$$

This energetic term corresponds to the so-called *first-order polarization energy* (or simply *electrostatic energy*) defined in SAPT,[16] which has been adopted as a benchmark electrostatic energy for the validation of recently developed short-range electrostatic potentials.

2.3. Electrostatic Potentials in Molecular Mechanics. To avoid the usage of continuous charge distributions, the MM methods typically invoke the multipolar expansion as detailed in the Supporting Information (SI), which approximates the zeroth-order energy defined in eq 9. Formally, the multipolar electrostatic energy $\Delta E_{ele,mp}^0$ is affected by two different error sources. On the one hand, the underlying expansion must be truncated at some order ($l_{max} = 0, 1, 2, ...$), resulting thus in a certain truncation error. On the other, when $\rho_A^0(\mathbf{r})$ and $\rho_B^0(\mathbf{r})$ overlap to a significant extent, the rigorous application of the multipole expansion is impeded so that its usage at short distances implies some charge penetration error, which is normally assumed to be dominant. Nevertheless, the multipole-based potentials are still largely useful in many cases, and they enhance convergence by distributing multipoles throughout the molecule at the atomic sites and/or bond centers.[9,35,36]

The MM electrostatic potentials can be classified into two groups. On one side, MM methods such as AMBER,[37] CHARMM,[38] GROMOS,[39] and OPLS[40] adopt simple electrostatic formulas with point charges (i.e., monopoles, with $l_{max} = 0$) that are ultimately the result of a fitting procedure against the molecular electrostatic potential (ESP). On the other side, more sophisticated methods, such as NEMO,[41] AMOEBA,[42] or the QTAIM-based FFLUX,[43,44] include higher order multipoles (frequently up to the quadrupoles, $l_{max} = 2$) in order to capture the anisotropy of the distribution of electrons in space. These potentials are generally built from the QM density matrix of the molecule of interest by means of the distributed multipole analysis (DMA)[36] or similar procedures. In addition, some methods (e.g., AMOEBA or NEMO) also refine the DMA multipoles to better reproduce the ESP values. In this way, the resulting charges/multipoles may include in an effective way both high-order multipolar contributions and CP effects. Actually, the performance of the MM potentials is examined statistically as a whole (i.e., using the full MM potential including all bonded and nonbonded terms) by various energetic and structural validation tests. A quite different approach is followed by the FFLUX force field. It makes use of QTAIM multipoles in contrast to the more widespread DMA methodologies and estimates them by means of a machine learning technique depending on each atom's environment.

In comparison with the atomic/multipolar methods that are massively employed in current simulation packages, the electrostatic MM potentials that go beyond the multipolar approximation are much less consolidated. In this category, we

find different methodologies such as SIBFA,[45] EFP,[46] and AMOEBA+[24] that complement the multipolar formulas with other (so-called damping) functions to capture very-short-range electrostatics and to remove the CP error. In this way, these potentials (whose general form is shown in the SI) seek to reproduce ΔE_{ele}^0 as evaluated by SAPT or similar methodologies.[24]

2.4. Charge Penetration Energy. The CP energy E_{pen} between two molecules has been defined[47] as the difference between the exact zeroth-order electrostatic energy ΔE_{ele}^0 and its multipolar analogue $\Delta E_{ele,mp}^0$,

$$E_{pen} = \Delta E_{ele}^0 - \Delta E_{ele,mp}^0 \tag{10}$$

Conceptually, this straightforward definition of E_{pen} is satisfactory. It also shows that E_{pen} is not only an interfragment quantity but rather an energy that presents also intramolecular contributions according to the real space partitioning of the whole ΔE_{ele}^0. In this respect, the energetic definition suggests that the CP energy is not limited to *the change in the electrostatic interaction between two atoms due to their electron cloud overlap and the associated loss of nuclear screening.*[48]

The rigorous evaluation of E_{pen} for different systems at varying intermolecular separations would allow a deeper analysis of electrostatics and, eventually, the development of more accurate potentials. However, as noticed by Bojarowski et al.,[47] *different methods of obtaining multipole moments lead to different radii of (pseudo)convergence, different levels of multipole expansions at which (pseudo)convergence is achieved, and different values of penetration energy.* Therefore, the value of the CP energy as evaluated with eq 10 may depend on the particular method used to derive the multipoles. Moreover, the usage of truncated expansions introduces some additional truncation error so that both truncation and penetration effects become somewhat mixed in the resulting E_{pen} values.[49]

An alternative to evaluate E_{pen} has been proposed by Kairys and Jensen.[50] Having noticed the relationship between the CP energy and the magnitude of the orbital overlap, they attempt to recover such an effect from scratch, with a derivation of E_{pen} independently from the multipolar model used to estimate electrostatics at first stage. However, the authors find that the inherent dependence on the set of molecular orbitals used may lead to different CP values.

2.4.1. A Novel IQF Definition of the Charge Penetration Energy. By combining both the Bader partitioning scheme ($\mathbb{R}^3 = \Omega_A + \Omega_B$) with a total zeroth-order density decomposition ($\rho^0(\mathbf{r}) = \rho_A^0(\mathbf{r}) + \rho_B^0(\mathbf{r})$), the following energy terms are obtained:

(i) the intramolecular interaction due to ρ_A^0 or ρ_B^0 inside a given molecular basin Ω_A or Ω_B, leading to $E_{ele}^A(\rho_A^0, \rho_A^0)$, $E_{ele}^B(\rho_A^0, \rho_A^0)$, $E_{ele}^A(\rho_B^0, \rho_B^0)$, and $E_{ele}^B(\rho_B^0, \rho_B^0)$.

(ii) the intramolecular interaction between the two monomeric densities inside a given basin: $E_{ele}^A(\rho_A^0, \rho_B^0)$ and $E_{ele}^B(\rho_A^0, \rho_B^0)$.

(iii) the intermolecular electrostatic energy between the same density pieces: $E_{ele}^{AB}(\rho_A^0, \rho_A^0)$ and $E_{ele}^{AB}(\rho_B^0, \rho_B^0)$.

(iv) the intermolecular interaction between ρ_A^0 and ρ_B^0 in opposite molecular basins: $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ and $E_{ele}^{BA}(\rho_A^0, \rho_B^0)$.

Hence, the total electrostatic energy of a complex AB can be written as

$$
\begin{aligned}
E_{ele}^0 = &\ E_{ele}^A(\rho_A^0, \rho_A^0) + E_{ele}^A(\rho_B^0, \rho_B^0) + E_{ele}^A(\rho_A^0, \rho_B^0) \\
&+ E_{ele}^B(\rho_A^0, \rho_A^0) + E_{ele}^B(\rho_B^0, \rho_B^0) + E_{ele}^B(\rho_A^0, \rho_B^0) \\
&+ E_{ele}^{AB}(\rho_A^0, \rho_A^0) + E_{ele}^{AB}(\rho_B^0, \rho_B^0) + E_{ele}^{AB}(\rho_A^0, \rho_B^0) + E_{ele}^{BA}(\rho_A^0, \rho_B^0)
\end{aligned}
\tag{11}
$$

In the notation used above the two interacting densities are encompassed by parentheses, while the basins they are integrated in are identified by the corresponding superscripts in the given order (only one if both are the same). Hence, for instance, the term $E_{ele}^{BA}(\rho_A^0, \rho_B^0)$ stands for $\int_{\Omega_B} d\mathbf{r}_1 \int_{\Omega_A} d\mathbf{r}_2 \rho_A^0(\mathbf{r}_1)\rho_B^0(\mathbf{r}_2) r_{12}^{-1}$ and $E_{ele}^B(\rho_A^0, \rho_A^0)$ corresponds to $\frac{1}{2}\int_{\Omega_B} d\mathbf{r}_1 \int_{\Omega_B} d\mathbf{r}_2 \rho_A^0(\mathbf{r}_1)\rho_A^0(\mathbf{r}_2) r_{12}^{-1}$.

When the above double decomposition is applied to the electrostatic energies of the separate fragments, such as A, in the final complex, the electrostatic energy of the original species becomes

$$E_{ele}(\rho_A^0, \rho_A^0) = E_{ele}^A(\rho_A^0, \rho_A^0) + E_{ele}^B(\rho_A^0, \rho_A^0) + E_{ele}^{AB}(\rho_A^0, \rho_A^0) \tag{12}$$

Note that this partitioning is derived from the AB zeroth-order (i.e., Hartree product) wave function and that geometry relaxation effects are not considered. By subtracting from eq 11 the previous fragment energies, the corresponding electrostatic contribution to the formation energy of the complex is obtained,

$$
\begin{aligned}
\Delta E_{ele}^0 = &\ E_{ele}^A(\rho_A^0, \rho_B^0) + E_{ele}^B(\rho_A^0, \rho_B^0) + E_{ele}^{AB}(\rho_A^0, \rho_B^0) \\
&+ E_{ele}^{BA}(\rho_A^0, \rho_B^0)
\end{aligned}
\tag{13}
$$

Among the surviving terms in eq 13, $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ reveals itself as the ordinary interaction term between the two monomers A and B. It matches ΔE_{ele}^0 at long distances, while the other three terms would present a similar behavior of increasing in magnitude when shortening the intermolecular distances R_{AB} and canceling out in the opposite situation. Thus, those three terms can be directly related with the interpenetration of molecular densities and grouped in the IQF-like electrostatic charge penetration energy

$$E_{ele,pen}^{IQF} = E_{ele}^A(\rho_A^0, \rho_B^0) + E_{ele}^B(\rho_A^0, \rho_B^0) + E_{ele}^{BA}(\rho_A^0, \rho_B^0) \tag{14}$$

This term fulfills $\lim_{R_{AB}\to\infty} E_{ele,pen}^{IQF} = 0$ (and so its three components), while $\lim_{R_{AB}\to\infty} E_{ele}^{AB}(\rho_A^0, \rho_B^0) = \Delta E_{ele}^0$. Figure 1 represents the previous four terms between the partitioned ρ_A^0 and ρ_B^0 adding up to ΔE_{ele}^0 and compares them to the $E_{ele}^{0,AB}$ term between the total densities in each basin.

3. COMPUTATIONAL DETAILS

3.1. Molecular Geometries and Reference Interaction Energies. All the QM and classical electrostatic calculations were performed on the molecular geometries retrieved from the S66 database,[29] which contains a set of 66 complexes featuring the most common noncovalent interactions in biomolecules. These can be classified depending on the atoms involved into polar, nonpolar, and mixed. Analogously, the different complexes have been grouped into H-bond,
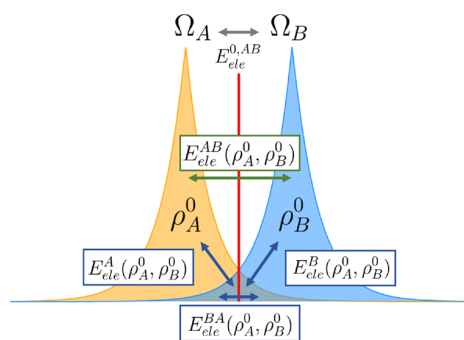
Figure 1. Graphical scheme of the four contributions giving rise to ΔE_{ele}^0, where three of them (in dark blue) comprise the IQF electrostatic penetration energy and the remaining one (dark green) accounts for the interaction of ρ_A^0 and ρ_B^0 lying in the molecular basins Ω_A and Ω_B, respectively. The zeroth-order IQF pairwise term $E_{ele}^{0,AB}$ has been also included to highlight its difference with the previous $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$, as it accounts for an interaction between total densities inside each basin (the original ρ_A^0 or ρ_B^0 and the tail from the other that has penetrated into another domain).

dispersion, and mixed according to the main interactions they experience (see Table S1). For representing both the atomic interactions and the subsets of complexes, a color code has been utilized: magenta for H-bond/polar, yellow for mixed, and blue for dispersion/nonpolar. In addition to the S66 set, a selection of 12 representative complexes from the S66x8 database,[30] which is an extension of the former to eight different fractions of the equilibrium intermolecular distances, were also considered. The benchmark CCSD(T)/CBS interaction energies collected in S66 were employed as the reference values for comparative purposes.

3.2. HF-D3 Calculations. HF/cc-pVTZ calculations were carried out on the S66 and the S66x8 geometries using the GAMESS-US package.[51] Grimme's D3 dispersion potential as implemented in the DFT-D3 code[52] was employed to incorporate the dispersion energy. Additionally, in order to correctly reproduce the asymptotic behavior of the dispersion

energy at small distances, the Becke–Johnson damping function was chosen.[53]

We selected HF because it lacks entirely dispersion energy and thereby yields a straight physical partitioning of energy in combination with the D3 potential. We also note in passing that HF-D3 has been shown to describe correctly and efficiently the structure and energetics of biomolecules[54] and that a variant of DFT-SAPT has been also developed in which the costly *ab initio* dispersion calculations are replaced by a reparametrized D3 potential.[55] In addition, the HF-D3/cc-pVTZ energies reproduce quite well the reference CCSD(T)/CBS energies of the S66 structures (see Figure S1).

3.3. IQA Energy Decomposition Analysis. The decomposition of the QM and the electrostatic energies derived from continuous charge densities were performed with the PROMOLDEN code.[56] The integration settings comprised β-spheres with radii of 60% of the distance between each nucleus and its closest critical point. Within them, Lebedev angular grids with 974 points were used, along with Euler–McLaurin radial quadratures with 382 radial points. A bipolar expansion of r_{12}^{-1} was selected with an l_{max} of 6. On the other hand, the outer part of the basins (i.e., outside the β-spheres) employed the same angular and radial quadratures, albeit increasing their respective points up to 5810 and 512, with a maximum radius of 15 au. In this case, r_{12}^{-1} was expanded by means of a Laplace expansion with $l_{max} = 10$.

3.4. Point-Charge and Multipolar Calculations. Atomic charges were computed for the separate monomers in the S66 structures by means of the restrained electrostatic potential (RESP) method following the General Amber Force Field (GAFF)[57] prescriptions with a HF/6-31G* level of theory. In the case of the atomic multipoles, two different sets were employed. On the one hand, AMOEBA multipoles were derived up to the quadrupoles ($l_{max} = 2$) following its corresponding parametrization protocol.[42,58] On the other, QTAIM multipoles were obtained by means of the PROMOLDEN program with an $l_{max} = 2$. Both the AMOEBA and the QTAIM multipolar energies were obtained with the MPOLINT code.[59]
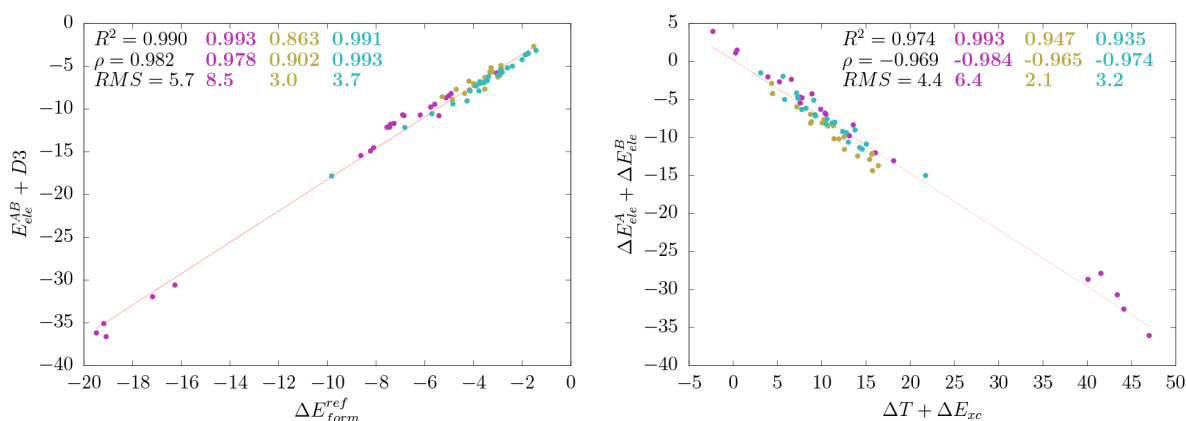
Figure 2. Left: correlation between the dispersion-augmented IQF intermolecular electrostatic energy $E_{ele}^{AB} + D3$ and the reference binding energies ΔE_{form}^{ref}. Right: anticorrelation featured by the intrafragment electrostatic contribution to formation $\Delta E_{ele}^A + \Delta E_{ele}^B$ and the total kinetic plus exchange-correlation contributions $\Delta T + \Delta E_{xc}$. The statistical analysis comprises the coefficient of determination R^2, Spearman's rank correlation coefficient ρ, and the root-mean-square error RMS. Data corresponding to the whole set of complexes is depicted in black and that ascribed to the H-bond group is in magenta, while mixed and dispersion complexes are in yellow and blue, respectively. All the energies are in kcal mol^{-1}.
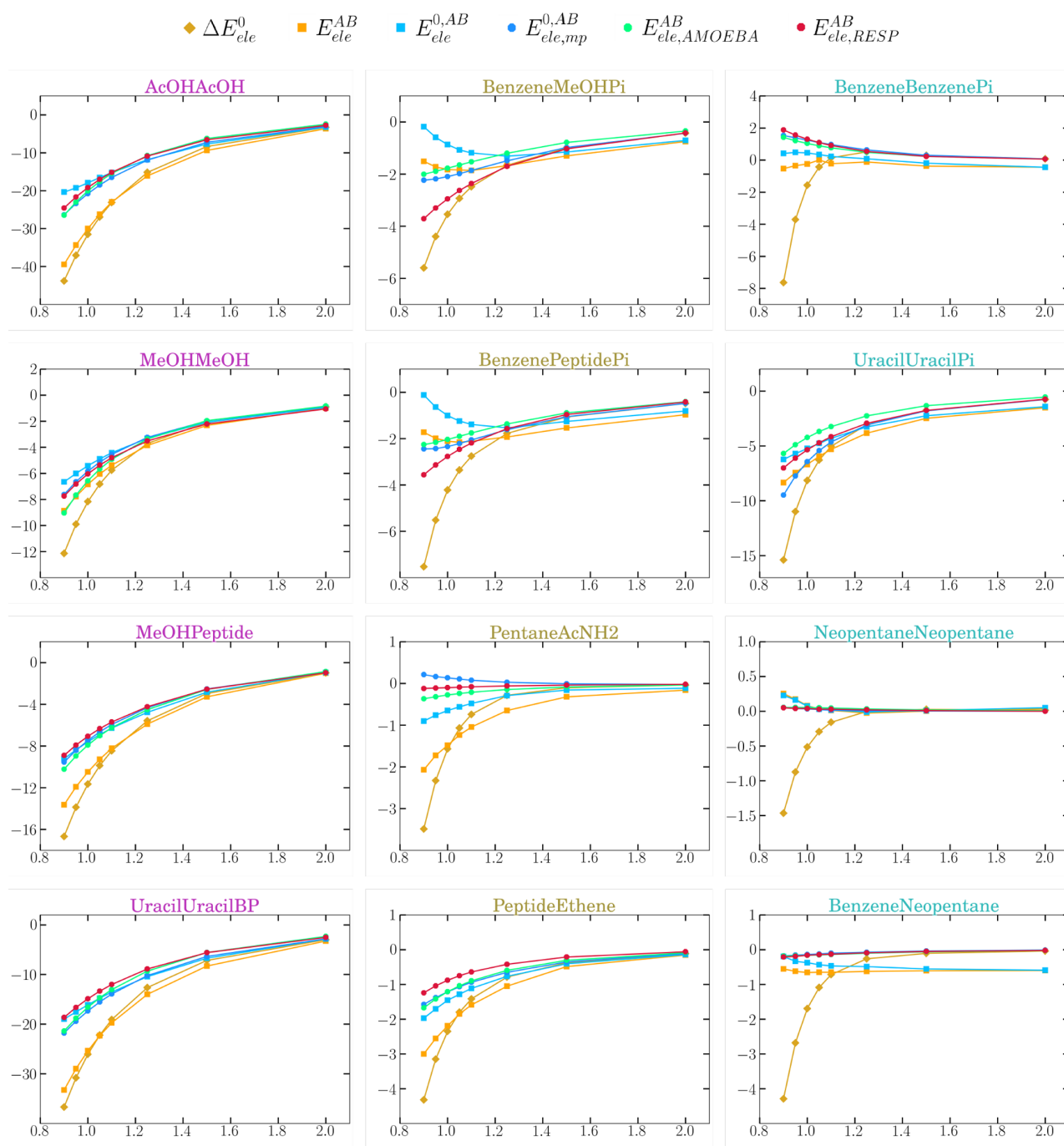
Figure 3. Intermolecular electrostatic interactions for a subset of the S66x8 complexes as provided by IQF (either exactly E_{ele}^{AB} or under the zeroth-order approximation $E_{ele}^{0,AB}$), zeroth-order QTAIM multipoles $E_{ele,mp}^{0,AB}$, AMOEBA multipolar energies $E_{ele,AMOEBA}^{AB}$, and RESP atomic charges $E_{ele,RESP}^{AB}$. Additionally, the zeroth-order electrostatic contribution to formation ΔE_{ele}^0 is also included. The complexes are colored and displayed in columns according to the group they belong to, namely, H-bond, mixed, or dispersion, respectively. The energies (Y-axis) are given in kcal mol^{-1}, and the abscissas represent the intermolecular distances relative to the equilibrium ones (R_{AB}/R_{eq}).

Additionally, a set of 12 S66x8 complexes was tested under the AMOEBA+ CP-corrected potentials.[24] For this, TINKER was used to calculate the respective CP energies as the difference between the CP-corrected multipoles and the multipolar energies previously derived. The parameters of the damping functions were directly taken from the literature.[24]

3.5. Graphs and Statistical Analyses. Octave[60] and GNUplot[61] were, in turn, used to perform the statistical analyses and the correlation plots, while Python's Matplotlib[62] was chosen for the rest of the representations.

4. RESULTS AND DISCUSSION

4.1. IQF-D3 Partitioning and Pairwise Approximation. The IQF-D3 decomposition of the HF/cc-pVTZ binding energies for the S66 complexes has been discussed at length in previous work.[32] Herein, we focus on the decomposition of the electrostatic descriptors into intra- and interfragment components. Interestingly, we found that the combination of the interfragment electrostatic interaction energy E_{ele}^{AB} with the D3 dispersion potential yields pairwise energies that are quite well correlated with the S66 benchmark values, the coefficient of determination being $R^2 = 0.990$ with *RMS* errors of 5.7 kcal mol^{-1} (see Figure 2 and Table S2). Thus, the IQF E_{ele}^{AB} descriptors in conjunction with the D3 potential capture the essential electrostatic and dispersion interactions that determine the relative stability of the noncovalent complexes. When addressing both terms independently (Figure S2), we find that the pure electrostatic E_{ele}^{AB} term exhibits a satisfactory overall correlation ($R^2 = 0.943$) due to the fundamental role of electrostatics in H-bond complexes. On the other hand, the D3 descriptor has a null global correlation with the S66 reference energies, although it is reasonable ($R^2 = 0.820$) for the dispersion complexes as expected. However, the mixed complexes are not well-described by either the electrostatic or the dispersion energies separately, and their combination becomes critical.

In contrast to the ability of the E_{ele}^{AB} + D3 descriptors to capture the main features of noncovalent binding, the combination of ΔE_{ele}, which includes both the intra- and the intermolecular electrostatic effects, with the D3 potential deteriorates the global correlation ($R^2 = 0.888$) and results in larger *RMS* errors (17.3 kcal mol^{-1}). The full IQF decomposition (eq 8) explains this unbalanced description because the intrafragment electrostatic energies, which contribute to the deformation energies, tend to cancel out with the QM energy terms (electronic kinetic energy and exchange-correlation) that are not required in the simple electrostatic + dispersion picture (see Figure 2 right). Therefore, the pairwise E_{ele}^{AB} terms arise as the most relevant IQF electrostatic descriptors of noncovalent binding.

4.2. Validating and Analyzing the Zeroth-Order Approximation. The electrostatic IQF terms can be readily evaluated under the zeroth-order approximation (i.e., $\rho^0(\mathbf{r}) = \rho_A^0(\mathbf{r}) + \rho_B^0(\mathbf{r})$). Thus, it turns out that the interaction energies E_{ele}^{AB} can be replaced effectively by their zeroth-order counterparts. Indeed, the pairwise $E_{ele}^{0,AB}$ + D3 energies have low *RMS* errors (3.1 kcal mol^{-1}) and maintain a good correlation ($R^2 = 0.971$) with respect to the benchmark data (Table S3). This behavior is also satisfactory within the S66 subsets: $R^2 = 0.989$ and 0.988 for the polar H-bonded systems and the dispersion-dominated complexes, respectively, albeit the correlation is somewhat reduced in the case of the mixed complexes ($R^2 = 0.755$). Further support for the use of the zeroth-order energies comes from the atomic level, where a high degree of coincidence between the diatomic zeroth-order $E_{ele}^{0,IJ}$ and fully relaxed E_{ele}^{IJ} energies is also found at the equilibrium geometries ($R^2 = 0.995$, see the SI).

When addressing the distance dependence of the previous term (see Figure 3), both E_{ele}^{AB} and $E_{ele}^{0,AB}$ follow the same trends at varying intermolecular separations R_{AB} (given as relative to the equilibrium distances R_{eq}). As expected, they start diverging at short distances due to the strengthening of charge polarization, charge-penetration, and charge-transfer effects that attenuate the pairwise electrostatic forces. The magnitude of these effects is clearly system-dependent, as well as the shape and slope of the E_{ele}^{AB} and $E_{ele}^{0,AB}$ curves, revealing thus further details about the role of electrostatics in these complexes. Thus, the electrostatic stabilization of the four H-bond complexes and others (e.g., the π-complex of the uracil dimer) is continuously reinforced upon shortening the monomer−monomer distance, reflecting the major electrostatic control of these systems. In contrast, the T-shaped benzene complexes with methanol or *N*-methylacetamide reach an electrostatic minimum at a distance longer than the equilibrium one while the small electrostatic energies of the dispersion dimers (i.e., +1, −1 kcal/mol) change very little along the curves (some small leaps are due to residual errors arising in the numerical integration over the atomic basins).

In Figure 3 the deviation between the global ΔE_{ele}^0 energies and the interfragment $E_{ele}^{0,AB}$ anticipates the underlying CP effects associated with the density overlap. For the H-bond and some of the mixed complexes, the two curves decrease with lowering separation, but they split gradually for $R_{AB}/R_{eq} < 1.6$. The global ΔE_{ele}^0 stabilization nearly doubles $E_{ele}^{0,AB}$ at R_{eq}, showing thus the large impact of intramolecular electrostatics as defined in the IQF framework. For the π-complexes (benzene−dimer, benzene−methanol, ...) or the weakly interacting neopentane dimer, the inter- vs intramolecular balance is differently modulated because the deviation between the global and the interfragment electrostatics becomes significant only at very close distances (e.g., $R_{AB}/R_{eq} < 1.1$), which are indicative of mutual overlap. In these systems, ΔE_{ele}^0 is thus reinforced by several kcal mol^{-1}, which are ascribed to the intrafragment electrostatic stabilization achieved by the fragment-overlap (i.e., CP) effects. Such effects have a minor influence on the small $E_{ele}^{0,AB}$ energies (<1−2 kcal mol^{-1}), which tend to remain nearly constant or become slightly attenuated. As shown below (Section 4.5), the IQF analysis of the CP energy gives further insight about the behavior of $E_{ele}^{0,AB}$ and ΔE_{ele}^0 with R_{AB}/R_{eq}.

4.3. Comparison between $E_{ele}^{0,AB}$ and Pairwise MM Energies. The pairwise approximation that emerges from the IQF-D3 decomposition and the validity of the zeroth-order approximation for the electrostatic interactions provide an insightful theoretical support for the construction of noncovalent MM potentials. In this scenario, $E_{ele}^{0,AB}$ can be seen as the most suitable IQF descriptor to assess the approximate electrostatic potentials. Hence, we calculated the interfragment electrostatic energies using the RESP atomic charges and the AMOEBA multipoles, as well as the QTAIM multipoles up to the quadrupoles.

According to the statistical data in Table 1, either the RESP atomic charges or the QTAIM/AMOEBA multipoles give interfragment electrostatic energies that correlate considerably well with $E_{ele}^{0,AB}$ ($R^2 > 0.9$ and *RMS* errors ∼ 1 kcal mol^{-1}) for the full S66 set and also for the H-bond/dispersion subsets. These point-charge/multipolar electrostatic energies are less satisfactory for the less abundant mixed complexes, although the multipolar potentials yield a more accurate description ($R^2 \simeq 0.6−0.8$) than the RESP charges ($R^2 \simeq 0.5$). In addition to ΔE_{ele}^0 and the fully relaxed and zeroth-order IQF pairwise

Table 1. Statistical Measurements Comprising the Coefficient of Determination R^2, Spearman's Rank Correlation Coefficient ρ, and the Root Mean Square Error RMS for the Correlation between $E_{ele}^{0,AB}$ and either the QTAIM or AMOEBA Multipoles ($l_{max} = 2$) or the RESP Point Charges ($l_{max} = 0$)

multipolar approximation	complex type	R^2	ρ	RMS
QTAIM	global	0.970	0.958	1.0
	H-bond	0.956	0.904	1.4
	mixed	0.644	0.768	0.9
	dispersion	0.955	0.795	0.5
AOMEBA	global	0.953	0.972	1.3
	H-bond	0.904	0.841	2.0
	mixed	0.800	0.845	0.7
	dispersion	0.939	0.893	0.4
RESP	global	0.974	0.962	0.8
	H-bond	0.981	0.918	0.7
	mixed	0.456	0.687	1.1
	dispersion	0.948	0.831	0.3

terms, Figure 3 also displays the distance dependence of the QTAIM/AMOEBA/RESP energies, that results quite close to that of the interfragment $E_{ele}^{0,AB}$ energies. Nevertheless, a closer inspection reveals that the QTAIM/AMOEBA/RESP energies tend to overestimate the stabilizing/destabilizing character of $E_{ele}^{0,AB}$ for the H-bond/dispersion dimers, respectively.

The good agreement between the multipolar and the RESP energies in Table 1 and in Figure 3 suggests that the RESP fitting procedure may incorporate in an effective way higher order effects even at short distances. In addition, our results point out that the pure QTAIM multipoles can be employed in the construction of accurate electrostatic potentials, free from the inclusion of other effects that may be present when the DMA multipoles are fitted against the molecular ESP. In fact, the QTAIM multipoles, which are already considered in the FFLUX force field, readily reproduce the ESP without the need of any constraint.[63]

4.4. Comparing Diatomic Electrostatic Interactions. IQA permits an unambiguous decomposition of the continuous-density intermolecular interaction energy into a sum of atomic and diatomic terms that enables a thorough analysis of the global molecular properties based on their atomic origins, and a close comparison with the various MM descriptions at this atomic level.

As expected, the IQA diatomic terms correlate almost perfectly with the QTAIM multipolar ones $E_{ele,mp}^{0,IJ}$ (see Figure 4). On the contrary, the AMOEBA and RESP energies are significantly less correlated (R^2 of 0.7 and 0.4, respectively) and have large RMS errors. For example, the largest discrepancies between $E_{ele}^{0,IJ}$ and the QTAIM-multipolar $E_{ele,mp}^{0,IJ}$ in the acetic acid dimer (about 6 kcal mol^{-1}) arise from the atoms involved in the OH·O H-bonds, the rest of pair interactions having much lower differences (<0.5 kcal mol^{-1}; see Tables S7−S9). When comparing $E_{ele}^{0,IJ}$ and $E_{ele,AMOEBA}^{IJ}$ (or $E_{ele,RESP}^{IJ}$), the largest discrepancies amount to hundreds of kcal mol^{-1} and involve not only short polar contacts but methyl C atoms too (see Tables S10−S15).

The dissimilarity between the $E_{ele,mp}^{0,IJ}$ energies and the $E_{ele,AMOEBA}^{IJ}/E_{ele,RESP}^{IJ}$ values was not entirely unexpected given that the RESP charges are derived from the molecular ESP and the AMOEBA multipoles are obtained by the DMA protocol. In fact, a difference of 1 order of magnitude between the atom−atom electrostatic interactions from IQA and MM potentials has also been noticed previously.[64] The present results show in further detail the actual discrepancies between the various atomic representations and suggest that, although the diverse atomic multipoles employed in classical potentials yield similar molecular electrostatic energies, the atomic decomposition is more questionable, which, in turn, can negatively affect the interpretation of localized electrostatic interactions and/or result in artifacts while dealing with QM and MM short-range electrostatics in hybrid QM/MM methodologies.

4.5. Charge Penetration under the QTAIM Scrutiny. Following the prescriptions introduced in Theory and Methods, the zeroth-order electrostatic formation energy ΔE_{ele}^0 of each S66 complex was decomposed by combining its real space partition into nonoverlapping atomic basins with the zeroth-order density approximation ($\rho^0(\mathbf{r}) = \rho_A^0(\mathbf{r}) + \rho_B^0(\mathbf{r})$). This strategy leads to the IQF-based charge penetration energies, $E_{ele,pen}^{IQF}$, resulting from the sum of the intramolecular terms $E_{ele}^A(\rho_A^0, \rho_B^0)$ and $E_{ele}^B(\rho_A^0, \rho_B^0)$, as those accounting for the interaction of both densities inside the same basin, and the intermolecular energy $E_{ele}^{BA}(\rho_A^0, \rho_B^0)$ between the tails of each molecular density that penetrate into the opposite basin, as described in eq 14. This constitutes an *effective* penetration energy in the sense that the molecular identity between two overlapping fragments becomes necessarily blurred so that fragment properties are dependent upon the scheme followed to dissect the global charge density into its constituents. Nevertheless, the topological analysis of ρ^0 yields a consistent identification of molecular fragments so that we believe that the associated charge-penetration analysis can give useful insight into the electrostatics of noncovalent complexes.

The application of eq 14 to ΔE_{ele}^0 results in the energy contributions shown in Figure 5. On the one hand, the interfragment energy $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ is formally not affected by charge penetration and plays a stabilizing role in all the H-bond complexes (slightly repulsive in the dispersion complexes). On the other hand, the IQF penetration term $E_{ele,pen}^{IQF}$ turns out to be of equal importance in the H-bond complexes or even more relevant in the dispersion subset for which penetration energy describes the major part of ΔE_{ele}^0.

The decomposition of the penetration energy shows that it arises mainly from the stabilizing interactions between ρ_A^0 and ρ_B^0 inside the same basin. This is an *intramolecular* effect as reflected by the magnitude of the $E_{ele}^B(\rho_A^0, \rho_B^0)$ and $E_{ele}^A(\rho_A^0, \rho_B^0)$ energies. As shown by the integration of ρ_A^0 or ρ_B^0 in the corresponding basins, the mutual CP values range, for instance, from 0.035 e in the neopentane dimer to 0.099 e in the case of the acetic acid dimer. These fractional charges involve the e−e repulsion between the fragment electron densities occupying the same space, such as $\rho_{e,A}^0(\mathbf{r})$, $\rho_{e,B}^0(\mathbf{r})$ such that $\mathbf{r} \in \Omega_A$ (or equivalently in region Ω_B), and the attraction experienced by
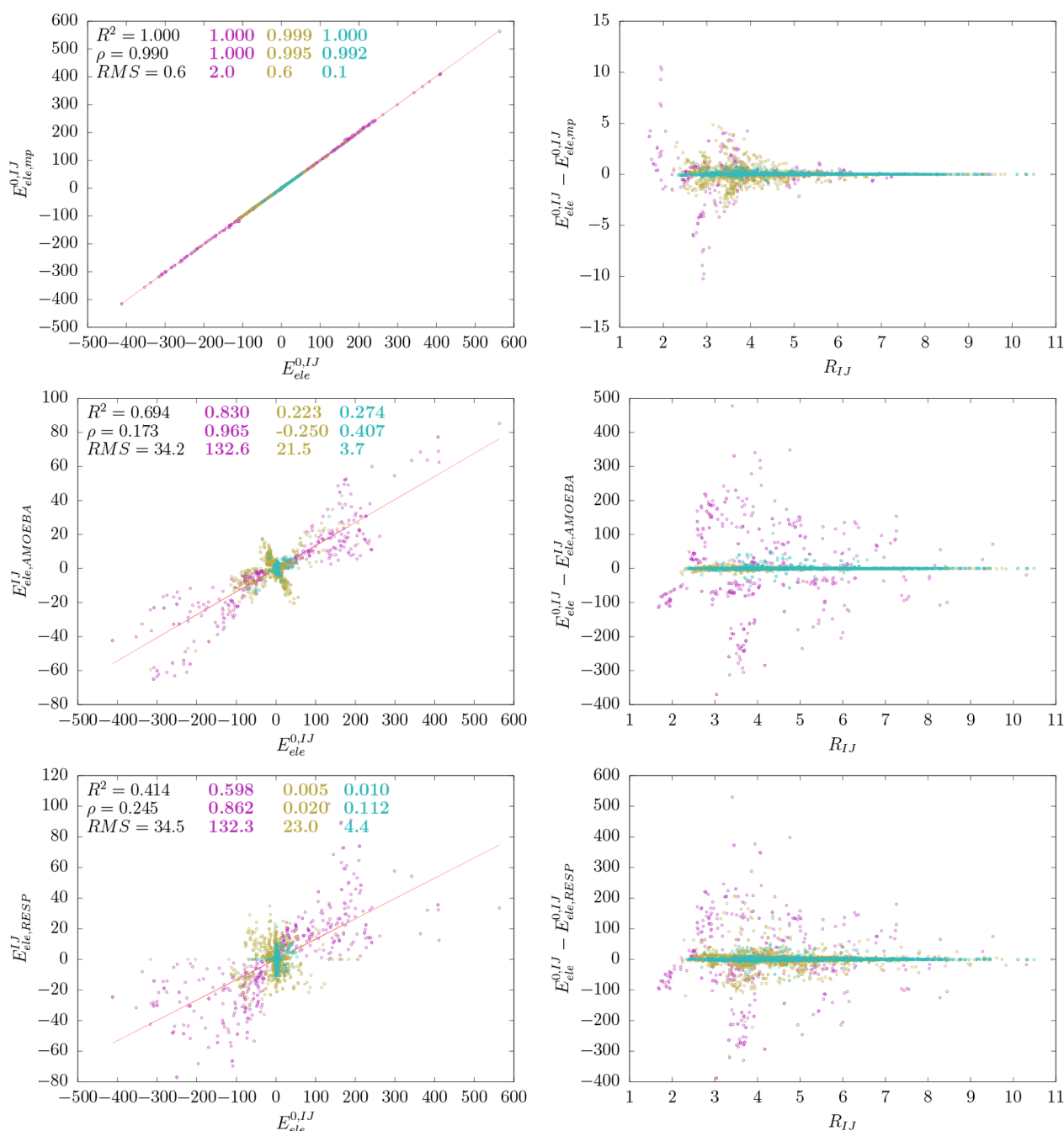
Figure 4. Comparison of the $E_{ele,mp}^{0,IJ}$, $E_{ele,AMOEBA}^{0,IJ}$ and $E_{ele,RESP}^{IJ}$ energies with the $E_{ele}^{0,IJ}$ term (kcal mol^{-1}). On the left are the correlation plots and, on the right, each difference as a function of the interatomic distance (Å).

the nuclei of one fragment $\{Z_I\}_{I \in A}$ (or $\{Z_J\}_{J \in B}$) and the fraction of electrons from the other that has penetrated into the former $\rho_{e,B}^0(\mathbf{r} \in \Omega_A)$ (or similarly $\rho_{e,A}^0(\mathbf{r} \in \Omega_B)$). In light of these results, e−n attraction greatly overcomes e−e repulsion between different zeroth-order densities inside the same basin and gives rise to the significant stabilizing energies observed. There is also a minor repulsive contribution owing to the purely electronic repulsion between the penetrating ρ_A^0 into Ω_B and the ρ_B^0 tail in Ω_A, which is measured by $E_{ele}^{BA}(\rho_A^0, \rho_B^0)$.

Further insight can be gained by analyzing the distance dependence of the various energy terms as shown in Figure 6. The plots confirm that the three components of E_{pen}^{IQF} tend to zero when $R_{AB}/R_{eq} > 1.5$ and further highlight the role of the intrafragment terms. Interestingly, the $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ energy, formally lacking penetration effects, is modulated by the degree of the interfragment overlap so that the decreasing trend in $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ is damped out or inverted at the shortest distances. This is not entirely unexpected given that, as two
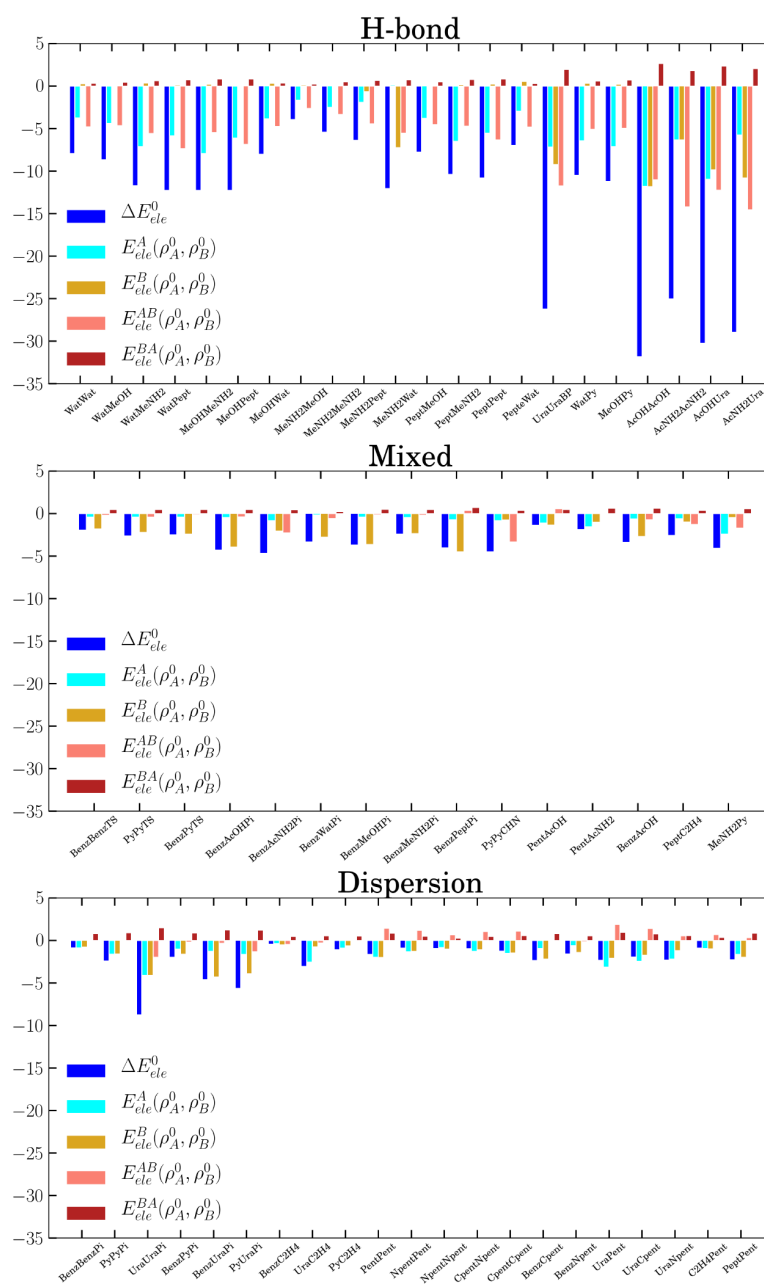
Figure 5. Decomposition of ΔE_{ele}^0 into $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ and the three IQF penetration terms $E_{ele}^A(\rho_A^0, \rho_B^0)$, $E_{ele}^B(\rho_A^0, \rho_B^0)$, and $E_{ele}^{BA}(\rho_A^0, \rho_B^0)$. Energies are given in kcal mol^{-1}.

initially separated atomic basins (e.g, $\Omega_{I \in A}$ and $\Omega_{J \in B}$) approach one another, their volume, shape, and electron population evolve along the R_{AB}/R_{eq} curve in response to the density overlap. We note, however, that the deviation of $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ with respect to the interfragment electrostatic energy $E_{ele}^{0,AB}$ may constitute a useful index about the specific impact of penetration effects on the pairwise electrostatics. At this point, an important caveat should be noted. Within the QTAIM framework, the $E_{ele}^{0,AB}$ energy includes a fraction of stabilizing penetration energy for $R_{AB}/R_{eq} < 1.2$ given that the loss of some electronic ρ_A^0 density from the basins of the monomer A

is partially compensated by the penetration of ρ_B^0 into the same basin. The fixed multipoles/charges in the classical potentials somehow mimic this behavior so that they remain closer to the $E_{ele}^{0,AB}$ descriptors than to $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ around the equilibrium distance.

Finally, Figure 7 compares the IQF penetration term and other relevant energetic quantities with the analogue term derived from the AMOEBA+ model as a function of the intermolecular distance. Thus, the combination of the multipolar $E_{ele,AMOEBA}^{AB}$ energies with the CP correction[24] $E_{pen}^{AMOEBA+}$ results in the $\Delta E_{ele}^{AMOEBA+}$ energies that approach
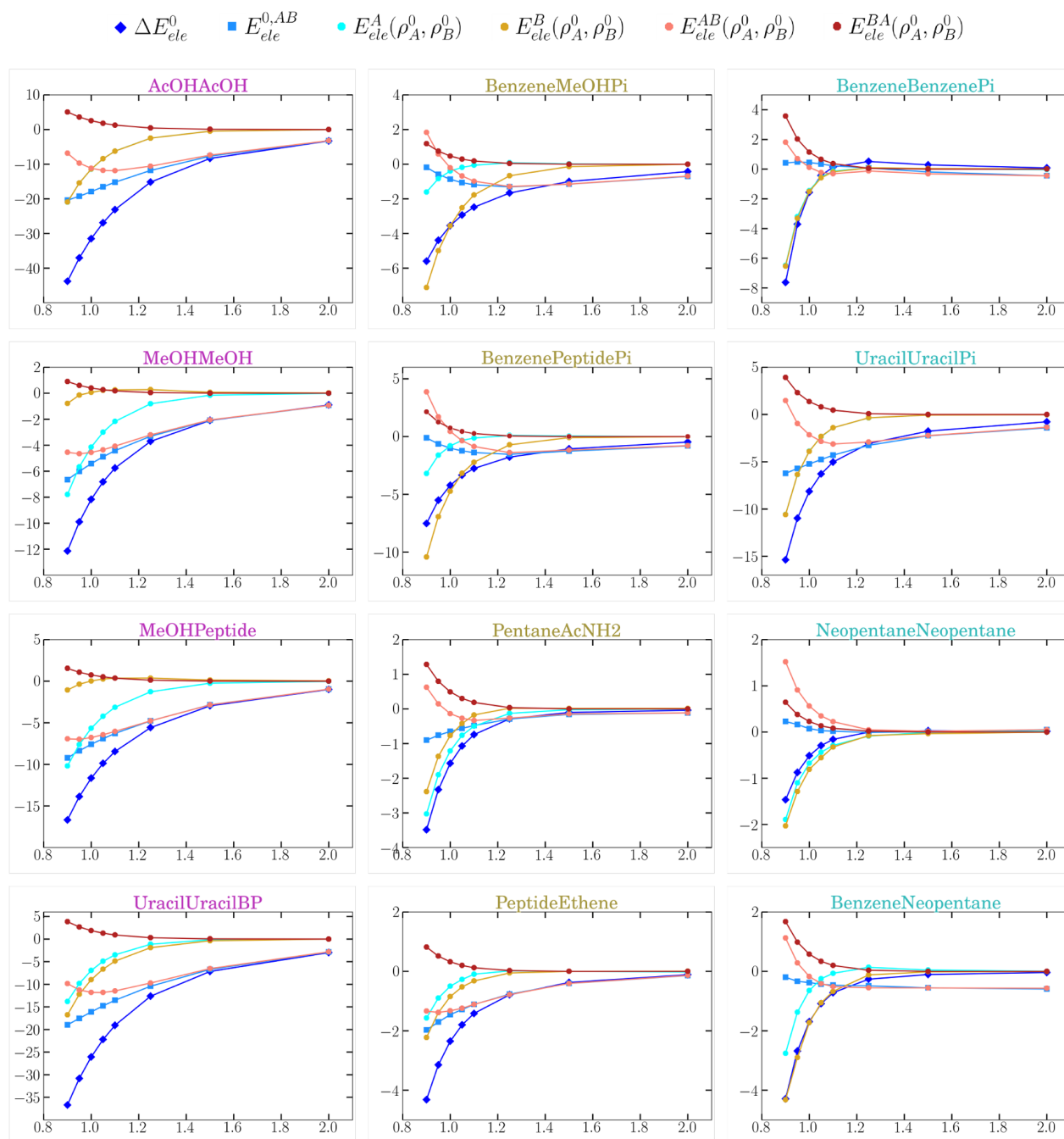
Figure 6. Evolution of the energy terms from eq 13, along with the $E_{ele}^{0,AB}$ pair term as a function of the distance for the set of S66x8 systems chosen. The complexes are grouped in three columns as belonging to the H-bond, mixed, or dispersion subsets, respectively.

to the reference ΔE_{ele}^0, which is equivalent to the SAPT electrostatic energy. In effect, Figure 7 shows that $\Delta E_{ele}^{AMOEBA+}$ nearly matches ΔE_{ele}^0. Concerning the CP energies, it is important to note again that the AMOEBA+ reference for measuring the CP energy is different from that provided by the IQF-QTAIM approach. Nevertheless, the two penetration energies exhibit a similar behavior with R_{AB}, particularly for the more stable H-bond complexes, which resemble also the variations experienced by the intramolecular CP terms,

$E_{ele}^A(\rho_A^0, \rho_B^0)$ and $E_{ele}^B(\rho_A^0, \rho_B^0)$. Therefore, we conclude that the AMOEBA+ CP and similar corrections account mainly for intramolecular electrostatics.

5. CONCLUDING REMARKS

In this work we have analyzed the short-range electrostatic interactions in the S66 and S66x8 data sets through a hierarchy of approximations at both the molecular and the atomic levels. We have shown first that the IQA/IQF decomposition augmented with the D3 dispersion terms gives support to
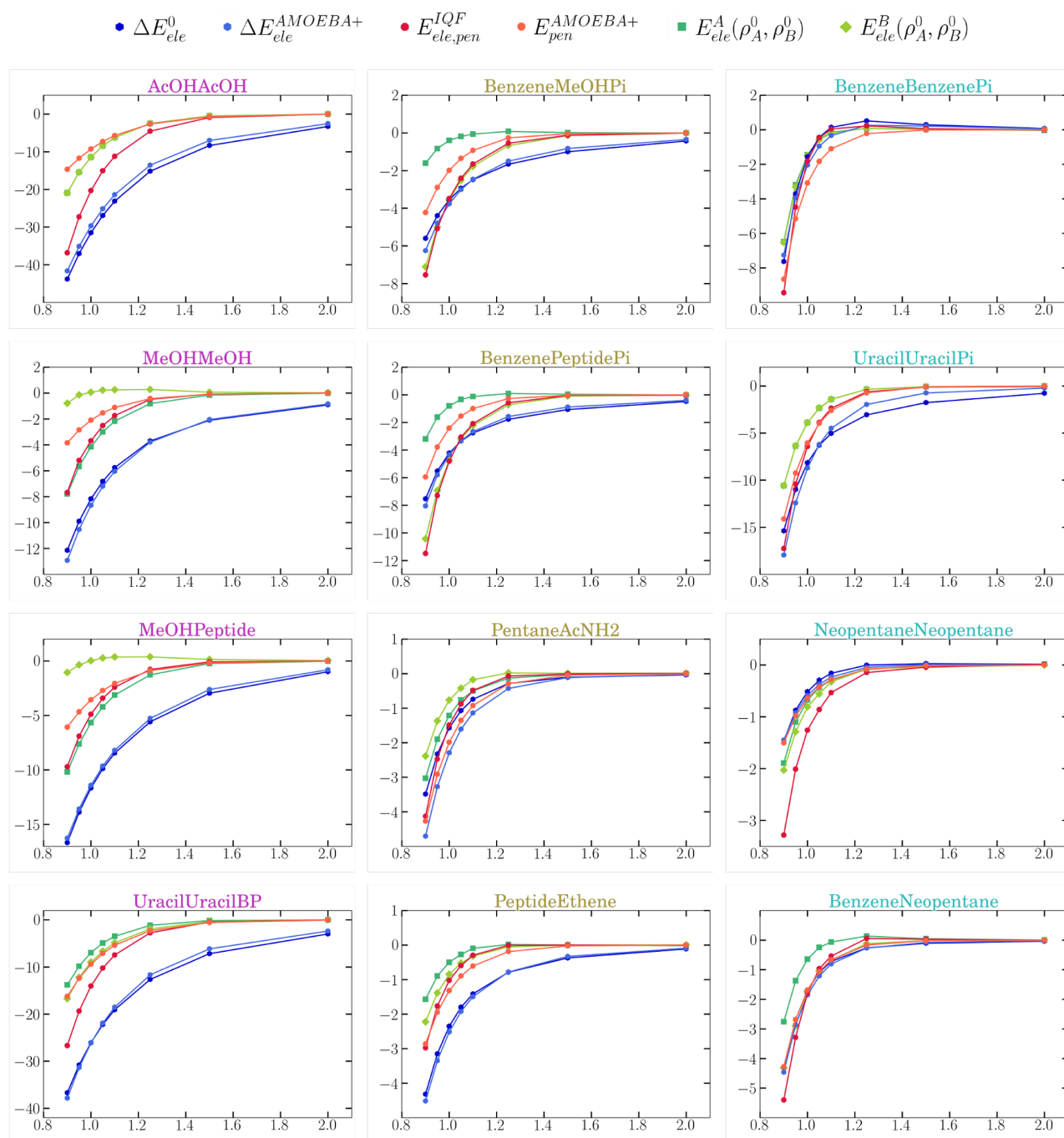
Figure 7. Comparison between the AMOEBA+ model and the zeroth-order IQF energies for our model S66x8 complexes. The complexes have been displayed according to the group they belong to (either H-bond, mixed, or dispersion). Distances (X-axis) are relative to the equilibrium ones (R_{AB}/R_{eq}) and energies (Y-axis) are in kcal mol^{-1}.

the pairwise approach adopted by many MM potentials. In this respect, the interfragment energies E_{ele}^{AB} derived from the IQF partitioning suffice to capture the essential electrostatic effects explaining the binding of the weakly interacting complexes. Moreover, the same role can be played by the equivalent $E_{ele}^{0,AB}$ values, which are obtained from the unrelaxed densities of the isolated monomers (i.e., the zeroth-order approximation).

According to our results, the intermolecular $E_{ele}^{0,AB}$ energy turns out to be the most appropriate IQF descriptor to analyze and/or compare with electrostatic MM potentials. In particular, we have considered two widely used potentials relying on the RESP atomic charges or the AMOEBA distributed atomic multipoles, respectively, as well as the multipolar potential up to the quadrupoles derived directly from the QTAIM basins. The three MM pairwise approximations correlate satisfactorily with the zeroth-order IQF term

at varying intermolecular distances and exhibit small *RMS* errors. However, when the $E_{ele}^{0,AB}$ values are further decomposed into diatomic contributions, large discrepancies between the RESP or the AMOEBA atom−atom interactions and their zeroth-order IQA counterparts are unveiled. Although this is understandable in terms of the specific details of the RESP/AMOEBA charge/multipole derivations, it contrasts sharply with the nearly perfect match between the QTAIM atomic multipolar energies and the IQA reference values. Hence, MM potentials based on the QTAIM multipoles—such as the QCT-based FFLUX—may provide a more consistent description of electrostatic interactions at both the molecular and the atomic levels.

Besides forging links between the IQF/IQA quantities and the MM electrostatic potentials, we have studied the charge penetration effects that arise from the mutual interpenetration of the zeroth-order molecular densities in their opposite QTAIM basins as built from the final ρ^0 of the complex. This QTAIM perspective allows us to dissect the CP energy into different contributions that emphasize its intramolecular character, which, in turn, is dominated by the attraction between the nuclei of fragment A (B) and the penetrating tail of density B (A). In this way we may clarify some practical issues related with the CP corrections for MM potentials. For example, adding CP corrections to MM potentials like RESP/AMOEBA, which target the zeroth-order interfragment electrostatic energy, results, necessarily, in an unbalanced description. This aspect, which has been overlooked in previous works,[10,48,65] implies also that the electrostatic energy employed in popular MM force fields (AMBER, CHARMM, ...) cannot be compared with the global ΔE_{ele}^0 energy derived from continuous charge distributions, but with its interfragment component. On the other hand, CP corrections have been derived to improve the description of the QM−MM electrostatic interactions in hybrid QM/MM methodologies.[12] In this case, such corrections should mitigate short-range electrostatic artifacts, particularly those associated with the QM−MM covalent linkages. However, considering the highly dissimilar interatomic electrostatic energies produced by the QM densities and the RESP/AMOEBA potentials, the usage of electrostatic parameters more akin to the QM densities at the atomic level may have a larger impact in improving the QM−MM electrostatics.

Finally, concerning the novel MM potentials inspired by the QM SAPT methodology, it is clear that the multipolar electrostatics (interfragment) must be augmented by the CP potentials (intrafragment) if one seeks to reproduce the global electrostatics ΔE_{ele}^0. Nevertheless, the IQF/IQA approach (and other EDAs) points out that the intramolecular electrostatic energy is closely related with other energy changes induced by fragment overlap (e.g., deformation and interfragment exchange-correlation energy), suggesting thus that the separate treatment of these effects by means of independent potential terms might be inefficient and hamper parameter development and transferability.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00263.

(1) Multipole expansion formulation, (2) a brief description of the CP corrections, (3) classification of the S66 complexes, (4) HF-D3−CCSD(T) comparison, (5) statistical analysis of the fully relaxed and zeroth-order electrostatic descriptions, (6) correlation study between the interatomic E_{ele}^{IJ} and $E_{ele}^{0,IJ}$ terms, and (7) atomic interactions within the acetic acid dimer (PDF)

AUTHOR INFORMATION

Corresponding Authors

Fernando Jiménez-Grávalos − *Departamento de Química Física y Analítica, Universidad de Oviedo, E-33006 Oviedo, Spain*; ⊙ orcid.org/0000-0002-9181-5234; Email: jimenezfernando@uniovi.es

Dimas Suárez − *Departamento de Química Física y Analítica, Universidad de Oviedo, E-33006 Oviedo, Spain*; ⊙ orcid.org/0000-0001-8003-2309; Email: dimas@uniovi.es

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.1c00263

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Ren, P.; Chun, J.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J.; Baker, N. A. Biomolecular electrostatics and solvation: a computational perspective. *Q. Rev. Biophys.* **2012**, *45*, 427−491.

(2) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical Electrostatics for Biomolecular Simulations. *Chem. Rev.* **2014**, *114*, 779−814.

(3) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369*, 253−287.

(4) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; IOP Publishing Ltd.: 1988.

(5) Pollock, E.; Glosli, J. Comments on P3M, FMM, and the Ewald method for large periodic Coulombic systems. *Comput. Phys. Commun.* **1996**, *95*, 93−110.

(6) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(7) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(8) Herce, D. H.; Perera, L.; Darden, T. A.; Sagui, C. Surface solvation for an ion in a water cluster. *J. Chem. Phys.* **2005**, *122*, 024513.

(9) Stone, A. *The Theory of Intermolecular Forces*; Oxford University Press: 2013.

(10) Bojarowski, S. A.; Kumar, P.; Dominiak, P. M. A Universal and Straightforward Approach to Include Penetration Effects in Electrostatic Interaction Energy Estimation. *ChemPhysChem* **2016**, *17*, 2455−2460.

(11) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. Evaluation of charge penetration between distributed multipolar expansions. *J. Chem. Phys.* **2000**, *112*, 7300−7306.

(12) Wang, B.; Truhlar, D. G. Including Charge Penetration Effects in Molecular Modeling. *J. Chem. Theory Comput.* **2010**, *6*, 3330−3342.

(13) Wang, B.; Truhlar, D. G. Screened Electrostatic Interactions in Molecular Mechanics. *J. Chem. Theory Comput.* **2014**, *10*, 4480−4487.

(14) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 2609−2618.

(15) Francisco, E.; Martín Pendás, A. Energy Partition Analyses: Symmetry-Adapted Perturbation Theory and Other Techniques. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; Otero de la Roza, A., DiLabio, G. A., Eds.; Elsevier: 2017; Chapter 2, pp 27−64.

(16) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887−1930.

(17) Patkowski, K. Recent developments in symmetry-adapted perturbation theory. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *10*, e1452.

(18) Kitaura, K.; Morokuma, K. A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int. J. Quantum Chem.* **1976**, *10*, 325−340.

(19) Bickelhaupt, F. M.; Baerends, E. J. Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry. *Reviews in Computational Chemistry*; John Wiley Sons, Ltd.: 2000; pp 1−86.

(20) Zhao, L.; von Hopffgarten, M.; Andrada, D. M.; Frenking, G. Energy decomposition analysis. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1345.

(21) Blanco, M. A.; Martin Pendas, A.; Francisco, E. Interacting Quantum Atoms: A Correlated Energy Decomposition Scheme Based on the Quantum Theory of Atoms in Molecules. *J. Chem. Theory Comput.* **2005**, *1*, 1096−1109.

(22) Francisco, E.; Martin Pendas, A.; Blanco, M. A. A Molecular Energy Decomposition Scheme for Atoms in Molecules. *J. Chem. Theory Comput.* **2006**, *2*, 90−102.

(23) Naseem-Khan, S.; Gresh, N.; Misquitta, A. J.; Piquemal, J.-P. Assessment of SAPT and Supermolecular EDA Approaches for the Development of Separable and Polarizable Force Fields. *J. Chem. Theory Comput.* **2021**, *17*, 2759−2774.

(24) Liu, C.; Piquemal, J.-P.; Ren, P. AMOEBA+ Classical Potential for Modeling Molecular Interactions. *J. Chem. Theory Comput.* **2019**, *15*, 4122−4139.

(25) Vandenbrande, S.; Waroquier, M.; Speybroeck, V. V.; Verstraelen, T. The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions. *J. Chem. Theory Comput.* **2017**, *13*, 161−179.

(26) Mao, Y.; Demerdash, O.; Head-Gordon, M.; Head-Gordon, T. Assessing Ion−Water Interactions in the AMOEBA Force Field Using Energy Decomposition Analysis of Electronic Structure Calculations. *J. Chem. Theory Comput.* **2016**, *12*, 5422−5437.

(27) Pendás, A. M.; Casals-Sainz, J. L.; Francisco, E. On Electrostatics, Covalency, and Chemical Dashes: Physical Interactions versus Chemical Bonds. *Chem. - Eur. J.* **2019**, *25*, 309−314.

(28) Massa, L.; Keith, T.; Cheng, Y.; Matta, C. F. The kernel energy method applied to quantum theory of atoms in molecules − energies of interacting quantum atoms. *Chem. Phys. Lett.* **2019**, *734*, 136650.

(29) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.

(30) Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466−3470.

(31) Bader, R. *Atoms in molecules: a quantum theory*; Clarendon Press: Oxford; New York, 1990.

(32) Suárez, D.; Díaz, N.; Francisco, E.; Martín Pendás, A. Application of the Interacting Quantum Atoms Approach to the S66 and Ionic-Hydrogen-Bond Datasets for Noncovalent Interactions. *ChemPhysChem* **2018**, *19*, 973−987.

(33) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(34) Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101.

(35) Stone, A. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.* **1981**, *83*, 233−239.

(36) Stone, A.; Alderton, M. Distributed multipole analysis. *Mol. Phys.* **1985**, *56*, 1047−1064.

(37) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs: Comput. Mol. Sci.* **2013**, *3*, 198−210.

(38) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2009**, *31*, 671−690.

(39) Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F. A.; van Gunsteren, W. F. Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.* **2005**, *34*, 273−284.

(40) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(41) Engkvist, O.; Åstrand, P.-O.; Karlström, G. Accurate Intermolecular Potentials Obtained from Molecular Wave Functions: Bridging the Gap between Quantum Chemistry and Molecular Simulations. *Chem. Rev.* **2000**, *100*, 4087−4108.

(42) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.

(43) Popelier, P. L. A. QCTFF: On the construction of a novel protein force field. *Int. J. Quantum Chem.* **2015**, *115*, 1005−1011.

(44) Hughes, Z. E.; Thacker, J. C. R.; Wilson, A. L.; Popelier, P. L. A. Description of Potential Energy Surfaces of Molecules Using FFLUX Machine Learning Models. *J. Chem. Theory Comput.* **2019**, *15*, 116−126.

(45) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960−1986.

(46) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry. *J. Phys. Chem. A* **2001**, *105*, 293−307.

(47) Bojarowski, S. A.; Kumar, P.; Dominiak, P. M. Interplay of point multipole moments and charge penetration for intermolecular electrostatic interaction energies from the University at Buffalo pseudoatom databank model of electron density. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2017**, *73*, 598−609.

(48) Rackers, J. A.; Wang, Q.; Liu, C.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. An optimized charge penetration model for use with the AMOEBA force field. *Phys. Chem. Chem. Phys.* **2017**, *19*, 276−291.

(49) Kumar, P.; Bojarowski, S. A.; Jarzembska, K. N.; Domagała, S.; Vanommeslaeghe, K.; MacKerell, A. D.; Dominiak, P. M. A Comparative Study of Transferable Aspherical Pseudoatom Databank and Classical Force Fields for Predicting Electrostatic Interactions in Molecular Dimers. *J. Chem. Theory Comput.* **2014**, *10*, 1652−1664.

(50) Kairys, V.; Jensen, J. H. Evaluation of the charge penetration energy between non-orthogonal molecular orbitals using the Spherical Gaussian Overlap approximation. *Chem. Phys. Lett.* **1999**, *315*, 140−144.

(51) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(52) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(53) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(54) Goerigk, L.; Collyer, C. A.; Reimers, J. R. Recommending Hartree−Fock Theory with London-Dispersion and Basis-Set-Superposition Corrections for the Optimization or Quantum Refinement of Protein Structures. *J. Phys. Chem. B* **2014**, *118*, 14612−14626.

(55) Sedlak, R.; Řezáč, J. Empirical D3 Dispersion as a Replacement for ab Initio Dispersion Terms in Density Functional Theory-Based Symmetry-Adapted Perturbation Theory. *J. Chem. Theory Comput.* **2017**, *13*, 1638−1646.

(56) Martín Pendás, Á.; Francisco, E. PROMOLDEN. A QTAIM/IQA code (unpublished).

(57) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(58) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143−3161.

(59) Francisco, E. MPOLINT. A program to calculate Coulombic interactions by means of the multipole approximation (unpublished).

(60) Eaton, J. W.; Bateman, D.; Hauberg, S.; Wehbring, R. *GNU Octave version 4.4.0 manual: a high-level interactive language for numerical computations*; 2018.

(61) Williams, T.; Kelley, C.; et al. *Gnuplot 5.2: an Interactive Plotting Program*; 2019. http://gnuplot.sourceforge.net/.

(62) Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.

(63) Popelier, P. L. A.; Kosov, D. S. Atom−atom partitioning of intramolecular and intermolecular Coulomb energy. *J. Chem. Phys.* **2001**, *114*, 6539−6547.

(64) Yuan, Y.; Mills, M. J. L.; Popelier, P. L. A. Multipolar electrostatics for proteins: Atom-atom electrostatic energies in crambin. *J. Comput. Chem.* **2014**, *35*, 343−359.

(65) Zgarbová, M.; Otyepka, M.; Šponer, J.; Hobza, P.; Jurečka, P. Large-scale compensation of errors in pairwise-additive empirical force fields: comparison of AMBER intermolecular terms with rigorous DFT-SAPT calculations. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10476−10493.

Beryllium Bonding in the Light of Modern Quantum Chemical Topology Tools

José Luis Casals-Sainz, Fernando Jiménez-Grávalos, Aurora Costales, Evelio Francisco, and Ángel Martín Pendás*

Departamento de Química Física y Analítica. Universidad de Oviedo. Spain.

E-mail: ampendas@uniovi.es

---

*To whom correspondence should be addressed

Abstract

We apply several modern quantum chemical topology (QCT) tools to explore the chemical bonding in well established Beryllium bonds. By using the interacting quantum atoms (IQA) approach together with electron distribution functions (EDF) and the natural adaptive orbitals (NAdOs) picture we show that, in agreement with orbital-based analyses, the interaction in simple σ and π complexes formed by BeX_2 (X=H,F,Cl) with water, ammonia, ethylene and acetylene is dominated by electrostatic terms, albeit covalent contributions cannot be ignored. Our detailed analysis proves that several σ back-donation channels are relevant in these dimers, actually controlling the conformational preference in the π adducts. A number of one-electron Beryllium bonds are also studied. Orbital invariant real space arguments clearly show that the role of covalency and charge transfer cannot be ignored.

Introduction

The rise in the number of studies regarding non-covalent interactions in this first part of the twenty-first century is out of discussion.[1] Once chemists have mastered the art of building individual molecules, it was only a matter of time that the focus turned toward understanding the rules governing supramolecular assemblies. In this soft-matter[2] regime, it is weak interactions that are responsible for structure and ultimately function in territories as different as biology and crystal engineering.[3] It comes as no surprise then that, as the number of systems investigated grew, so did the number of specific weak bonds reported. The paradigmatic hydrogen bond (HB) category has thus been enlarged with a large set of *new* bonding motifs: dihydrogen, halogen, pnicogen, tetrel bonds have been defined and used to control supramolecular structures. And as it happened with the hydrogen bond,[4] several different, sometimes opposing views about the nature of these interactions have flourished.[5–9] On one end we find the electrostatic point of view, pioneered by Politzer and Murray.[10,11] According to this position, it is the electrostatic attraction between an electron depleted region on one fragment, the so-called σ-hole, and an electron rich one on the other that drives these non-covalent interactions. In this view, which can also be understood as a general-

ized Lewis acid/base framework or as a weak interaction version of Weinhold's donor-acceptor paradigm,[12] maps of the electrostatic potential (ESP) play a relevant role in the correct *docking* of σ-holes and electron rich domains. On the other hand, many theoretical studies, summarized by the IUPAC[13] have shown that covalency or dispersion play also a relevant role.

In 2009, following the idea that σ-holes are indeed playing the role of electron acceptors, the group of Manuel Yáñez and Otilia Mó proposed that Beryllium derivatives, known to be very good Lewis acceptors, should form complexes with Lewis bases.[14] Since then, these beryllium bonds have been added to the toolkit of new non-covalent links, and have been shown to be considerably strong and dominated by electrostatic interactions. Be-bonds have been found in both σ-[14] and π-complexes[15], and their density polarizing abilities have been used to modify the bonding abilities of other moieties, for instance inducing σ-holes in fluorine-containing systems.[16] As of today, the nature of the Be-bond has been established by a battery of methods that include standard molecular orbital ideas,[17] Fock-space energy decomposition analyses like the LMOEDA method,[18] Weinhold's natural bond orbitals (NBO),[12] quantum chemical topology (QCT) methods like the quantum theory of atoms in molecules (QTAIM) of Bader and coworkers,[19] or the electron localization function (ELF),[20] the natural orbitals for chemical valence of Ziegler, Michalak and Mitoraj[21,22], etc. Most of these methods have shown that electrostatics is an important player in explaining Be-bonds, but that covalent contributions are not negligible. However, Politzer, Murray and Clark have recently argued strongly against these *theoretical constructs*.[10] According to these authors, exchange, Pauli repulsions or orbitals are *simply mathematical constructs that are used to obtain an approximate solution of the multi-electron Schrödinger equation*. Only electrostatics, via the Hellmann-Feynman theorem, does account for bonding, and other effects, such as charge transfer lie only in the model: *the illusion of charge transfer is in the model; the reality of polarization is in the electronic density, the physical observable.*

We show here that QCT methods, using quantities which are in principle amenable to experimental determination, provide univocal answers to these questions. By employing reduced density matrices (which are Dirac observables) and QTAIM partitions (which can be and are actually deter-

mined in experiments) we: (i) gain access to energy components (through the interacting quantum atoms approach, IQA) including electrostatic and covalent terms; (ii) provide effective one-electron pictures that contain correlation effects via natural adaptive orbitals;[23] and (iii) uncover the distribution of the electron population, thus true charge transfers, with electron distribution functions (EDFs).

As far as we know, this is the first time that Be-bonds are studied using such a combined strategy. We have decided to focus on three sets of aggregates: σ-complexes, exemplified by the BeX_2-LB series, LB being a Lewis base like ammonia or water; π-complexes, like those formed by BeX_2 and ethylene or acetylene; and one-electron bonded complexes, with LiBe and BeLiBe as examples. Additional motivations that justify addressing the study of these (and other) beryllium compounds, trying to clarify the main characteristics of their chemical bonds, are purely practical. It is well known, for instance, that beryllium fluoride associates to ADP inhibiting protein action.[24–30] Reactivity patterns of half-sandwich complexes formed by Be are also interesting.[31,32] Finally, although it is not completely established, it seems that the interaction between Be^{2+} and several water molecules plays an important role on beryliosis or CBD (chronic beryllium disease), an very often fatal illness caused by this metal.[33,34] Theoretical analyses of chemical bonds formed by beryllium may help to clarify the possible causes associated to its toxicity. In this work, we focus on the nature of the stabilizing energy components, explaining how electrostatics, but not only electrostatics, accounts for their binding energy. We also consider charge transfer, interpreting it in a crystal clear way as a result of real space resonance of several electron configurations. We start by summarizing the conceptual framework to be used, turning to the computational details and an analysis of our results.

Methodology

Only Dirac observable densities are used in QCT, that starts with a physical partition of space into chemically meaningful regions.[19] We use the QTAIM partition, so that $R^3 = \bigcup_i^m A_i$, where A_i is an

attraction basin of the electron density field, $\rho(\mathbf{r})$, usually corresponding to an atom-in-a-molecule. We stress that the atoms of the QTAIM are nowadays obtained routinely both from computations as well as from X-ray diffraction experiments.[35] Much information can be obtained by examining the local topology of ρ,[36] but since this QTAIM local operating mode is well known and several works on Be-bonds have used it, we will not consider it in this work. The position space QTAIM partitioning may be directly inherited by all the reduced density matrices (nRDMs)[37]. For instance, the $A_1 A_2 \ldots A_n$ component of the n-th order reduced density (nRD) is simply defined as

$$\rho_n^{A_1 A_2 \ldots A_n}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n) = \rho_n(\mathbf{r}_1 \in \Omega_{A_1}, \mathbf{r}_2 \in \Omega_{A_2}, \ldots, \mathbf{r}_n \in \Omega_{A_n}).$$

Because the electronic energy for a Coulomb Hamiltonian depends only on the 1RDM and the 2RD, $E = \mathrm{Tr}\,\hat{h}\rho_1(\mathbf{r}_1, \mathbf{r}_1') + (1/2)\mathrm{Tr}\,r_{12}^{-1}\rho_2(\mathbf{r}_1, \mathbf{r}_2)$, we can write E as a sum of one- and two-domain components. This leads to the IQA decomposition,

$$E = \sum_A E_{\mathrm{self}}^A + \sum_{A>B} E_{\mathrm{int}}^{AB}. \tag{1}$$

In this expression, E_{self}^A is the self-energy of atom A, which adds all the energy terms that depend only on nuclei (n) and electrons (e) contained in domain A, while the pairwise additive interatomic energy, E_{int}^{AB}, gathers all contributions containing particles in the A and B regions. In this way, $E_{\mathrm{self}}^A = T^A + V_{\mathrm{ne}}^{AA} + V_{\mathrm{ee}}^{AA}$ and $E_{\mathrm{int}}^{AB} = V_{\mathrm{nn}}^{AB} + V_{\mathrm{ee}}^{AB} + V_{\mathrm{ne}}^{AB} + V_{\mathrm{ne}}^{BA}$, where we have used a clear notation that needs no more comments. If self-energies are measured with respect to given energetic references for each atom or fragment, E_0^A, then deformation energies arise as $E_{\mathrm{def}}^A = E_{\mathrm{self}}^A - E_0^A$.

We obtain a fruitful decomposition of E_{int}^{AB} if we further partition the 2RD into its Coulomb (J) and exchange-correlation (xc) components, $\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \rho_2^J(\mathbf{r}_1, \mathbf{r}_2) + \rho_2^{xc}(\mathbf{r}_1, \mathbf{r}_2)$. In this way, we can separate all terms in E_{int} depending on the one-particle density (that would correspond to the interaction among classical particles) from those calculated from the exchange-correlation density (with no analogue in classical mechanics), so that $E_{\mathrm{int}}^{AB} = E_{\mathrm{cl}}^{AB} + E_{\mathrm{xc}}^{AB}$. The contribution E_{cl}^{AB} is thus immediately associated to the electrostatic (ionic) energy component of a chemical bond, tending

asymptotically to $Q_A Q_B / R_{AB}$ for charged species, while E_{xc}^{AB} i.e., the exchange-correlation energy, represents a measure of covalency[38,39]. Notice that, regarding Politzer *et al.* considerations, E_{cl} is the purely classical electrostatic interaction of the modified, mutually polarized, molecular densities.

If not the energy-weighted RDs but the RDs themselves are coarse-grained, we get a general population analysis.[37] With this, we can decompose general n-th order cumulant densities (CDs) $\rho_n^c(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n)$ that integrate to the total number of electrons N into one-, two- or n-center terms: a partition of the $\rho_c^1 = \rho$ provides the standard QTAIM atomic populations, that of $\rho_2^c = \rho_{xc}$ gives rise to the well-known localization and delocalization indices which are real space covalent bond orders, and, in general, a decomposition of ρ_n^c reveals n-center bond orders. Doing this with the N-th order RD, $\Psi^*\Psi$, we can obtain the probability of finding a given number of electrons in each of the n A regions,[40–42] leading to EDFs. Finally, CDs can be partially coarse-grained, leaving one electron coordinate free of this process that describes a real space natural density of n-center bonding. Diagonalizing these densities we get sets of effective one electron functions, the natural adaptive orbitals (NAdOs),[23] together with their associated natural adaptive occupations. The latter decompose the electron population, the two-center, three-center, etc. bond orders into one-electron components as the order of the cumulant that is diagonalized increases. All these quantities can in principle be obtained from experiments (for instance with X-ray constrained wave functions[43,44]).

Computational details

We have performed IQA decompositions and analyzed EDFs and NAdOs in a set of relevant Be-bonded molecules that include the CO, H_2O and NH_3 complexes of BeX_2 (X=H,F,Cl) together with the BeLi, BeLiBe, BeNa, and $C_{10}H_6Be_2H_2^-$ systems. All geometry optimizations were carried out at the B3LYP//def2-tzVPD level using the GAMESS package.[45] Further single points calculations at the B3LYP/def2-qzVPD level using the PySCF suite[46] were performed for all subsequent analyses, using the scaling approach as described in Ref. 47 For BeLi, BeLiBe, and BeNa

CASSCF calculations using a aug-cc-pVQZ basis were preferred and obtained also with PySCF. The following notation is used: (electrons,orbitals). The orbital space for each system was selected using the density matrix embedding theory (DMET).[48] For the bath selection a cutoff of 0.01 was chosen and the following impurities were selected: in the case of BeLi and BeLiBe, the $2s, 2p$ orbitals for each atom, resulting in a $(3, 10)/(5, 15)$ space; for BeNa, the $2s, 2p$ orbitals for Be, and the $3s, 3p$ for the Na atom, resulting in a $(3, 9)$ space.

IQA integrations were performed using β-spheres with radii between 0.1 and 0.3 bohr. Restricted angular Lebedev quadratures with 5810 points and 451 points Gauss-Chebyshev mapped radial grids were used inside the β-spheres, with L expansions cut at $l = 10$. Outside the β-spheres, extended 5810-point Lebedev, 551- and 651- mapped radial point Gauss-Legendre quadratures, and L expansions up to $l = 12$ were selected. Total energies reconstructed from these IQA decompositions differ in less than 1.0 kcal/mol from those of the parent electronic structure codes, and since error cancelation does not occur in these numerical integrations, each of the computed interactions is considerably more accurate than this figure. This accuracy is enough for the aims of this paper. All IQA calculations were done with our in-house code PROMOLDEN, available upon request[49]. Electron distribution functions (EDF) were obtained using our in house EDF code,[50] and natural adaptive orbitals (NAdOs) with DENMAT.[51] In both cases the atomic overlap matrices (AOM) that are needed to feed these codes were obtained from PROMOLDEN.

Beryllium σ-complexes

The first Beryllium bonds analyzed[14] were σ complexes between BeX$_2$ (X=H,F,Cl) units and several Lewis bases like water or ammonia. Besides recognizing the important geometrical distortions suffered by the linear BeX$_2$ moiety upon complexation as well as systematizing the considerable strength of the interactions, their chemical bonding was interpreted in terms of local topological descriptors coming from the electron density or the ELF and via the natural bond orbital (NBO) analysis. This led to propose an important role of the LB lone pairs as donors to both the empty

p orbital of Be and to the σ^*_{BeX} antibond. Further analyses[52] have even reported some IQA data, but have failed, in our opinion, to provide much more than a description of the computed data. As we expect to show, charge transfer between the units is an important player that helps rationalizing conformer preference, leading to both LB$\rightarrow$ BeX$_2$ donation as well as to LB$\leftarrow$ BeX$_2$ back-donation channels which as far as we know have not been reported.

Table 1: Several IQA properties for BeX$_2$-Lewis base (LB) complexes. A and B denote BeX$_2$ and LB, respectively. E^A_{def} and E^B_{def} are given with respect to the total energies of the isolated fragments at the geometry they have in the complex, $E^{\text{AB}}_{\text{bind}} = E^{\text{AB}}_{\text{int}} + E^A_{\text{def}} + E^B_{\text{def}}$, and $E^{\text{AB}}_{\text{dis}} = E^{\text{AB}}_{\text{bind}} + E_{\text{relax}}$, where E_{relax} is the energy associated to the geometric relaxation of the fragments to their own optimal geometries. Energies in kcal/mol, Q_A in electrons and δ_{AB} in electron pairs.

System	$E^{\text{AB}}_{\text{int}}$	$E^{\text{AB}}_{\text{cl}}$	$E^{\text{AB}}_{\text{xc}}$	E^A_{def}	E^B_{def}	$E^{\text{AB}}_{\text{bind}}$	$E^{\text{AB}}_{\text{dis}}$	Q_A	δ^{AB}
H$_2$Be$\cdots$OH$_2$	-111.03	-54.30	-56.73	35.27	50.45	-25.31	-18.29	0.014	0.467
H$_2$Be$\cdots$NH$_3$	-118.76	-60.89	-57.88	34.17	52.45	-32.15	-22.70	-0.010	0.481
H$_2$Be$\cdots$CO	-103.18	-38.16	-65.02	44.94	45.90	-12.34	-6.24	0.076	0.593
H$_2$Be$\cdots$C$_2$H$_4{}^a$	-119.53	-39.71	-79.82	52.67	51.48	-15.38	-5.67	0.080	0.755
H$_2$Be$\cdots$C$_2$H$_4{}^b$	-36.14	-7.42	-28.73	10.81	21.95	-3.39	-1.71	-0.030	0.311
H$_2$Be$\cdots$C$_2$H$_2{}^a$	-139.34	-49.00	-90.34	60.25	59.51	-19.57	-8.85	0.103	0.826
H$_2$Be$\cdots$C$_2$H$_2{}^b$	-15.41	-1.84	-13.57	5.32	9.48	-0.62	-0.51	-0.014	0.162
F$_2$Be$\cdots$OH$_2$	-102.09	-52.19	-49.90	27.35	44.99	-29.75	-20.41	-0.024	0.387
F$_2$Be$\cdots$NH$_3$	-113.46	-59.95	-53.51	26.83	48.83	-37.79	-26.32	-0.045	0.422
F$_2$Be$\cdots$CO	-66.34	-23.72	-42.62	20.90	34.61	-10.83	-5.37	-0.020	0.372
F$_2$Be$\cdots$C$_2$H$_4{}^a$	-57.40	-16.33	-41.07	16.84	30.94	-9.62	-3.91	-0.047	0.399
F$_2$Be$\cdots$C$_2$H$_4{}^b$	-60.32	-19.17	-41.15	16.21	32.67	-11.44	-5.10	-0.050	0.394
F$_2$Be$\cdots$C$_2$H$_2{}^a$	-65.59	-21.35	-44.24	19.55	34.17	-11.87	-5.90	-0.041	0.416
F$_2$Be$\cdots$C$_2$H$_2{}^b$	-50.13	-14.48	-35.64	13.54	28.48	-8.11	-2.91	-0.045	0.346
Cl$_2$Be$\cdots$OH$_2$	-127.37	-67.44	-59.93	39.26	55.75	-32.35	-21.12	-0.001	0.502
Cl$_2$Be$\cdots$NH$_3$	-137.32	-74.89	-62.43	37.21	59.54	-40.56	-27.77	-0.027	0.527
Cl$_2$Be$\cdots$CO	-100.40	-40.29	-60.11	39.47	47.78	-13.15	-5.11	0.029	0.563
Cl$_2$Be$\cdots$C$_2$H$_4{}^a$	-78.91	-24.96	-53.95	25.53	42.44	-10.94	-1.13	-0.028	0.557
Cl$_2$Be$\cdots$C$_2$H$_4{}^b$	-76.68	-25.83	-50.85	22.43	42.41	-11.84	-2.94	-0.037	0.521
Cl$_2$Be$\cdots$C$_2$H$_2{}^a$	-93.21	-33.89	-59.32	31.52	47.64	-14.05	-3.04	-0.017	0.590
Cl$_2$Be$\cdots$C$_2$H$_2{}^b$	-64.45	-20.29	-44.16	19.98	36.77	-7.70	-0.70	-0.035	0.457

In order to keep the discussion succinct, we have only considered the H$_2$O, NH$_3$, and CO dimers with BeX$_2$ (X=H,F,Cl). In all cases the X-Be-X angle differs considerably from 180°, see the supplementary information. Relevant IQA data are reported on Table 1. As it is well known, all BeX$_2$ systems are considerably ionic with QTAIM charges of about 1.7 and -0.85 electrons for the

Be and X atoms respectively. In all the cases the L shell of Beryllium has been transferred to the X companion, as seen by the laplacian. Complexation polarizes the BeX_2 fragment, that gets bent leaving a large positive laplacian region toward which the lone pair of the LB points. All $Be\cdots LB$ bond critical points display positive laplacian. It is interesting to notice that the $BeX_2\cdots H_2O$ systems are planar, with a $\sigma + \pi$ distribution of the oxygen's lone pairs adequate for π donation, as suggested by NBO analyses.

Total charge transfer between the fragments is in general larger than in typical hydrogen bonded (HB) complexes,[53] getting as large as 0.08 electrons in $BeH_2\cdots CO$. Notice that the direction of charge transfer oscillates. A negatively charged LB necessarily implies that there exist other delocalization channels beside the $lp \to p_{Be}$ or $lp \to \sigma^*_{BeX}$. This back-donation dominates effectively in the BeH_2 complexes with water and in $BeX_2\cdots CO$ with X=H,Cl. We will return to the origin of this interesting observation.

As energy components are regarded, both electrostatic and covalent contributions are important in the stabilization of the complexes. In all the σ dimers it is the deformation energy of the Lewis base that dominates, as it is also the case in HB dimers. However, deformations in the present Be-bonds are several times larger than the ones found in HBs. The deformation of the LB in the water dimer, for instance, is only about 8 kcal/mol, to be compared with 50 kcal/mol in the BeH_2 complex. The distortions induced by mutual polarization of the fragments in Beryllium bonds seem to be rather intense.

Electron delocalization is extremely relevant. The A-B delocalization indices lie around $0.3 - 0.6$, being much larger than in HBs (one of the largest is found in the strongly bound FHF^- system, with a δ close to 0.2.[53]). This makes the stronger nature of distortions and interactions in Be-bonds clear as compared to HBs. As in usual (relatively) weak complexes, the sum of the fragments' deformation energies and E^{AB}_{xc}, which we have related several times to the exchange-repulsion terms in other approaches,[53,54] is positive, again larger (by one order of magnitude) than the values found in HBs. Delocalization thus does not compensate deformation. This behavior is typical of traditional charge transfer complexes or very ionic bonds, being in favor of electrostatic bonding

models. But, as we have also explained before,[53] covalency cannot be disregarded, since E_{cl}^{AB} is not stabilizing enough as to overcome deformations.

Table 2: Electron Distribution Functions (EDF) for some BeX$_2$-Lewis base (LB) complexes. A and B denote BeX$_2$ and LB, respectively. $p(BeX_2)$ denotes the probability of the nominal RSRS and $p(BeX_2^q)$ the probability of the RSRS of the fragment BeX$_2$ with a total charge q.

System	$p(BeX_2)$	$p(BeX_2^-)$	$p(BeX_2^+)$	$p(BeX_2^{2-})$	$p(BeX_2^{2+})$
H$_2$Be$\cdots$OH$_2$	0.795	0.092	0.103	0.004	0.005
H$_2$Be$\cdots$NH$_3$	0.788	0.106	0.096	0.004	0.004
H$_2$Be$\cdots$CO	0.741	0.093	0.153	0.003	0.011
H$_2$Be$\cdots$C$_2$H$_4$[a]	0.684	0.115	0.177	0.007	0.015
H$_2$Be$\cdots$C$_2$H$_4$[b]	0.855	0.084	0.057	0.003	0.001
H$_2$Be$\cdots$C$_2$H$_2$[a]	0.662	0.116	0.195	0.007	0.019
H$_2$Be$\cdots$C$_2$H$_2$[b]	0.922	0.045	0.032	0.001	0.000
F$_2$Be$\cdots$OH$_2$	0.826	0.094	0.073	0.004	0.003
F$_2$Be$\cdots$NH$_3$	0.811	0.111	0.070	0.005	0.003
F$_2$Be$\cdots$CO	0.830	0.092	0.072	0.003	0.003
F$_2$Be$\cdots$C$_2$H$_4$[a]	0.820	0.108	0.065	0.005	0.002
F$_2$Be$\cdots$C$_2$H$_4$[b]	0.822	0.108	0.063	0.005	0.002
F$_2$Be$\cdots$C$_2$H$_2$[a]	0.815	0.107	0.071	0.005	0.003
F$_2$Be$\cdots$C$_2$H$_2$[b]	0.842	0.097	0.056	0.004	0.002
Cl$_2$Be$\cdots$OH$_2$	0.784	0.103	0.102	0.005	0.006
Cl$_2$Be$\cdots$NH$_3$	0.772	0.121	0.095	0.006	0.005
Cl$_2$Be$\cdots$CO	0.758	0.125	0.104	0.004	0.009
Cl$_2$Be$\cdots$C$_2$H$_4$[a]	0.763	0.125	0.100	0.008	0.013
Cl$_2$Be$\cdots$C$_2$H$_4$[b]	0.775	0.123	0.090	0.015	0.025
Cl$_2$Be$\cdots$C$_2$H$_2$[a]	0.752	0.124	0.109	0.007	0.007
Cl$_2$Be$\cdots$C$_2$H$_2$[b]	0.800	0.111	0.080	0.005	0.003

The nature of charge transfer deserves specific consideration. Table 2 contains the most representative real space resonance structures (RSRSs) contributing to our Beryllium complexes. Notice that the overall inter-fragment charge transfer sign comes from the relative weight of the BeX$_2^-$ versus the BeX$_2^+$ structures, given the small values of more highly charged structures BeX$_2^{2-}$ and BeX$_2^{2+}$. The non-negligible probabilities of both the BeX$_2^-$ and BeX$_2^+$ RSRSs point towards polarized symmetric (thus covalent-like) delocalizations or to several independent charge transfer channels. We can distinguish between both possibilities by exploring the two-fragment natural

adaptive orbitals that decompose the overall delocalization index into effective one-electron components. Table 3 reports the most relevant contributions for each of the complexes, while Fig. 1 provides pictorial images in the $BeH_2 \cdots H_2O$ and $BeF_2 \cdots H_2O$ cases. There are several interesting points that deserve being discussed.

Table 3: Contributions to δ^{AB} greater than 0.01 of the two-fragment NAdOs of BeX$_2$-Lewis base (LB) systems. Each number represents the occupation of a natural adaptive orbital (NAdO). See Figure 1 for the cases of $BeH_2 \cdots H_2O$ and $BeF_2 \cdots H_2O$.

BeX$_2$ →	BeH$_2$			BeF$_2$			BeCl$_2$		
LB →	H$_2$O	NH$_3$	CO	H$_2$O	NH$_3$	CO	H$_2$O	NH$_3$	CO
	0.154	0.213	0.287	0.155	0.214	0.194	0.169	0.234	0.233
	0.121	0.126	0.217	0.066	0.056	0.066	0.103	0.084	0.151
	0.119	0.100	0.075	0.054	0.056	0.053	0.080	0.081	0.074
	0.048	0.018		0.038	0.035	0.036	0.051	0.043	0.065
	0.017	0.016		0.030	0.023		0.035	0.030	
				0.020	0.014		0.027	0.019	
							0.010	0.011	
							0.010		

Firstly, the dominant channel is always that involving the base σ lone pair. Its intensity is largest in the ammonia complexes, as expected, if we exclude the $BeH_2 \cdots CO$ system. In the second place, the LB to p_{Be} donation, the fourth component in both rows of Fig. 1, is never second in relevance, as NBO arguments suggest. Contrarily, its donating capacity is relatively low. This type of discrepancies between NBO and real space considerations is rather usual, stemming from the NBO's use of localized basis functions in a second order perturbation expansion. More interesting are the BeX$_2$ to LB back-donation channels. They delocalize charge directly onto the base, and are formed by the symmetric and antisymmetric combination of σ-like functions mainly localized on the X$_2$ moieties. Turning to a localized NAdOs scheme, if we would like to, they are simply the two σ Be-X bonds. We have found that the σ back-donation capacity of these channels is largest in BeH$_2$, with easily polarizable hydride-like entities. It decreases considerably in BeF$_2$, where the NAdOs are combinations of highly localized non-bonding p orbitals of the F atoms, and it increases again as the fluorines are substituted by more diffuse and polarizable Cl substituents. The relevance of the σ back-donation we are describing in the chemistry of these compounds

11

remains to be examined. To check the generality of the arguments here presented, we have also computed conformations of the σ H_2O complexes where the BeX_2 molecular plane is orthogonal to that of the H_2O molecule. The importance of the back-donation decreases considerably, as expected. In the BeH_2 case, for instance, the effect is so large that the BeH_2 fragment changes polarity, with a net charge of -0.005 a.u.

| $\delta^{AB} = 0.154$ | $\delta^{AB} = 0.121$ | $\delta^{AB} = 0.119$ | $\delta^{AB} = 0.048$ | $\delta^{AB} = 0.017$ | |
| $\delta^{AB} = 0.155$ | $\delta^{AB} = 0.066$ | $\delta^{AB} = 0.054$ | $\delta^{AB} = 0.038$ | $\delta^{AB} = 0.020$ | $\delta^{AB} = 0.030$ |
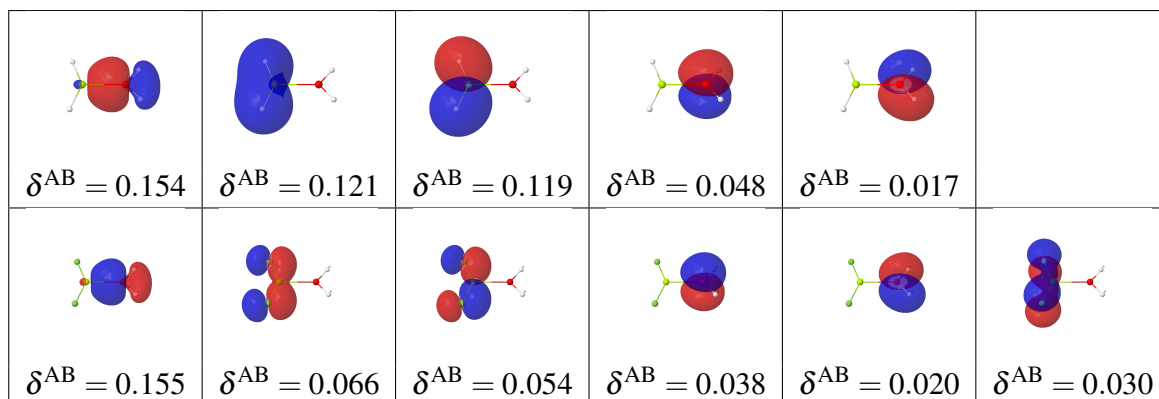
Figure 1: The most highly occupied NAdOs of the $BeH_2 \cdots H_2O$ (top) and $BeF_2 \cdots H_2O$ (bottom) systems. The isosurface shown corresponds to $|\phi| = 0.05$ a.u. Their contribution to the δ^{AB} delocalization index are shown below.

It is easy to check that a reasonable estimate of the total charge transfer can be obtained by: (i) classifying the NAdO channels into those basically localized in the BeX_2 or LB fragments; (ii) adding the occupation numbers of both classes and, finally; (iii) subtracting them to get a grand total. This allows to rationalize easily the oscillating net charge pattern of the complexes found in Table 1 and to construct a donating and back-donating scale of the LB and BeX_2 moieties. For the former CO$\approx$ NH_3 > H_2O, for the latter, BeH_2 > $BeCl_2$ > BeF_2.

π-bonded complexes

Soon after the BeX_2 σ complexes were examined it was found that the BeX_2 species would also embark on π bonding with molecules such as ethylene or acetylene. Although weaker in general than the previously analyzed σ dimers, these π complexes have binding energies typical of hydrogen bonded systems, displaying QTAIM critical points characteristic of π interactions and

being described from the NBO perspective through equivalent $\pi \to p_{\text{Be}}$ and $\pi \to \sigma^*_{\text{BeX}}$ donor-acceptor contributions. An interesting point that has not been thoroughly investigated is the origin of the preference between parallel/perpendicular conformations: BeF_2 and $BeCl_2$ prefer to locate their molecular planes coinciding with the plane that contains the C-C internuclear axis [parallel conformation, (a)] in acetylene complexes, but orthogonal to it [orthogonal conformation, (b)] in ethylene ones, while BeH_2 prefers the parallel conformation in both cases, the difference in energy between the two conformers being rather large in this last case. We have thus investigated both conformations and shown our IQA results also in Table 1.

A first insight is related to the net charge transfer in the dimers. In the BeH_2 case the most stable conformers are those in which this moiety is positively charged. In fact, the perpendicular conformers display much smaller charges (and of opposite signs). All IQA energetic indicators point toward a much stronger interaction (both electrostatic and covalent) as well as larger deformation energies in the (a) conformations. It is also interesting to notice that in these (a) isomers it is the BeH_2 moiety which is mostly deformed. This tells about the role of the hydrogens in the stabilization of the complex, leading again to backdonation issues.

Table 4: Contributions to δ^{AB} greater than 0.01 of the two-fragment NAdOs of $BeX_2 \cdots C_2H_4^{a,b}$ and $BeX_2\text{-}C_2H_2^{a,b}$ (X=H,F) systems. Each number represents the occupation of a natural adaptive orbital (NAdO). See Figure 2 for the cases $BeH_2 \cdots C_2H_4^{a,b}$ and $BeH_2 \cdots C_2H_2^{a,b}$.

BeH$_2$				BeF$_2$				BeCl$_2$			
$C_2H_4^a$	$C_2H_4^b$	$C_2H_2^a$	$C_2H_2^b$	$C_2H_4^a$	$C_2H_4^b$	$C_2H_2^a$	$C_2H_2^b$	$C_2H_4^a$	$C_2H_4^b$	$C_2H_2^a$	$C_2H_2^b$
0.348	0.158	0.393	0.079	0.197	0.199	0.186	0.170	0.234	0.229	0.222	0.198
0.248	0.080	0.236	0.048	0.060	0.057	0.066	0.057	0.116	0.082	0.129	0.083
0.093	0.047	0.099	0.019	0.057	0.040	0.060	0.033	0.080	0.063	0.082	0.049
		0.064		0.021	0.036	0.028	0.032	0.030	0.058	0.041	0.045
						0.021	0.029	0.010	0.024	0.031	0.045

When X=F,Cl, the situation changes. Now the hydrocarbon donation channels dominate, and in every case the net charge of the BeX_2 species is negative. The ability of the *p*-like lone pairs of X to participate in bonding interactions in both conformations makes the balance subtle with ethylene, while the conformer discrimination is more neat with acetylene. In all the conformers

E_{def} of the BeX_2 fragment is considerably smaller than that of the hydrocarbon. For the $BeCl_2$ ethylenic complex an interesting situation arises. The (b) conformer has a smaller total interaction energy than the (a) one, although it is the most stable of the two. Actually, the final energetic difference is close to that in electrostatic interactions, so that deformation energies and covalent contributions cancel their difference for both conformers.

Fig. 2 shows the most relevant NAdOs and their contribution to the inter-fragment delocalization indices for the parallel and orthogonal conformers of BeH_2 with ethylene and acetylene. The NAdOs occupation numbers for the three complexes BeX_2 (X=H,F,Cl) are found in Table 4. It is obvious that the (b) conformations quench the σ back-donation channels from the hydrides to the carbon atoms, which are dominant in the parallel ones and explain the charge transfer observed. Without these channels, the H-Be-H angle is considerably closer to $180°$, and the Be-C distance increases. As a result, also the $\pi \rightarrow BeH_2$ donation decreases in strength.

BeX_2 with X=F,Cl behave differently thanks to the p-symmetry orbitals of the halogen. We only analyze in detail the F complexes, since the conclusions are similar for X=Cl. Fig. 3 collects the relevant NAdOs. The ethylene compounds display four basic delocalization channels. The most intense one (i) is the $\pi_{C_2H_4} \rightarrow BeF_2$ that accounts for the angularization of BeF_2. The rest are $BeF_2 \rightarrow C_2H_4$ back-donation terms that involve the symmetric and antisymmetric combination of F p non-bonding orbitals, i.e. two localized p non-bonding orbitals (iia, iib), plus the symmetric π-symmetry bonding orbital formed by the in-phase combination of the three p_z functions corresponding to the F-Be-F atoms (iii). The balance determining the stability of the (a) and (b) conformers is subtle. In the orthogonal one (b, the most stable) and in agreement with intuition, channel (iii) becomes reinforced, while channels (iia, iib) weaken.

Finally, five channels become relevant in the acetylenic compounds. Fig. 3 shows that now a second $C_2H_2 \rightarrow BeF_2$ component in π conformation becomes relevant, but that all the rest contributions are reinforced by symmetry constraints in the parallel conformer.

We thus come to the conclusion that the combination of IQA with the NAdOs decomposition offers a consistent image of σ and π BeX_2 Beryllium bonds, in which electrostatic interactions

14
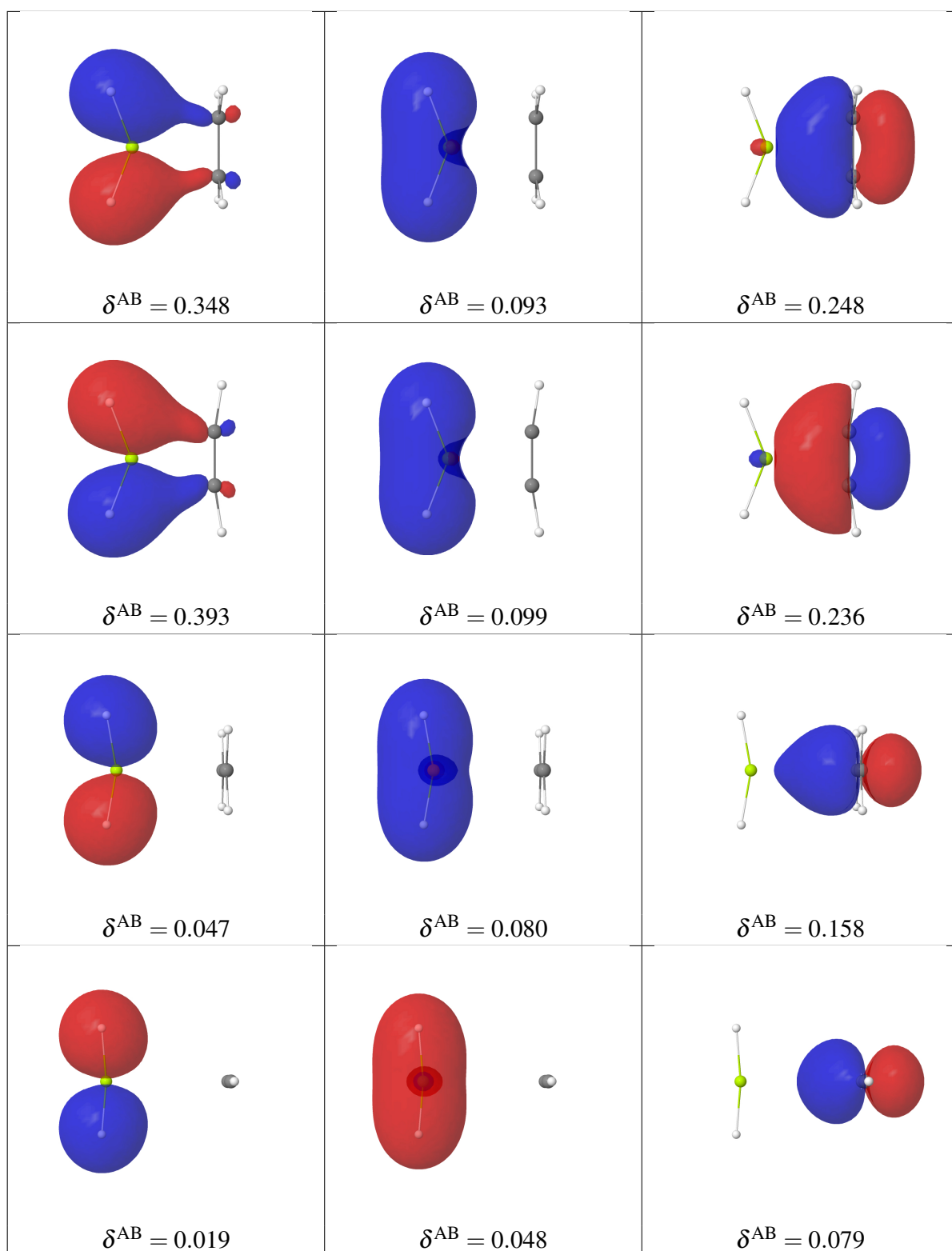
Figure 2: The three most highly occupied NAdOs of the $BeH_2 \cdots LB$ systems, with LB = $C_2H_4{}^a$ (1st row), $C_2H_2{}^a$ (2nd row), $C_2H_4{}^b$ (3rd row), and $C_2H_2{}^b$ (4th row). The isosurface shown corresponds to $|\phi| = 0.05$ a.u. Their contribution to the δ^{AB} delocalization index are shown below.

Figure 3: The most highly occupied NAdOs of the $BeF_2 \cdots$ LB systems, with LB = $C_2H_4{}^a$ (1strow), $C_2H_2{}^a$ (2nd row), $C_2H_4{}^b$ (3rd row), and $C_2H_2{}^b$ (4th row). The isosurface shown corresponds to $|\phi| = 0.05$ a.u. Their contribution to the δ^{AB} delocalization index are shown below.
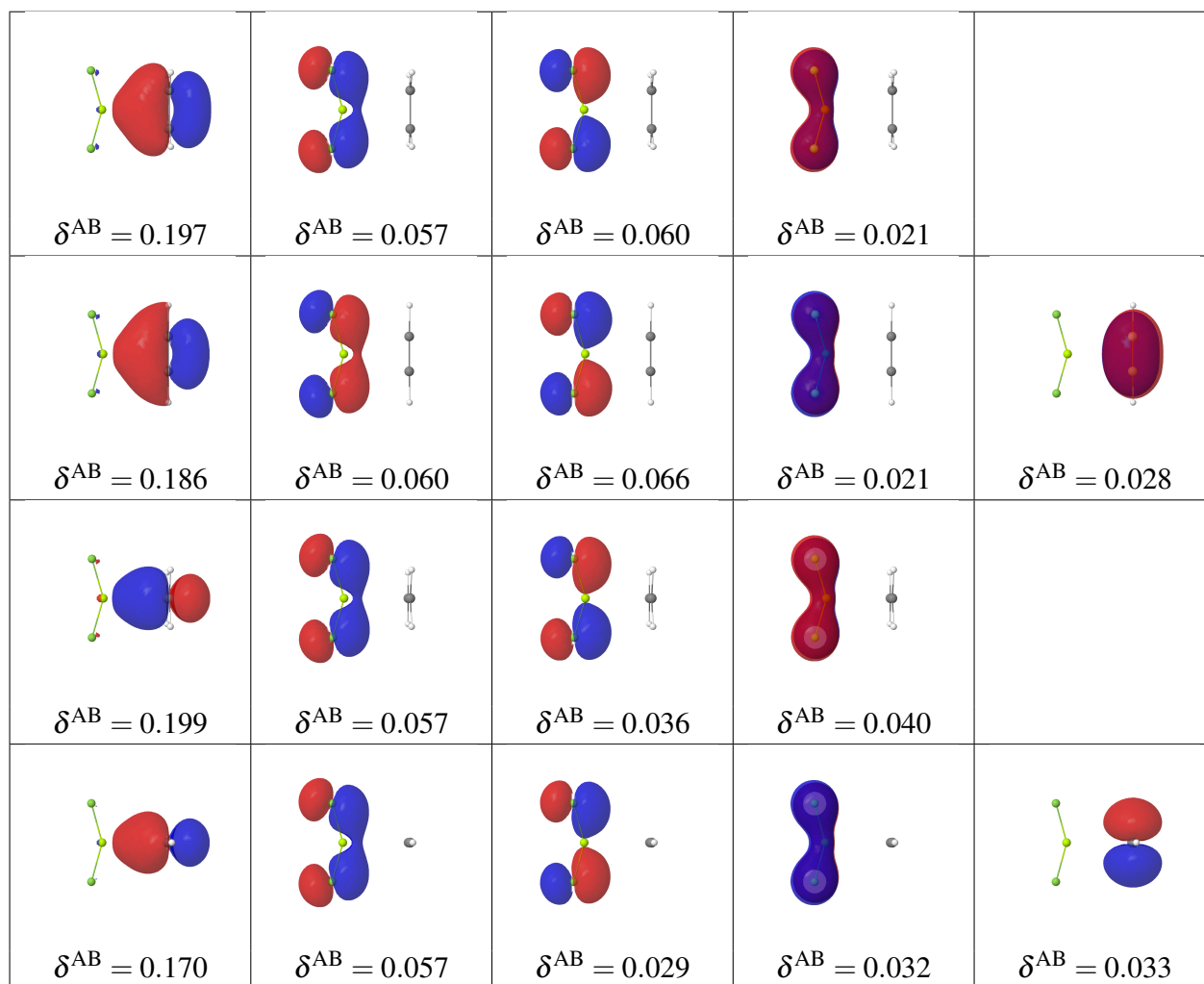
together with covalent contributions account for the bonding properties displayed by these interesting systems. We stress that the use of NBOs may skip important delocalization channels that are essential to understand the observed electron fluxes.

One-electron Beryllium bonds

Beryllium compounds may also become linked via one-electron bonds, which have been rediscovered and analyzed in recent times. We have computed the BeLi and BeNa diatomics, as well as the BeLiBe triatomic system and the 1,8-BeH-disubstituted naphtalene anion that was recently proposed as an example of intramolecular Be-Be one-electron bond,[55] all in their lowest doublet electronic states. Table 5 gathers several IQA properties.

Table 5: Some IQA data for the one-electron bonded LiBe, NaBe, and BeLiBe[a] linear molecules and the $C_{10}H_6Be_2H_2^-$ anion[b]. T^A, T^B, E_{self}^A, and E_{self}^B in atomic units; E_{int}^{AB}, E_{xc}^{AB}, and E_{cl}^{AB} in kcal/mol; Q's in electrons and δ_{AB} in electron pairs.

	T^A	T^B	E_{self}^A	E_{self}^B	E_{int}^{AB}	E_{xc}^{AB}	E_{cl}^{AB}	Q_A	Q_B	δ^{AB}
BeLi	14.723	7.418	-14.557	-7.285	-133.36	-37.47	-95.89	-0.707	0.708	0.466
BeNa	14.642	161.861	-14.594	-161.796	-54.92	-35.40	-19.52	-0.330	0.330	0.550
LiBe[a]	7.401	14.653	-7.286	-14.584	-68.74	-19.27	-49.46	0.738	-0.369	0.223
Be⋯Be[a]	14.653	14.653	-14.584	-14.584	0.42	-13.74	14.17	-0.369	-0.369	0.323
Be⋯Be[b]	14.316	14.316	-14.114	-14.114	138.43	-38.75	177.18	1.258	1.258	0.293

We first point out that, in agreement with electronegativity arguments, the Be atom gets a negative net charge that accommodates the positively charged alkali atom. This is immediately translated into the Be self-energy, which is progressively destabilized as its negative charge increases. This implies that negatively charging Be has an energy cost, as expected. This conclusion is contrary to that obtained if atomic energies as measured by minus the atomic kinetic energies are used. We warn against using the latter to get chemical insights. Interaction energies are revealing. The total LiBe interaction is not negligible, about -130 kcal/mol, and is dominated by a rather large electrostatic component (-95 kcal/mol), although covalency cannot be ignored. On moving to BeNa, the covalent component stays almost exactly the same within a couple of kcal/mol,

although the electrostatic term is greatly decreased. This is telling us about a rather common electron sharing mechanism, also revealed by the almost 1/2 delocalization index which points toward a single one-electron delocalization channel.

Fig. 4, where $\nabla^2\rho(\boldsymbol{r})$ is shown for the BeLi molecule, reveals an easy to recognize pattern in highly ionic molecules. The Li atom has lost its valence shell, and the $\nabla^2\rho < 0$ region of the Be atom is now heavily polarized toward the Li moiety. We have superimposed the electrostatic potential mapped onto the $\rho = 0.03$ isosurface to show how the standard concepts used in intermolecular interactions can be directly translated to the present examples, although we use a much larger value of ρ in this case. The Be atom has a core-like positive ESP region that corresponds well to a laplacian depletion zone, with electron rich regions to its sides. This agrees with a $s - p_z$-like hybridization of the one-electron bond that will come out clear in the following. Notice that the polarization of the Be $\nabla^2\rho < 0$ domain coincides with that of the ESP. On the contrary the Li core corresponds to a slightly polarized positive ESP region which faces the electron rich Be. This picture qualitatively points toward an important role of electrostatics in BeLi bonding. IQA easily quantifies it. Similar images are found for the BeNa case.

Even more interesting is the consideration of the BeLiBe triatomic. In this case, the sum of both $E_{\mathrm{xc}}^{\mathrm{LiBe}}$ energies adds to almost exactly the same value as in the LiBe diatomic. The same can be said about the total interaction energy. Each of the two LiBe interactions is *halved* with respect to LiBe. In chemical terms, we are sharing the electron between three regions, as we will see, and the delocalization index between each LiBe pair is again halved. This consistency is one of the important properties of IQA. In BeLiBe, the one-electron bond is tricentric: the three-center delocalization index ($\delta^{\mathrm{BeLiBe}} = 0.186$)[41] is one of the largest ever reported.

More insight about the electronic structure of these compounds is obtained by examining their EDFs. Table 6 summarizes de probability of the different real space resonance structures. In BeLi and BeNa it is rather clear that, except residual contributions, the distribution is dominated by two structures. This is a obviuos sign of a one-electron delocalization. In BeLi, the delocalized electron is heavily polarized toward the Be atom, $p(n_{\mathrm{Be}} = 5, n_{\mathrm{Li}} = 2) = 0.7$. This polarization
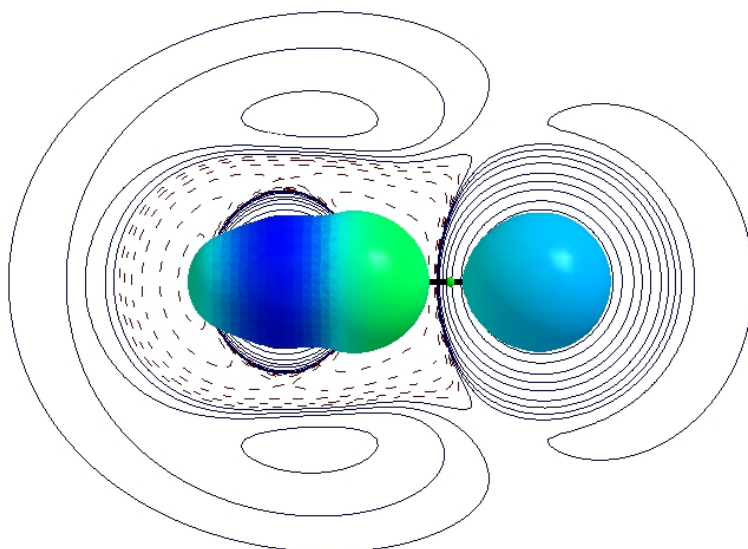
Figure 4: $\nabla^2\rho(\boldsymbol{r})$ for the BeLi molecule in a plane containing the nuclei. An isosurface of $\rho = 0.03$ a.u. with the electrostatic potential mapped onto it has been superimposed. The Be atom is on the left side of the figure, and the BeLi bond critical point of the electron density is marked as a small green sphere.

Table 6: Electron distribution functions for the one-electron bonded LiBe, NaBe, and BeLiBe linear molecules. The atoms are labelled in the order in which they are written. In this sense, n_A, n_B refer to the populations in Be and Li in the BeLi moiety. All data in atomic units.

BeLi			BeNa			BeLiBe			
n_A	n_B	$p(n_A, n_B)$	n_A	n_B	$p(n_A, n_B)$	n_A	n_B	n_C	$p(n_A, n_B, n_C)$
5	2	0.710	4	11	0.619	4	2	5	0.370
4	3	0.275	5	10	0.354	5	2	4	0.370
3	4	0.009	3	12	0.025	4	3	4	0.219
6	1	0.004	6	9	0.001	5	3	3	0.013
						3	3	5	0.013

decreases in BeNa. The BeLiBe data show again that the bonding electron is delocalized among the three atoms, slightly more polarized toward the Be ends and providing a very neat image of a three-center one-electron bond.

$$\delta^{\text{BeLi}} = 0.393 \qquad\qquad \delta^{\text{BeNa}} = 0.464 \qquad\qquad \delta^{\text{BeLiBe}} = 0.180$$
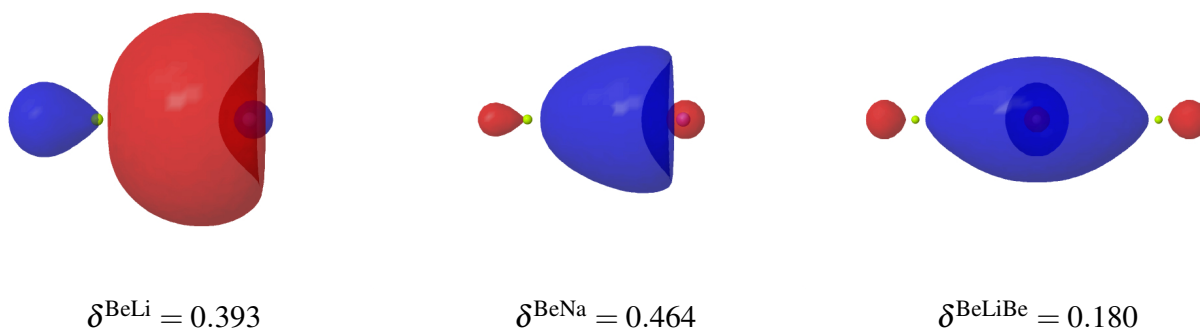
Figure 5: One-electron NAdOs for the BeLi (left), BeNa (middle), and BeLiBe (right) systems. The isosurface shown corresponds to $|\phi| = 0.05$ a.u. Their contribution to the two-center (BeLi and BeNa) or three-center delocalization indices (BeLiBe) are shown below.

A pictorial glimple of the one-electron bonds is shown in Fig. 5 where the main NAdO component is shown. Only one exchange channel (NAdO) accounts for almost all of the final electron delocalization. The only exception is BeLi, where a second smaller contribution, basically a polarized Be 2s orbital is also found. The one-electron bond is a combination of Be sp_z-like hybrids and polarized Li 2s or Na 3s functions, as expected.

We have also computed the 1,8-BeH disubstituted naphtalene anion, in which a relatively strong one-electron intramolecular Be-Be bond was recently reported.[55] In agreement with previous knowledge, a bond critical point between the Be atoms appears, with $\rho = 0.033$ a.u. and $\nabla^2\rho = -0.041$ a.u. As seen in Table 5 both Beryllium atoms are considerably positively charged, this leading to a large electrostatic destabilization. We thus have in this case a homonuclear-like interaction, which is only partially stabilized by a covalent term of about -39 kcal/mol, with a delocalization index close to that shown in BeLiBe. To understand the nature of the Be···Be stabilization it is necessary to consider the interaction of the two $-\text{C}_{\text{naph}}\text{BeH}$ fragments, where negatively charged groups (like the H atoms) compensate the Be···Be electrostatic destabilization. This can be done easily in IQA, resulting in a total interaction energy between the fragments of

about -62 kcal/mol, with only -8 kcal/mol coming from electrostatic interaction. Actually, the two $-C_{naph}BeH$ moieties display, as a whole, a net charge of only -0.469 extra electrons, meaning that half the anionic charge is delocalized between the two $-C_{naph}BeH$ fragments. All this tells about a rather important energy stabilization coming from the formation of the Be-Be bond.
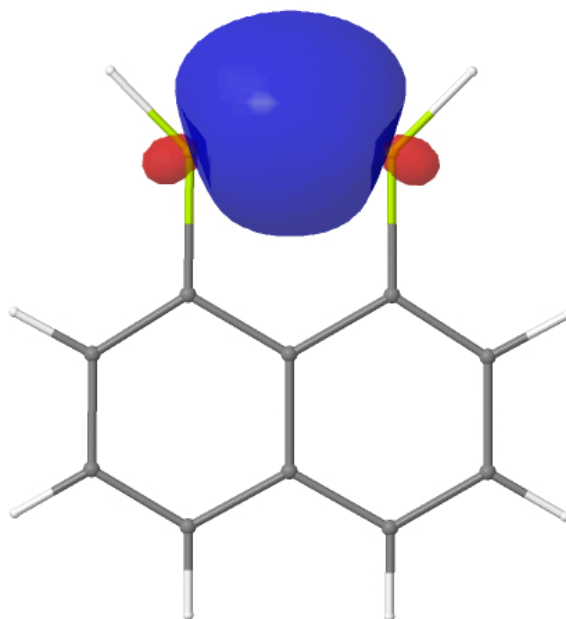
Figure 6: One-electron Be-Be NAdO for the $C_{10}H_6Be_2H_2^{-}$ anion. The isosurface shown corresponds to $|\phi| = 0.05$ a.u.

Fig. 6 shows the Be-Be one-electron NAdO obtained from our DFT calculation. It shows a very clear in-phase combination of sp^2-like hybrids, in agreement with previous knowledge. It accounts for most of the Be-Be delocalization index.

Altogether, these examples show that the same tools that can be used in the case of intermolecular or weaker interactions apply equally well to more strongly bound systems. Taking BeLi, for instance, charge transfer *is not in the model*, but a very real phenomenon. The probability that taking a *snapshot* of the system we find a Be^{-} moiety is 0.7. At the same time, the ESP, the density, or the laplacian of the density, all point in this direction. Evaluation of the exact electrostatics coming from the in-the-molecule electron density provides an important part of the BeLi interac-

tion energy, but falls short of its total value. The lacking ingredient is delocalization or covalency, which corrects the electrostatic interaction because some electrons that contribute to the density (and to E_{cl}) are counted several times due to their wandering nature, in other words, because the pair density is not contained in the density. Analyzing in detail how delocalization takes place we can go back to standard molecular orbital arguments, all from observable quantities. If this is accepted, smaller intermolecular charge transfers do also have to, without models.

Conclusions

We have used in this contribution several recent quantum chemical topology tools, including the interacting quantum atoms (IQA) approach and the effective one-electron pictures provided by two-fragment natural adaptive orbitals or electron distribution functions, to shed new light on the nature of the interactions in some typical Beryllium-bonded systems. These tools add to the traditional local topological approaches, complementing them. The real space energetic partitioning of IQA shows that, in agreement with previous works, it is the electrostatic interaction between the bonded fragments that accounts for the basis energetics as well as conformational preferences in the compounds examined. However, in the absence of delocalization (covalent) contributions, the deformation energy of the fragments is not overcome by electrostatics. In this sense, and similarly to what was found in simple hydrogen-bonded systems,[53] covalency is essential for the stability of the aggregates. An overlooked aspect related to the total charge transfer has been analyzed in detail. This real space charge transfer is free from the criticisms posed by Politzer *et al.*[10]. Our results clearly show that there exist important delocalization channels that are not properly accounted for by the natural bond orbital (NBO) formalism, involving σ back-donation from BeX_2 moieties to the Lewis base which can even be dominant, leading to negative net LB charges. The strength of the different forward and back-bonding channels is predictable from the characteristics of the fragments involved, leading to a chemically appealing rationalization of conformational preference. An analysis of some toy Beryllium containing molecules displaying one-electron bonds has

also been undertaken. EDFs and NAdOs show that the one-electron bonding model accurately describes their bonding.

Supporting Information

Optimized geometries, Electron Distribution Functions (EDFs), and Natural Adaptive Orbitals (NAdOs) of the molecules studied in this work.

Acknowledgments

References

(1) Scnheider, H.-J. Binding Mechanisms in Supramolecular Complexes. *Angew. Chem. Int. Ed.* **2009**, *48*, 3924–3977.

(2) Hobza, P.; Müller-Dethlefs, K. *Non-Covalent Interactions: Theory and Experiment*; Royal Society of Chemistry: Cambridge, 2010.

(3) Novoa, J., Ed. *Intermolecular Interactions in Crystals: Fundamentals of Crystal Engineering*; The Royal Society of Chemistry: Cambridge, 2017.

(4) Jeffrey, G. *An Introduction to Hydrogen Bonding*; 1997: New York, Oxford University Press.

(5) Desiraju, G.; Steiner, T. *The Weak Hydrogen Bond in Structural Chemisry and Biology*; Oxford University Press: New York, 1999.

(6) Buckingham, A. D.; Fowler, P. W.; Jutson, J. M. Theoretical Studies of van der Waals Molecules and Intermolecular Forces. *Chem. Rev.* **1988**, *88*, 963–988.

(7) Spackman, M. A. A Simple Quantitative Model of Hydrogen Bonding. *J. Chem. Phys.* **1986**, *85*, 6587–6601.

(8) Weinhold, F. Nature of H-Bonding in Clusters, Liquids, and Enzymes: an Ab Initio, Natural Bond Orbital Perspective. *J. Mol. Struct.* **1997**, *398–399*, 181–197.

(9) Isaacs, E. F.; Shukla, A.; Platzman, P. M.; Hamann, D. R.; Barbiellini, B.; Tulk, C. A. Covalency of the Hydrogen Bond in Ice: A Direct X-Ray Measurement. *Phys. Rev. Lett.* **1999**, *82*, 600–603.

(10) Politzer, P.; Murray, J. S.; Clark, T. Mathematical Modeling and Physical Reality in Noncovalent Interactions. *J. Mol. Model.* **2015**, *21*, 52.

(11) Murray, J. S.; Politzer, P. The Electrostatic Potential: An Overview. *WIREs Comput. Mol. Sci.* **2011**, *1*, 153–163.

(12) Weinhold, F.; Landis, C. *Valency and Bonding. A Natural Bond Orbital Donor-Acceptor Perspective*; Cambridge Univ. Press, 2005.

(13) Arunan, E.; Desiraju, G. R.; Klein, R. A.; Sadlej, J.; Scheiner, S.; I., A.; C., C. D.; Crabtree, R. H.; Dannenberg, J. J.; Hobza, P. et al. Defining the Hydrogen Bond: An Account. *Pure Appl. Chem.* **2011**, 1619–1636.

(14) Yañez, M.; Sanz, P.; Mó, O.; Alkorta, I.; Elguero, J. Beryllium Bonds, Do They Exist? *J. Chem. Theory Comput.* **2009**, *5*, 2763–2771.

(15) Fernández Villanueva, E.; Mó, O.; Yañez, M. On the Existence and Characterization of $\pi-$Beryllium Bonds. *Phys. Chem. Chem. Phys.* **2014**, *16*, 17531–17536.

(16) Brea, O.; Mó, O.; Yañez, M.; Alkorta, I.; Elguero, J. Creating $\sigma-$Holes through the Formation of Beryllium Bonds. *Chemistry - A European Journal* **2015**, *21*, 12676–12682.

(17) Gimarc, B. M. *Molecular Structure and Bonding. The Qualitative Molecular Orbital Approach*; Academic Press: New York, 1979.

(18) Su, P.; Li, H. Energy Decomposition Analysis of Covalent Bonds and Intermolecular Interactions. *J. Chem. Phys.* **2009**, *131*, 014102.

(19) Bader, R. F. W. *Atoms in Molecules*; Oxford University Press: Oxford, 1990.

(20) D., B. A.; Edgecombe, K. E. A Simple Measure of Electron Localization in Atomic and Molecular Systems. *J. Chem. Phys.* **1990**, *92*, 5397–5403.

(21) Mitoraj, M.; Michalak, A. Natural Orbitals for Chemical Valence as Descriptors of Chemical Bonding in Transition Metal Complexes. *J. Mol. Model.* **2007**, *13*, 347–355.

(22) Michalak, A.; Mitoraj, M.; Ziegler, T. Bond Orbitals from Chemical Valence Theory. *J. Phys. Chem A.* **2008**, *112*, 1933–1939.

(23) Menéndez, M.; Álvarez-Boto, R.; Francisco, E.; Martín Pendás, A. One-Electron Images in Real Space: Natural Adaptive Orbitals. *J. Comput. Chem.* **2015**, *36*, 833–843.

(24) Chen, X.; Grammer, J.; Lawson, J.; Cooke, R.; Pate, R., E. Yount A Novel Restricted Photoaffinity Spin-Labeled Non-Nucleoside ATP Analogue as a Covalently Attached Reporter Group of the Active Site of Myosin Subfragment 1. *Biochemistry* **2002**, *41*, 2609–2620.

(25) Shibuya, H.; Kondo, K.; Kimura, N.; Maruta, S. Formation and Characterization of Kinesin.ADP.Fluorometal Complexes. *J. Biochem.* **2002**, *132*, 573–579.

(26) Danko, S.; Yamasaki, K.; Daiho, T.; Suzuki, H. Distinct Natures of Beryllium Fluoride-bound, Aluminum Fluoride-bound, and Magnesium Fluoride-bound Stable Analogues of an ADP-insensitive Phosphoenzyme Intermediate of Sarcoplasmic Reticulum $Ca^{2+}-$ATPase CHANGES IN CATALYTIC AND TRANSPORT SITES DURING PHOSPHOENZYME HYDROLYSIS. *J. Biol. Chem.* **2004**, *279*, 14991–14998.

(27) Sauna, Z.; Nandigama, K.; Ambudkar, S. Exploiting Reaction Intermediates of the ATPase Reaction to Elucidate the Mechanism of Transport by P-glycoprotein (ABCB1). *J. Biol. Chem.* **2006**, *281*, 26501–26511.

(28) Tiago, T.; Simao, S.; Aureliano, M.; Martin-Romero, F.; Gutierrez-Merino, C. Inhibition of Skeletal Muscle S1-Myosin ATPase by Peroxynitrite. *Biochemistry* **2006**, *45*, 3794–3804.

(29) Orban, J.; Lorinczy, D.; Hild, G.; Nyitrai, M. Noncooperative Stabilization Effect of Phalloidin on ADP.BeF$_x-$ and ADP.AlF$_4-$Actin Filaments. *Biochemistry* **2008**, *47*, 4530–4534.

(30) Liu, F.; Putnam, A.; Jankowsky, E. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci.* **2008**, *105*, 20209–20214.

(31) Barbaro, P.; Cecconi, F.; Dakternieks, D.; Dominguez, S.; Duthie, A.; Ghilardi, C.; Midollini, S.; Orlandini, A.; Vacca, A. Beryllium(II) Complexes of the Kläui Tripodal Ligand Cyclopentadienyltris(diethylphosphito-P)cobaltate($-$). *Inorg. Chem.* **2003**, *40*, 2725–2729.

(32) Conejo, M.; Fernández, R.; Carmona, E.; Andersen, R.; Gutiérrez-Puebla, E.; Monge, M. Synthetic, Reactivity, and Structural Studies on Half-Sandwich (η^5-C$_5$Me$_5$)Be and Related Compounds: Halide, Alkyl, and Iminoacyl Derivatives. *Chem. Eur. J.* **2003**, *9*, 4462–4471.

(33) Schmidbaur, H. *Coord. Chem. Rev.* **2001**, *215*, 223–241.

(34) Scott, B.; McCleskey, T.; Chaudhary, A.; Hong-Geller, E.; Gnanakaran, S. *Chem. Commun.* **2008**, *371*, 2837–2847.

(35) Gatti, C., Macchi, P., Eds. *Modern Charge-Density Analysis*; Springer, Dordrecht., 2012.

(36) Matta, C. F., Boyd, R. J., Eds. *The Quantum Theory of Atoms in Molecules*; Wiley-VCH Verlag GmbH & Co. KGaA, 2007.

(37) Francisco, E.; Martín Pendás, A.; García-Revilla, M.; Álvarez Boto, R. A Hierarchy of Chemical Bonding Indices in Real Space from Reduced Density Matrices and Cumulants. *Comput. Theor. Chem.* **2013**, *1003*, 71–78.

(38) Blanco, M. A.; Martín Pendás, A.; Francisco, E. Interacting Quantum Atoms: A Correlated Energy Decomposition Scheme based on the Quantum Theory of Atoms in Molecules. *J. Chem. Theory Comput.* **2005**, *1*, 1096–1109.

(39) Francisco, E.; Martín Pendás, A.; Blanco, M. A. A Molecular Energy Decomposition Scheme for Atoms in Molecules. *J. Chem. Theory Comput.* **2006**, *2*, 90–102.

(40) Cancès, E.; Keriven, R.; Lodier, F.; Savin, A. How Electrons Guard the Space: Shape Optimization with Probability Distribution Criteria. *Theor. Chem. Acc.* **2004**, *111*, 373–380.

(41) Francisco, E.; Martín Pendás, A.; Blanco, M. A. Electron Number Probability Distributions for Correlated Wave Functions. *J. Chem. Phys.* **2007**, *126*, 094102.

(42) Martín Pendás, A.; Francisco, E.; Blanco, M. A. Spin Resolved Electron Number Distribution Functions: How spins Couple in Real Space. *J. Chem. Phys.* **2007**, *127*, 144103.

(43) Jayatilaka, D.; Grimwood, D. J. Wavefunction Derived from Experiment. I. Motivation and Theory. *Acta Cryst. A.* **2001**, *57*, 76–86.

(44) Grabowsky, S.; Genoni, A.; Bürgi, H.-B. Quantum Crystallography. *Chem. Sci.* **2017**, *8*, 4159–4176.

(45) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. et al. General Atomic and Molecular Electronic Structure System. *Journal of Computational Chemistry* **1993**, *14*, 1347–1363.

(46) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J.; Sharma, S.; Wouters, S. et al. The Python-based Simulations of Chemistry Framework (PySCF). **2017**,

(47) Francisco, E.; Casals-Sainz, J. L.; Rocha-Rinza, T.; Martín Pendás, Partitioning the DFT Exhange-Correlation Energy in Line with the Interacting Quantum Atoms Approach. *Theor. Chem. Acc.* **2016**, *135*, 170–177.

(48) Knizia, G.; Chan, G. K.-L. Density Matrix Embedding: A Simple Alternative to Dynamical Mean-Field Theory. *Phys. Rev. Lett.* **2012**, *109*, 186404.

(49) Martín Pendás, A.; Francisco, E. PROMOLDEN: A QTAIM/IQA code (Available from the authors upon request by writing to ampendas@uniovi.es).

(50) Francisco, E.; Martín Pendás, A. Electron Number Distribution Functions from Molecular Wavefunctions. Version 2. *Comput. Phys. Commun.* **2014**, *185*, 2663–2682.

(51) Francisco, E.; Martín Pendás, A. DENMAT: Density Matrices and Natural Adaptive Orbitals (Available from the authors upon request by writing to evelio@uniovi.es).

(52) Eskandari, K. Nature of Beryllium Bonds in View of Interacting Quantum Atoms and Natural Energy Decomposition Analyses. *Comput. Theor. Chem.* **2016**, *1090*, 74–79.

(53) Martín Pendás, A.; Blanco, M. A.; Francisco, E. The Nature of the Hydrogen Bond: A Synthesis from the Interacting Quantum Atoms Picture. *J. Chem. Phys.* **2006**, *125*, 184112.

(54) Martín Pendás, A.; Blanco, M. A.; Francisco, E. Steric Repulsions, Rotation Barriers, and Stereoelectronic Effects: A Real Space Perspective. *J. Comput. Chem.* **2009**, *30*, 98–109.

(55) Brea, O.; Mó, O.; Yañez, M.; Alkorta, I.; Elguero, J. On the Existence of Intramolecular One-Electron Be-Be Bonds. *Chem. Commun.* **2016**, *52*, 9656–9659.

Challenging the electrostatic σ-hole picture of halogen bonding using minimal models and the interacting quantum atoms approach

Fernando Jiménez-Grávalos[a], Miguel Gallegos[a], Ángel Martín Pendás[a,*], Alexander S. Novikov[b,**]

[a]*Departament of Analytical and Physical Chemistry, University of Oviedo, E-33006, Oviedo, Spain.*
[b]*Saint Petersburg State University, Universitetskaya Nab. 7/9, 199034 Saint Petersburg, Russian Federation*

Abstract

Among the different non-covalent interactions, halogen bonds have captured wide attention in the last years. Their stability has been rationalised in electrostatic terms by appealing to the σ-hole concept, a charge-depleted region that is able to interact favourably with electron rich moieties. This interpretation has been questioned, and in this work a set of anionic halogen model systems are used to shed some light on this issue. We use the Interacting Quantum Atoms (IQA) method, which provides an orbital invariant energy decomposition in which pure electrostatic terms are well isolated, and we complement our insights with the analysis of electrostatic potentials as well as with traditional descriptors of charge accumulation like the Laplacian of the electron density. The total electrostatic interaction between the interacting species is surprisingly destabilising in many of the systems examined, demonstrating that although σ-holes might be qualitatively helpful, much care has to be taken in ascribing the stability of these systems to electrostatics. It is clearly shown that electron delocalisation is essential to understand the stability of the complexes. The evolution of atomic charges as the aggregates form reveals a charge transfer picture in which the central, σ-hole bearing halogen acts as a mere spectator. These systems may then be not far from engaging in a classical 3c-4e interaction. Since the presence of a σ-hole as characterised by the electrostatic potential mapped on a suitable molecular envelope isosurface does not guarantee attractive electrostatic interactions, we encourage to employ a wider perspective that takes into account the full charge distribution.

---

*To whom correspondence should be addressed: ampendas@uniovi.es
**To whom correspondence should be addressed: a.s.novikov@spbu.ru

1. Introduction

The consideration of non-covalent interactions (NCIs) is one of the priority development directions in modern chemistry and related areas of natural science. The attention of the scientific community to this field has grown explosively in the past decade, as these weak contacts play key roles in most chemical[1–3] and biochemical[4,5] processes. Non-covalent interactions determine the existence of molecular solids and the properties of molecular systems in the gas and liquid phases, and their control has found wide application in nanoscience,[6,7] material science,[8,9] medicine[10,11] or catalysis,[12–16] to name just a few fields. As progress accumulates, the number of purportedly new NCIs beyond the ubiquitous van der Waals attractions has increased considerably, and the NCI zoo now includes beryllium, tetrel, pnictogen, chalcogen, etc. bonds.[17–20] In many of these, electrostatic interactions have been targeted as fundamental, and the concept of σ-holes has become instrumental,[21–23] particularly in halogen bonds.[24,25]

It has thus become generally accepted that the charge density anisotropy of bonded halogen atoms leading to σ-holes is the driving force behind halogen bonding. This has lead to emphasise the role of the molecular electrostatic potential (ESP) in understanding these interactions,[26] and to assume that their energetics should be well described by classical contributions.[27] Nonetheless, such a traditional view has been challenged in the last few years with the use of other complementary techniques.

In this respect, numerous computational studies have appeared trying to shed light on the nature of halogen-bond (XB) interactions. Most are based on orbital-based energetic decompositions that use natural bond orbitals (NBO), Kohn-Sham based energy decomposition analyses (EDA), block-localised wave functions (BLWs), in the realm of valence bond (VB) theory, or even symmetry adapted perturbation theory (SAPT).[22,25,28–33] Less effort has been put in rationalising halogen bonding with orbital invariant real space-based alternatives, such as the interacting quantum atoms (IQA) scheme.[34,35] Fock-space analyses already warned about the importance of non-purely electrostatic contributions, such as charge transfer, polarisation, or dispersion in different complexes.[22,25,28–33] On the other hand, the IQA picture has also revealed how in some instances covalent effects, as revealed by non-negligible exchange-correlation energies, may have been skipped.[32,36–44]

In this work, we have decided to adhere to Occam's razor arguments [Johannes Poncius' commentary on John Duns Scotus' Opus Oxoniense, book III, dist. 34, q. 1. in John Duns

Scotus Opera Omnia, vol. 15, Ed. Luke Wadding, Louvain (1639), reprinted Paris: Vives, (1894) p.483a] and thus to choose the simplest possible model systems of halogen bonding in order to isolate as well as possible the driving forces behind their stability through the IQA partitioning. Since XBs comprise a σ-hole bearing halogen atom acting as a Lewis acid that interacts with a Lewis base counterpart, we opted to explore the σ-hole halogen interactions within negatively charged $X_1 - Y_2 - X_3$ species, using different combinations of chlorine, bromine and iodine as the X and Y moieties.

According to the early work of Sakurai[45] and Desijaru,[46] and the seminal paper by Espinosa and coworkers on halogen-halogen interactions,[47] there would be two distinct geometrical arrangements for two interacting halogens in crystals. The first one, called type I, refers to a symmetric, often linear arrangement that occurs mainly around a crystallographic inversion centre and is generally regarded to be dispersion driven. In the second, or type II, the interaction occurs at angles close to 90°, and the positively-charged σ-hole from one atom approaches the negative lone pair belt of the other, so that electrostatic terms are expected to dominate. Type II XBs are often found on crystallographic screw axes and glide planes. In order to capture the mutual effect of maximised van der Waals forces in type I and electrostatic interactions in type II geometries, we selected linear geometries between a halogen diatomic molecule and an approaching halide ion.

To decrease the number of degrees of freedom as much as possible, we have inspected both the potential energy surface minima, corresponding to the symmetric $[X_1 - Y_2 - X_3]^-$ trihalide configurations as found in the gas phase (Figure 1b), and constrained $X_1 - Y_2 \cdots X_3^-$ asymmetric geometries that mimic the halogen arrangements found in crystals (Figure 1a). These systems are close to those studied by Wolters and Bickelhaupt with Kohn-Sham EDAs,[48] that were also chosen thanks to their capturing of the essence of σ-holes and XBs.

Thus, and with all the previous simplifications, we seek to compare the role of σ-holes in XBs from different perspectives: the qualitative picture provided by the ESP and other scalar fields like the Laplacian of the electron density, and the quantitative image given by the IQA energetic decomposition. IQA has now been used to shed light on a wide variety of phenomena, including some recent controversies such the nature of bonding in anion-π interactions or the character of the Na-B bond in $NaBH_3^-$ to cite a couple of them.[49,50] This together with its orbital-invariant nature makes it an appropriate choice for our study. In doing so, we expect to gain deeper insight into the nature of halogen bonding as well as to check whether the electrostatic view

survives quantitative analyses.

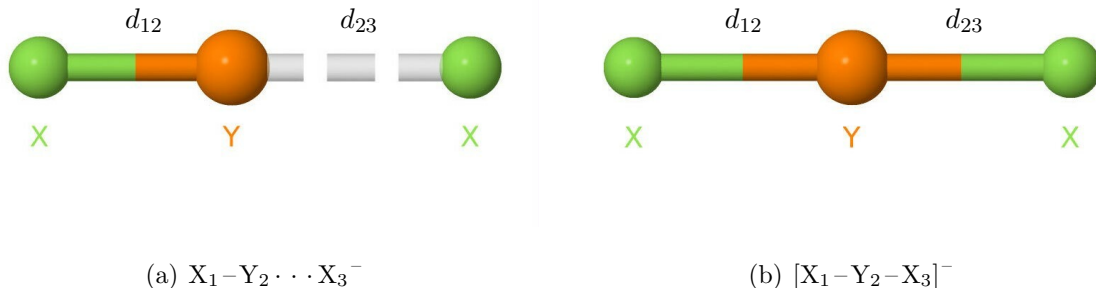(a) $X_1 - Y_2 \cdots X_3{}^-$ (b) $[X_1 - Y_2 - X_3]^-$

Figure 1: Geometries considered in this work: constrained asymmetric complex (left), with $d_{12} < d_{23}$, and symmetric trihalide (right) structure, with $d_{12} = d_{23}$.

2. Methodology

Quantum Chemical Topology (QCT) refers to a set of wave function analysis techniques that take advantage of the topology induced by the gradient field of orbital invariant descriptors.[51] By construction, QCT is independent on the electronic structure method selected to build the system's wave function. Amongst the most well-known QCT procedures we find the Quantum Theory of Atoms in Molecules (QTAIM) of Bader et al.,[52] which provides an atomic-like exhaustive partition of real space and of the expectation values of any quantum mechanical operator. In the QTAIM, atoms are defined as the attractor basins of the electron density field, $\rho(\boldsymbol{r})$. A general energy partition scheme in QCT that leads to an atomic and interatomic decomposition of the molecular energy when applied to the QTAIM is the interacting quantum atoms (IQA) approach.[34,35] IQA/QTAIM provides physically rigorous domain kinetic energies, in contrast to those offered by other spatial decompositions, and it needs only the first- and second-order reduced density matrices to compute its different energy terms. According to this scheme, and making use of its usual language, the total energy E is split into a sum of atomic (net) energies E_{net}^A and pairwise additive interaction terms E_{int}^{AB} as

$$E = \sum_A E_{net}^A + \sum_{A<B} E_{int}^{AB}. \tag{1}$$

The atomic net-energy comprises, on the one hand, the kinetic energy of the electrons contained in a particular basin, their mutual interaction, and the attraction with the particular

nucleus they are associated with. On the contrary, E_{int}^{AB} gathers the interactions of all the electrons and nuclei located in a basin A with those located in basin B.

There are two sources for the interaction energy, namely the classical electrostatic interaction between the total (nuclear and electronic) charge densities lying in the A and B regions, and a correction to that term which is purely quantum mechanical in nature. Accordingly, the interaction energy can be separated into a Coulomb, or classical contribution E_{cl}^{AB} —that has been shown to correspond to the ionic component of a chemical interaction— and an exchange-correlation one E_{xc}^{AB} —that measures covalency:

$$E_{int}^{AB} = E_{cl}^{AB} + E_{xc}^{AB}. \tag{2}$$

Related with the exchange-correlation energy is the delocalisation index, DI. It corresponds to the integration of the exchange-correlation density over two basins and is the real space equivalent of bond order, whose value can be interpreted as the number of electron pairs that two particular basins share.[53]

Since the union of QTAIM regions is again a proper QTAIM domain, we can gather sets of QTAIM atoms to form groups $\mathcal{G}$ and rewrite the IQA partitioning in terms of them:

$$E = \sum_{\mathcal{G}} E_{net}^{\mathcal{G}} + \sum_{\mathcal{G} < \mathcal{H}} E_{int}^{\mathcal{G}\mathcal{H}}, \tag{3}$$

with

$$E_{net}^{\mathcal{G}} = \sum_{A \in \mathcal{G}} E_{net}^{A} + \sum_{\substack{A < B \\ A,B \in \mathcal{G}}} E_{int}^{AB} \tag{4}$$

$$E_{int}^{\mathcal{G}\mathcal{H}} = \sum_{\substack{A < B \\ A \in \mathcal{G}, B \in \mathcal{H}}} E_{int}^{AB}. \tag{5}$$

Here, $\mathcal{G}$ and $\mathcal{H}$ stand for the different groups in which the molecule has been divided. The same grouping can be performed for each of the different IQA energy components, such as the classical or exchange-correlation contributions.

The unavailability of a second-order reduced density matrix in DFT prevents the exact decomposition of the exchange-correlation energy in an IQA-like manner, hence DFT would not be in principle compatible with IQA. Nevertheless, several working approximations have been introduce to by-pass this limitation. Here we will consider a scaling technique that modifies both intra- and interatomic contributions.[54]

Given a particular exchange-correlation functional with a non-hybrid part $\epsilon(\mathbf{r})$, the total DFT exchange-correlation (xc) energy can be easily computed as

$$E_{xc}^{DFT} = \int_{\mathbb{R}^3} \rho(\mathbf{r})\epsilon(\mathbf{r})d\mathbf{r} + a_0 E_x^{KS}, \tag{6}$$

where a_0 denotes the fraction of the pure Hartree-Fock (HF) exchange E_x^{KS}, which is directly computed from the Kohn-Sham (KS) orbitals when a hybrid functional is used.

By delimiting the previous integral to a particular spacial basin (e.g., Bader's atomic regions, with $\mathbb{R}^3 = \cup_A \Omega_A$), a splitting of the total xc energy is achieved, but such terms involve both IQA intra- and inter-basin contributions. To overcome this problem a scaling technique based on the E_{xc}^{DFT}/E_x^{KS} ratio over each basin was successfully proposed.[54] First, the parameters λ_A must be computed for each atomic basin as

$$\lambda_A = \frac{E_{xc,add}^{DFT,A}}{E_{x,add}^{KS,A}} = a_0 + \frac{1}{E_{x,add}^{KS,A}} \int_{\Omega_A} \rho(\mathbf{r})\epsilon(\mathbf{r})d\mathbf{r}, \tag{7}$$

where *add* stands for the additivity of the terms, as they sum up to the molecular E_{xc}^{DFT} and involve inter- and intra-atomic components. The Hartree-Fock exchange is calculated with the sum

$$E_{x,add}^{KS,A} = E_x^{KS,A} + \frac{1}{2}\sum_{B\neq A} E_x^{KS,AB}, \tag{8}$$

where

$$E_x^{KS,AB} = \int_{\Omega_A} d\mathbf{r}_1 \int_{\Omega_B} d\mathbf{r}_2 r_{12}^{-1} \rho_x^{KS}(\mathbf{r}_1, \mathbf{r}_2), \tag{9}$$

that corresponds to the intra-atomic $E_x^{KS,A}$ for $\Omega_B \equiv \Omega_A$ and only requires the set of KS molecular orbitals to construct the exchange density $\rho_x^{KS}(\mathbf{r}_1, \mathbf{r}_2)$.

Once the set of λ_A parameters is available, these are utilised to approximate the intra- $(B = A)$ or inter-atomic $(B \neq A)$ DFT xc energies as follows:

$$\tilde{E}_{xc}^{AB} = \frac{1}{2}\left[\lambda_A + \lambda_B\right] E_{x,add}^{KS,A}. \tag{10}$$

Consequently, the xc energy is split into

$$E_{xc}^{DFT} = \sum_A \tilde{E}_{xc}^A + \sum_{B\neq A} \tilde{E}_{xc}^{AB}. \tag{11}$$

Traditionally, σ-hole driven halogen bonding has been characterised in terms of the positive ESP region that develops along the bonding axis in heavy halogens that usually points toward the lone pair torus of their bonding partners.[55,56] Any real space measure of electron localisation

like the electron localisation function (ELF),[57] or of local charge concentration and depletion, like the Laplacian of the electron density $\nabla^2 \rho$, can be used to explore and detect σ-holes trustfully.[58–60] As the Laplacian is regarded, its ability to uncover the shell structure of low Z atoms makes it a very appealing descriptor. Even for ($Z \geq 20$) atoms, for which the Laplacian does not resolve their valence shells appropriately, the visualisation of $\nabla^2 \rho$ still provides clear pictures of σ-hole behaviour.

3. Computational details

Geometry optimisations and single point calculations were carried out with the Gaussian09 (G09) package[61] at the M06-2X/x2c-TZVPPall level of theory, with Grimme's D3 dispersion correction[62] and an ultrafine grid. The M06-2X density functional was specifically developed and parameterised for a correct description of non-covalent interactions (especially in the case of main group chemical elements)[63] and was also validated for these purposes in several benchmark studies.[64–66] We have chosen this functional according to our previous experience and its successful performance in a number of halogen bond studies in various similar supramolecular systems.[67–71]

In order to avoid problems with relativistic effects, we used the specially developed segmented contracted all-electron relativistic basis sets x2c-TZVPPall and the Ahlrich's family basis sets def2-TZVPP with all-electron spin-free X2C correction for iodine.[72] We have also previous experience on the successful application of such basis sets in studies of non-covalent interactions[73,74] and, in particular, in chemical systems containing heavy atoms (noble metals, iodine).[75,76] Additionally, the RESC scalar relativistic approach,[77] as implemented in G09, was employed for all compounds that include iodine.

All the optimisations converged towards the trihalides $[\mathrm{X_1-Y_2-X_3}]^-$, that were properly characterised through harmonic frequency analysis. To obtain asymmetric geometries $\mathrm{X_1-Y_2 \cdots X_3}^-$ resembling the arrangements found in crystals, a scan over the $\mathrm{Y_2-X_3}$ distance was performed with a frozen $\mathrm{X_1-Y_2}$ separation equal to the experimental diatomic equilibrium one, retrieved from the NIST database.[78] This has proven to be a reliably procedure to examine pure halogen bonding effects without the influence of further species.

Analogously, and in order to test the reliability of the DFT results, a comparison between the energy components of M06-2X and a more reliable CCSD reference was also performed. For these correlation-consistent calculations, PySCF[79] was used to obtain the density matrices

required by the subsequent IQA analyses. The x2c-TZVPPall and def2-TZVPP were used in these calculations for a selected set of $X_1-Y_2\cdots X_3^-$ systems. Also, relativistic effects were accounted for iodine compounds, in this case with the all-electron spin-free X2C correction. This precludes a complete IQA reconstruction of the energy, but does not affect the interatomic interaction components, which constitute our focus.

The wave functions obtained from the previous G09 or PySCF calculations were input to the PROMOLDEN code[80] to carry out the IQA energetic analyses. For these, β-spheres with radii between 0.30 and 0.35 a.u. were employed along with 5810-point Lebedev angular grids and Gauss-Chebychev of 2nd kind radial quadratures with 451 radial points. The expansion of r_{12}^{-1} was performed up to a L_{max} of 10. Non β-sphere grids were performed using the same angular quadratures but 551 points Clenshaw-Curtis radial quadrature was chosen in this case and the L_{max} was raised to 12. PROMOLDEN does not actually recover relativistic effects, that would affect mainly some intra-basin energy contributions like the atomic kinetic energy. Since we are mainly interested in interatomic interaction energies, which remain basically unperturbed after our relativistic calculations, we think that the reported interactions involving iodine atoms are not seriously affected by this approximation.

Finally, the molecular and atomic representations have been obtained with the AIMAll[81] and Jmol[82] codes. The rest of the figures were done with the help of Python's Matplotlib.[83]

4. Results and Discussion

4.1. Systems

As mentioned in the Introduction, we have focused our efforts on pure negatively-charged halogen species that are intended to capture the essence of σ-hole interactions. These systems bind linearly through the extra σ-like lone pair of the halide anion. A total of eighteen of these systems have been considered in this work. These comprise the trihalides as well as the constrained X_1-Y_2 dimers attacked by a X_3 halogen anion in a linear arrangement,[47] where X,Y=Cl, Br, I: $Cl-Cl\cdots Cl^-$, $[Cl-Cl-Cl]^-$, $Cl-Br\cdots Cl^-$, $[Cl-Br-Cl]^-$, $Cl-I\cdots Cl^-$, $[Cl-I-Cl]^-$, $Br-Br\cdots Br^-$, $[Br-Br-Br]^-$, $Br-Cl\cdots Br^-$, $[Br-Cl-Br]^-$, $Br-I\cdots Br^-$, $[Br-I-Br]^-$, $I-Cl\cdots I^-$, $[I-Cl-I]^-$, $I-Br\cdots I^-$, $[I-Br-I]^-$, $I-I\cdots I^-$, $[I-I-I]^-$. We leave aside fluorine, since this element is normally not considered to get involved in halogen bonds as a result from its low polarisability. All optimised and constrained geometries and energies can be found in the Supporting Information (SI), Table S1.

8

4.2. Atomic charges and interatomic interaction energies

Sigma holes are expected to enhance halogen interactions by means of favourable electrostatics between the halogen that develops the σ-hole and the attacking, electron-rich moiety. However useful any argument based on the docking of ESP isosurfaces may be, it is always the total electrostatic interaction between the approaching entities that has to be taken into account. This is something that IQA provides directly. As we are going to show, a lock-and-key ESP isosurface match may actually hide a globally destabilising electrostatic contribution. The IQA interaction energy E_{int}^{AB}, and its classical (Coulomb, E_{class}) and non-classical (exchange-correlation, E_{xc}) components give directly the grand total interactions, which can be perceived as what one would obtain if we brought together all the ESP pictures at all possible different isovalues, reconstructing an onion from its peels. Another way to state this idea comes from recognising that a region characterised by a positive ESP value on the van der Waals envelope surface may well display a negative ESP value on another surface, so that the global electrostatic is by no means obvious from the exam of the first surface.

Table 1 contains the total interaction energies, along with the classical and exchange-correlation contributions, for every pair of atoms present in the complexes. The QTAIM atomic charges are also depicted. Although the data present in Table 1 correspond to the M06-2X density functional, we have validated the reliability of our results by performing reference CCSD calculations on a selected number of systems: $Cl-Cl\cdots Cl^-$, $Cl-Br\cdots Cl^-$, $Cl-I\cdots Cl^-$ and $Br-Br\cdots Br^-$, with both the x2c-TZVPPall and the def2-TZVPP basis sets. The data can be found in the S2 and S3 Tables of the SI. Small differences in charges and larger energetic discrepancies can be found on comparing the DFT and CCSD data but, overall, the M06-2X/x2c-TZVPPall level of theory so far used is found to be accurate enough. All trends found in Table 1 are maintained. For comparison purposes also, M06-2X results in the optimised *in vacuo* X_1-Y_2 diatomics can also be found in Table 2.

Table 1: Atomic charges (a.u.) as well as delocalisation indices (DI) and interaction energies (in kcal mol^{-1}) for each atomic pair in the complexes $X_1 - Y_2 - X_3$ at the M06-2X/x2c-TZVPPall level. Negative charges are coloured in red, positive in blue and those close to zero in green.

Complex	Pair	DI	E_{int}	E_{xc}	E_{class}
$Cl_1{}^{-0.30} - Cl_2{}^{+0.04} \cdots Cl_3{}^{-0.74}$	1-2	1.278	-161.439	-182.490	21.051
	2-3	0.542	-66.779	-60.613	-6.166
	1-3	0.149	6.863	-5.443	12.306
$[Cl_1{}^{-0.50} - Cl_2{}^{+0.01} - Cl_3{}^{-0.50}]$	1-2	0.872	-104.724	-107.164	2.440
	2-3	0.872	-104.502	-107.169	2.667
	1-3	0.211	7.861	-7.646	15.507
$Cl_1{}^{-0.46} - Br_2{}^{+0.16} \cdots Cl_3{}^{-0.70}$	1-2	1.169	-159.842	-156.455	-3.387
	2-3	0.674	-94.116	-76.965	-17.151
	1-3	0.152	15.019	-5.443	20.462
$[Cl_1{}^{-0.56} - Br_2{}^{+0.12} - Cl_3{}^{-0.56}]$	1-2	0.877	-113.753	-104.485	-9.268
	2-3	0.877	-113.916	-104.551	-9.366
	1-3	0.179	13.500	-6.234	19.734
$Cl_1{}^{-0.61} - I_2{}^{+0.34} \cdots Cl_3{}^{-0.73}$	1-2	1.061	-174.495	-130.451	-44.044
	2-3	0.696	-115.651	-75.887	-39.764
	1-3	0.118	25.321	-4.003	29.324
$[Cl_1{}^{-0.65} - I_2{}^{+0.30} - Cl_3{}^{-0.65}]$	1-2	0.836	-128.549	-93.217	-35.332
	2-3	0.835	-128.068	-93.114	-34.955
	1-3	0.129	22.938	-4.229	27.167
$Br_1{}^{-0.22} - Cl_2{}^{-0.10} \cdots Br_3{}^{-0.69}$	1-2	1.243	-142.250	-165.847	23.597
	2-3	0.590	-58.304	-64.176	5.872
	1-3	0.192	-0.491	-6.571	6.080
$[Br_1{}^{-0.45} - Cl_2{}^{-0.10} - Br_3{}^{-0.45}]$	1-2	0.883	-93.221	-103.455	10.234
	2-3	0.881	-93.034	-103.235	10.201
	1-3	0.251	2.311	-8.476	10.787
$Br_1{}^{-0.39} - Br_2{}^{+0.03} \cdots Br_3{}^{-0.64}$	1-2	1.164	-132.598	-147.432	14.835
	2-3	0.729	-83.199	-81.060	-2.139
	1-3	0.191	7.152	-6.407	13.559

$[\text{Br}_1{}^{-0.50}-\text{Br}_2{}^{+0.01}-\text{Br}_3{}^{-0.50}]$	1-2	0.883	-98.024	-102.562	4.538
	2-3	0.881	-98.239	-102.832	4.593
	1-3	0.251	7.136	-7.004	14.141
$\text{Br}_1{}^{-0.52}-\text{I}_2{}^{+0.21}\cdots\text{Br}_3{}^{-0.69}$	1-2	1.099	-139.381	-127.528	-11.854
	2-3	0.698	-92.092	-71.456	-20.636
	1-3	0.147	16.688	-4.536	21.224
$[\text{Br}_1{}^{-0.59}-\text{I}_2{}^{+0.18}-\text{Br}_3{}^{-0.59}]$	1-2	0.857	-104.986	-90.242	-14.744
	2-3	0.857	-105.079	-90.311	-14.768
	1-3	0.161	15.456	-4.857	20.312
$\text{I}_1{}^{-0.06}-\text{Cl}_2{}^{-0.28}\cdots\text{I}_3{}^{-0.66}$	1-2	1.208	-141.696	-147.381	5.685
	2-3	0.542	-37.715	-53.691	15.975
	1-3	0.24	-9.262	-7.409	-1.853
$[\text{I}_1{}^{-0.37}-\text{Cl}_2{}^{-0.26}-\text{I}_3{}^{-0.37}]$	1-2	0.853	-80.019	-92.052	12.033
	2-3	0.851	-79.783	-91.862	12.080
	1-3	0.317	-3.618	-9.734	6.116
$\text{I}_1{}^{-0.24}-\text{Br}_2{}^{-0.15}\cdots\text{I}_3{}^{-0.61}$	1-2	1.168	-117.898	-135.958	18.060
	2-3	0.661	-58.258	-66.461	8.204
	1-3	0.237	-1.622	-7.156	5.533
$[\text{I}_1{}^{-0.42}-\text{Br}_2{}^{-0.16}-\text{I}_3{}^{-0.42}]$	1-2	0.872	-79.271	-91.298	12.027
	2-3	0.873	-79.488	-91.397	11.909
	1-3	0.274	0.301	-8.108	8.409
$\text{I}_1{}^{-0.41}-\text{I}_2{}^{+0.04}\cdots\text{I}_3{}^{-0.63}$	1-2	1.129	-112.387	-123.099	10.712
	2-3	0.719	-73.745	-69.966	-3.779
	1-3	0.192	7.315	-5.485	12.801
$[\text{I}_1{}^{-0.51}-\text{I}_2{}^{+0.02}-\text{I}_3{}^{-0.51}]$	1-2	0.879	-84.009	-87.520	3.512
	2-3	0.879	-84.689	-87.474	2.785
	1-3	0.209	7.356	-5.846	13.202

Table 2: Atomic charges (a.u.), delocalisation indices (DI), exchange-correlation, classical and total interaction energies (kcal mol^{-1}) for the *in vacuo* diatomic molecules $X_1 - Y_2$. Negative charges are coloured in red, positive in blue and those close to zero in green.

Diatomic molecule	DI	E_{int}	E_{xc}	E_{class}
$Cl^{0.00} - Cl^{0.00}$	1.440	-170.066	-199.990	29.923
$Cl^{-0.14} - Br^{+0.14}$	1.407	-164.855	-181.860	17.006
$Cl^{-0.33} - I^{+0.33}$	1.323	-178.780	-155.429	-23.351
$Br^{0.00} - Br^{0.00}$	1.415	-148.233	-172.478	24.245
$Br^{-0.20} - I^{+0.20}$	1.367	-146.100	-151.289	5.188
$I^{0.00} - I^{0.00}$	1.395	-125.317	-144.553	19.236

A most remarkable feature found in both the trihalides and the halogen-bonded moieties is the considerable value of the charge difference in the edge atoms $\Delta Q(\text{X3}) \sim -\Delta Q(\text{X1})$, measured from the isolated $X_1 - Y_2$ reference. This can be naïvely interpreted as a direct charge transfer from the attacking halide X_3 to the farthermost halogen atom X_1. In the symmetric trihalide systems, the central Y_2 atom is almost neutral if X=Y, and carries a positive or negative charge, consistent with the difference of electronegativity between X and Y, otherwise. In any case, the anionic character of all the molecules is heavily carried by the external X atoms. Interestingly, this large density polarisations are also found in the asymmetric $X_1 - Y_2 \cdots X_3^-$ systems that we use to model halogen bonds, albeit the attacking X_3^- now retains a larger part of the anionic negative charge. Despite the distance between Y_2 and X_3, that we will call d_{23}, is longer than d_{12} in the $X_1 - Y_2 \cdots X_3^-$ systems mimicking halogen bonding (see Table S1), the central charge is pretty similar for Y_2 in both $[X_1 - Y_2 - X_3]^-$ and $X_1 - Y_2 \cdots X_3^-$. More interestingly, this central charge resembles to a large extent that found in the corresponding isolated diatomic molecule. This shows very neatly that this edge to edge polarisation is rather general and that the central atom plays an spectator role as far as electrostatics is concerned. We will consider more deeply this issue in the next section.

Contrary to the common assumption that the covalent or exchange-correlation contribution is not very relevant in comparison with its electrostatic counterpart in σ-hole interactions, IQA shows that such an energy component is of great importance in all the cases studied. This is in line with previous assessment of the role of both type of interactions in other σ-hole instances.[32,36–44] A close inspection of the pair interaction energies and their components reveals that the 2-3 pairs are many times destabilised by the electrostatic term and that, even when those energies are stabilising, the non-classical ones always dominate. The closer the third X_3 atom is to Y_2, the more attractive the $X_3 - Y_2$ interactions become, but at the expense of increasing the repulsive character of the Coulomb interactions while the xc term becomes even more dominant. It shall be seen in following sections that the grouped attractive interaction of fragment 1-2 with atom 3 is explained only by the E_{xc} component in almost all the cases studied, whilst the group E_{class} terms are destabilising.

Paying attention to the relation between the electrostatic interaction and the monopole $Q^A Q^B / R_{AB}$ contribution, our results show that, in general, oppositely charged basins lead to stabilising electrostatics, although multipolar terms often modify this initial assumption. For example, the first two chlorine atoms of $Cl - Cl \cdots Cl^-$ present an E_{class} of $\sim +21$ kcal mol^{-1}

although the involved atoms carry opposite-sign net charges. Multipolar contributions are only dominant when one of the two species displays small net charges.

It is also relevant to consider how the X_1-Y_2 pair becomes modified when passing from the isolated diatomic to the $X_1-Y_2\cdots X_3{}^-$ systems, since the distance between the first two atoms remains fixed in this process. It is clear that, in general, $E_{int}(X_1, Y_2)$ becomes less favourable as the X_3 atom approaches. The smallest destabilisations in E_{int} are found for systems with X=Cl and Y=I, ranging from 4.3 ($Cl-I\cdots Cl^-$) to 12.9 kcal mol^{-1} ($I-I\cdots I^-$); the largest ones in the opposite X=I and Y=Cl situations, the I-Cl interaction in $I-Cl\cdots I^-$ being that which is mostly destabilised (by 31.1 kcal mol^{-1}). Concerning its components, $|E_{xc}(X_1Y_2)|$ decreases upon interaction, an easy to interpret result since the central Y_2 atom now shares its valence electrons among two, and not one species. In most cases, the xc destabilisation is larger than that found in the total E_{int}. On the other hand, $E_{class}(X_1, Y_2)$ shows a richer behaviour. In some cases the electrostatic interaction weakens, but in others it strengthens substantially. This is in accordance with the charge transfer that the diatomic molecule undergoes upon complex formation. For instance, in Cl-Br, $Q(Cl) = -0.14e$. When moving to the complexes, in $Cl-Br\cdots Cl^-$ bromine maintains roughly the same charge, while chlorine gains about 0.46 e and, accordingly, their mutual classical energy stabilises by about 20 kcal mol^{-1}. Contrarily, in $Br-Cl\cdots Br^-$, bromine bears a charge of -0.45 e, but, at the same time, chlorine gains 0.10 extra electrons, giving rise to a electrostatic destabilisation of 6.6 kcal mol^{-1}. Similar trends are found for the rest of the $X_1-Y_2\cdots X_3{}^-$ systems.

As d_{23} is considerably smaller in the trihalides, their associated E_{xc} (and so E_{int}) values are rather larger than those present in the asymmetric complexes. The most stabilising E_{class} is found in $Cl-I\cdots Cl^-$, where the more compact and polarisant Cl^- ion interacts with the diffuse, polarisable I atom. The same behaviour is found in Br, although in a less pronounced fashion. Since the interaction of Y with both X is symmetric, and so are distances d_{12} and d_{23}, the X_1-Y_2 bond is highly weakened, with an E_{int} destabilisation ranging from 41.1 to 98.8 kcal mol^{-1}. The major source of destabilisation is again E_{xc}, in accordance with the loss in electron sharing (57.0-92.8 kcal mol^{-1}), while E_{class} turns out again to be the most interesting contribution to the 1-2 pair interaction energy. For all systems but $[I-Cl-I]^-$ and $[I-Br-I]^-$ the electrostatic interaction energy of pairs 1-2 stabilises from 6.7 (in $[Br-Cl-Br]^-$) to 27.5 kcal mol^{-1} (in $[Cl-Cl-Cl]^-$). $[I-Cl-I]^-$ and $[I-Br-I]^-$, by contrast, show an E_{class} destabilisation of 6.8 and 35.4 kcal mol^{-1} respectively. This is as a result of the large multipolar distortions

suffered by the iodine atoms.

All the previous findings concerning the charge distribution and the high relative importance of the xc energy in the systems here studied are in line with the Pimentel-Rundle model,[84,85]. In real space, the effects of the 3c-4e bond translates into anomalously high delocalisation indices between the edge atoms,[86] and it also predicts net charges equal to $-0.5\,e$ for the external atoms. All this is in very good agreement with what we have obtained in the $[X_1-Y_2-X_3]^-$ species as shown in Fig. 1. More interestingly, these insights remain valid with slight modifications in the $X_1-Y_2\cdots X_3^-$ complexes. The incipient formation of a 3c-4e-like interactions in halogen bonds is thus a possibility that should be further studied with standard orbital techniques. For instance, Oliveira et al. used relative stretching force constants in this regard.[87] The presence of such a bond type justifies the real space image and strengthens the importance of quantum mechanical contributions on top of the dominant electrostatic point of view.

4.3. Charge transfer along $Y_2-X_3^-$ stretching

One of the most striking features highlighted in the previous section is the large charge transfer following the approach of the X_3^- anion. It leads to the farthermost X_1 atom concentrating the charge excess while the central atom Y_2 maintaining a quite similar charge to that it carries in the isolated diatomic molecule. In order to assess whether this observation is the result of a concerted charge transfer mechanism in which the central atom is not directly affected, or one with several steps in which CT and polarisation are decoupled, several scans over the Y_2-X_3 distance were performed for the systems $Cl-Cl\cdots Cl^-$, $Cl-Br\cdots Cl^-$, $Cl-I\cdots Cl^-$ and $Br-Br\cdots Br^-$. Figure 2 depicts the evolution of the atomic charge for each atom in the $Cl-Br\cdots Cl^-$ complex (the rest of the selected systems and the CCSD reference calculations can be found in Table S4).

In the distance range sampled, it is rather clear that the charge of the central atom does not change appreciably while, on the contrary, the external atoms display mirror evolutions. Similar results are found for the rest of the systems considered. This experiment demonstrates that the X_3^- to X_1-Y_2 charge transfer is in this region a one-step process, which cannot be easily explained without the assistance of orbital interactions. Whether this is a particular feature of halide attacks or a general property of halogen bonds remains a very important point to be investigated. If these results are found to be general, the full halogen bonding electrostatic model will have to be reconsidered. It is interesting to note that the evolution of $Q(X_3)$ with

15

Figure 2: Atomic charges (in a.u.) as a function of the $Br \cdots Cl$ distance.
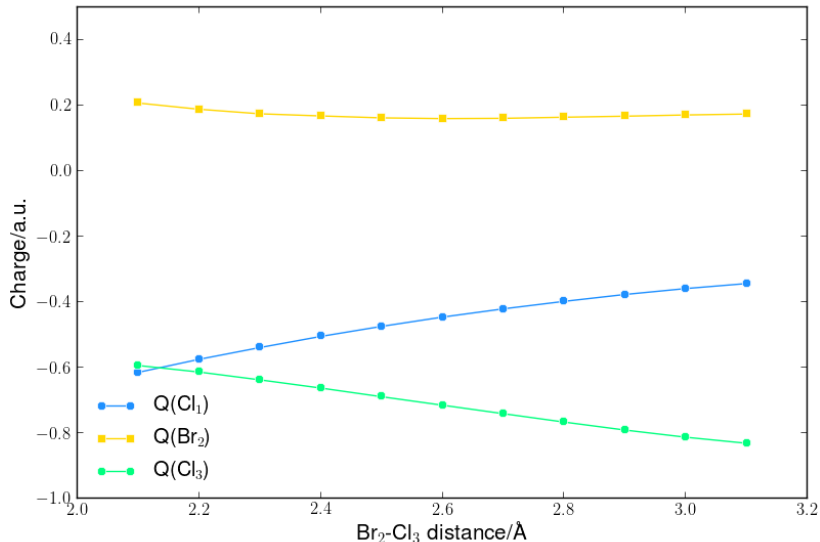
distance is convex in the case of Y=Cl but that it changes to concave when more bulky atoms become involved. Thus, for Y=Br the trend is quasi-linear (such as the case depicted in Fig. 2), and clearly concave between external Cl and central I (also more charged terminal Cl). Such a behaviour reveals the rate at which the charge transfer takes place, that results more rapid when more accessible (i.e., compact) atoms occupy the central position.

4.4. IQA formation energies

The formation energies (E_{form}) that provide the stabilisation energy of the trihalide or asymmetric complexes from the isolated diatomics and halide ions can also be partitioned within the IQA approach. E_{form} thus becomes a sum of the deformation energy of the fragments E_{def} —the energetic cost needed to prepare the fragment from its *in vacuum* energy to the geometric and electronic state found within the complex— and the interaction energy between the two fragments. Additionally, an IQA equivalent of Pauli repulsion in weakly interacting systems can be built by adding the fragments total deformation to their mutual exchange-correlation energy. This is usually known as the exchange-correlation repulsion term, XCR = $E_{def}(X_1 - Y_2) + E_{def}(X_3) + E_{xc}(X_1 - Y_2, X_3)$. This IQA fragment partition can be found in Table 3.

The set of formation energies provided by IQA (E_{form}^{IQA}) are very close to the ones calculated from the G09 package (E_{form} in Table S1), although subject to a certain numerical error. Notice

16

Table 3: IQA formation energies obtained from the balance between the deformation of fragments X_1-Y_2 and X_3, and their mutual interaction, together with their exchange-correlation repulsion terms XCR. All data in kcal mol^{-1}.

Complex	E_{form}^{IQA}	$E_{def}(X_1-Y_2/X_3)$	E_{int}	E_{xc}	E_{class}	XCR
Cl−Cl···Cl$^-$	-21.608	14.178 / 24.131	-59.917	-66.057	6.140	-27.748
[Cl−Cl−Cl]$^-$	-28.764	14.287 / 53.591	-96.641	-114.815	18.174	-46.937
Cl−Br···Cl$^-$	-32.643	14.100 / 32.354	-79.097	-82.408	3.311	-35.954
[Cl−Br−Cl]$^-$	-39.277	11.384 / 49.755	-100.416	-110.785	10.369	-49.646
Cl−I···Cl$^-$	-35.470	21.592 / 33.268	-90.330	-79.889	-10.440	-25.030
[Cl−I−Cl]$^-$	-40.276	21.018 / 43.837	-105.130	-97.342	-7.788	-32.488
Br−Cl···Br$^-$	-22.142	10.669 / 25.985	-58.795	-70.747	11.952	-34.094
[Br−Cl−Br]$^-$	-29.010	6.754 / 54.959	-90.723	-111.711	20.988	-46.998
Br−Br···Br$^-$	-32.530	9.778 / 33.739	-76.047	-87.467	11.421	-43.951
[Br−Br−Br]$^-$	-36.768	4.658 / 49.676	-91.102	-109.836	18.734	-55.502
Br−I···Br$^-$	-29.772	14.066 / 31.566	-75.405	-75.992	0.588	-30.360
[Br−I−Br]$^-$	-35.384	10.219 / 44.020	-89.623	-95.168	5.545	-40.929
I−Cl···I$^-$	-16.528	3.934 / 26.530	-46.993	-61.051	14.058	-30.587
[I−Cl−I]$^-$	-21.562	-0.093 / 61.931	-83.401	-101.596	18.195	-39.758
I−Br···I$^-$	-23.592	4.283 / 32.005	-59.880	-73.617	13.737	-37.329
[I−Br−I]$^-$	-26.751	-2.072 / 54.507	-79.187	-99.505	20.318	-47.069
I−I···I$^-$	-27.159	7.186 / 32.084	-66.430	-75.452	9.022	-36.181
[I−I−I]$^-$	-29.709	1.892 / 45.732	-77.333	-93.320	15.988	-45.696

that our IQA deformations do not include properly relativistic effects, although this does not affect the E_{int} values reported.

The E_{def} values of the attacking anions are considerably bigger than those of the diatomic fragment. This can be explained by the larger number of relaxation channels (or degrees of freedom) of the latter. Interestingly, the E_{def}'s of $[I-Cl-I]^-$ and $[I-Br-I]^-$ systems are negative, a fact related with the energy stabilisation of any fragment that gains electrons. After all, halogens have high electron affinities. Notice how the deformation of the X_1-Y_2 group decreases as we go down the table. Iodine containing systems display very small E_{def}'s. As the halide

approaches to the halogen diatomic, two energetically opposed factors balance the deformation in the $X_1 - Y_2$ system: an increasing electronic cost and an also increasing electron affinity-driven charge transfer stabilisation. This two factors operate in unison in the halide deformation, since it gets depopulated as well as electronically deformed.
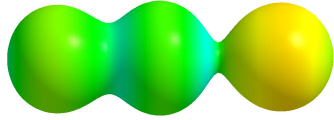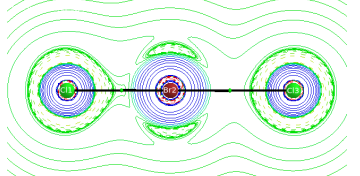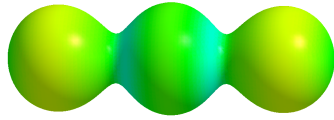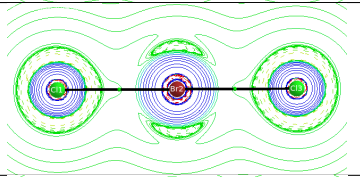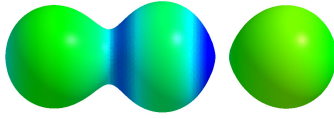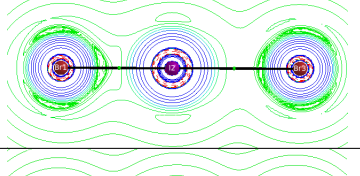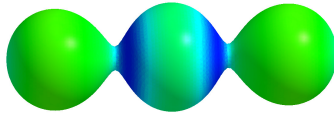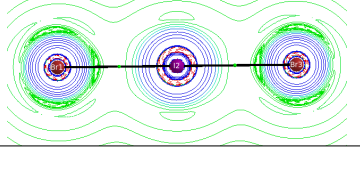
Turning to the electrostatic part of the interaction between the two fragments, E_{class} is seen to correlate with E_{form}. Indeed, the asymmetric complexes $X_1 - Y_2 \cdots X_3^-$, that experience smaller deformations and lower classical interaction energies, present slightly destabilising or even stabilising E_{class} values depending on the polarisant-polarisable character of the central and σ-hole featuring atom and the attacking halide. Such behaviour is generally consistent with lower formation energies, and points out the relevance of electrostatics in describing these XB systems. In fact, as pointed out by Clark et al.,[88,89] the ESP calculated at specific points is shown to correlate well with E_{form}. Nevertheless, as can be appreciated from Table 3, the complex formation cannot be explained in terms of electrostatics only; by contrast, it is the non-classical E_{xc} the component that accounts for most of the total E_{int} as it is in all cases negative and represents a stronger interaction between fragments $X_1 - Y_2$ and X_3^-.

These findings contrast with those obtained with EDA by Wolters and Bickelhaupt,[48] who reported stabilising electrostatic energies for all the trihalides here considered. This is what is normally expected from the use of interpenetrating electron densities. In EDA or SAPT, when the density of a fragment spreads over a region close to the nuclei of the other fragment, a large stabilising contribution appears which is absent in exhaustive spatial partitionings. In our opinion, large EDA electrostatic stabilisations simply signal the inadequacy of the interpenetrating model, for an electron close to the nucleus of another fragment should be associated to the latter and not to the former. Otherwise, the trends in the evolution of the different EDA contributions run roughly parallel to our IQA results.

The last energy descriptor reported here accounts for the Pauli repulsion at long range regime.[90] The fact that XCR is negative in all the cases explored is a very strong indicator of the non-perturbative regime of these interactions and of the fundamental role played by orbital interactions, in standard molecular orbital parlance, or covalency or exchange-correlation effects in real space language. Negative XCRs imply that exchange-correlation overcomes deformation, signalling the importance of electron delocalisation overcoming Pauli repulsion. We observe, once and again, that these simple XB models cannot be understood fully in terms of electrostatics.

4.5. Other topological descriptors: The Laplacian of ρ and the electrostatic potential

Table 4: Electrostatic potential on the 0.05 a.u. isosurface of ρ (left) and Laplacian of the electron density on a molecular plane (right) for $Cl-Br\cdots Cl$, $[Cl-Br-Cl]^-$, $[Br-I\cdots Br-]$ and $[Br-I-Br]^-$. A red to blue palette has been used ranging from -0.22 to +0.22 a.u. for the ESP and from 1.00 to +1.00 a.u. in the case of the Laplacian.

System	ESP on $\rho = 0.05$	$\nabla^2\rho$ on a plane
$\mathbf{Cl-Br\cdots Cl^-}$		
$\mathbf{[Cl-Br-Cl]^-}$		
$\mathbf{Br-I\cdots Br^-}$		
$\mathbf{[Br-I-Br]^-}$		

In order to complete the above quantitative energetic analysis, we now turn to the usual qualitative insights that can be obtained from inspecting both the electrostatic potential (ESP) and the Laplacian of the electron density, $\nabla^2\rho(\boldsymbol{r})$. The complete set of representations can be found inn the SI (Tables S5-S9). Here, we show two representative $X_1-Y_2-X_3^-$ systems: $Cl-Br-Cl^-$ and $Br-I-Br^-$. Table 4 presents the ESP mapped onto a $\rho(\boldsymbol{r}) = 0.05$ a.u. isosurface, along with 2D Laplacian maps using solid lines for positive values and dashed lines for negative ones.

σ-holes are clearly present as blueish ESP regions along the bonding direction. A very clear evolution from the more polarising Cl atom to the more polarisable I moiety is clearly seen. The Laplacian counts the correct number of shells only for chlorine, but can be clearly used to unveil the σ-holes in all cases: unprotected, depletion regions, together with torus-like lone pair accumulation regions, are notorious along the molecular axis in the central atoms. Tables

S5-S9 show that the iodine lone pairs can even be discerned as the red, less positive regions of $\nabla^2 \rho$ when of the anions and the diatomic species are depicted, disclose the characteristic clear directionality in the σ-holes. For instance, in the case of the compact chlorine atoms, a flattening of the external negative Laplacian regions is visible when close to another polarising entity. Conversely, when chlorine occupies the external positions (i.e., X=Cl) the outermost concentration distorts towards the central atom (e.g., $Cl-Br\cdots Cl^-$).

We will not proceed further. Using the ESP and $\nabla^2 \rho(\boldsymbol{r})$ (or the ELF) has become mainstream and needs no more consideration. Both can be used to sense and detect σ-holes. However we should not forget that the information they provide is qualitative in nature. For example, (i) the most positive ESP is found for I atoms, independently of whether they occupy the central position or not, and (ii) the Laplacian detects different shells for different atoms, so that quantification becomes very difficult with it. Moreover, we insist that standard ESP representations do not account for the whole charge distribution and that their bare use might be misleading. As an example, the most intense σ-holes as sensed from the ESP are actually those of iodine that give rise to the lowest interatomic E_{class} values (see Table 1). However, when I occupies the edge position (Table S7) and thus it becomes the attacking halide for complexes $X_1-Y_2\cdots X_3$, its electrostatic energies with central and more electronegative Cl and Br atoms are highly destabilising. This is another example of the incomplete picture that the ESP provides.

5. Conclusions

We examine in this work a minimal model of halogen bond (XB) interactions in an attempt to understand their nature beyond the σ-holes electrostatic picture under an orbital-invariant perspective. To that end, we have investigated a set of negatively charged halogen triatomic species built with different combinations of Cl, Br and I. Given the controversial character of halogen bonding with F containing species, fluorine has not been considered. For all possible Cl, Br, I combinations, two linear geometries have been selected: the trihalide optimised one together with another that mimics an XB. The latter is obtained by constraining the interatomic distance of two of the atoms at their experimental gas-phase value while a halide anion attacks the constrained diatomic. We have used the interacting quantum atoms (IQA) approach to dissect the energetics of the systems, complementing it with other traditional descriptors like electrostatic potential (ESP) or electron density Laplacian maps.

Both the ESP and the Laplacian analyses reveal the traditional signature of σ-holes: positive regions of the ESP when mapped onto an isodensity or van der Waals surface, and charge depletion areas of the Laplacian field. A neat distinction between different atoms also emerges. Chlorine preserves its atomic shell structure to a large extent, while bromine and iodine are largely deformed according to their more polarisable nature. Similarly, the more polarisant-polarisable difference in the atomic pair that interacts, the clearer the σ-hole develops.

The QTAIM's charges assigned to each atom within the anionic species unveil a profound density polarisation and concomitant charge transfer, that in light of a $Y_2 - X_3$ distance scan takes place gradually with participation of just the charge donor and its farthermost partner, while the central atom remains a spectator. These findings support the assumption that these systems present 3c-4e bonds, which spread the anionic charge over the $X_1 X_2$ terminal atoms, that present anomalously large delocalisation indexes.

The electrostatic σ-hole model is shown to be useful as a qualitative tool but far from the real energetics of the complexes. Indeed, the presence of such a positive region in the ESP does not necessarily imply global favourable electrostatics between the interacting species, that account for the whole charge distribution, and such a simple interpretation may become misleading. Thus, we have shown that in all but those cases where the XB involves a very polarisant-polarisable interacting pair the full electrostatic interaction between the halide and the remaining diatomic is destabilising. This contrasts other energy decomposition analyses

characterised by interpenetrating densities, which tend to provide stabilising electrostatic contributions due to the non-physical overlap of the densities of the two fragments.

Taking into account the non-negligible deformation energies suffered by both entities, it is the covalent (exchange-correlation) interaction that justifies the overall stability of the complexes. This is in line with what prior studies had suggested in other systems of interest. By reducing the complexity of the systems here studied to an absolute minimum, we believe to have shown convincingly that electrostatics alone is not enough to understand the stability of halogen bonded systems. Together with the evolution of charge transfer, the essential role of electron delocalisation, i.e., of orbital interactions, is thus highlighted.

6. Acknowledgements

7. Data availability statement

The data that support the findings of the present study are available in both the article and its supplementary material. Additional information can be requested from the authors.

8. ORCID

Fernando Jiménez-Grávalos https://orcid.org/0000-0002-9181-5234

Miguel Gallegos https://orcid.org/0000-0001-7472-8158

Ángel Martín Pendás https://orcid.org/0000-0002-4471-4000

Alexander S. Novikov https://orcid.org/0000-0001-9913-5324

References

[1] B. S. Sumerlin, *Science* **2018**, *362*, 150–151.

[2] P. Hobza, J. Řezáč, *Chemical Reviews* **2016**, *116*, 4911–4912.

[3] C. F. J. Faul, *Accounts of Chemical Research* **2014**, *47*, 3428–3438.

[4] D. Y. Lee, Z. Khatun, J.-H. Lee, Y. kyu Lee, I. In, *Biomacromolecules* **2011**, *12*, 336–341.

[5] J. J. Strunk, I. Gregor, Y. Becker, P. Lamken, S. Lata, A. Reichel, J. Enderlein, J. Piehler, *Bioconjugate Chemistry* **2009**, *20*, 41–46.

[6] J. Hermann, R. A. DiStasio, A. Tkatchenko, *Chemical Reviews* **2017**, *117*, 4714–4758.

[7] B. Rybtchinski, *ACS Nano* **2011**, *5*, 6791–6818.

[8] W. Zhan, T. Wei, Q. Yu, H. Chen, *ACS Applied Materials & Interfaces* **2018**, *10*, 36585–36601.

[9] C. Sutton, C. Risko, J.-L. Brédas, *Chemistry of Materials* **2015**, *28*, 3–16.

[10] J. S. Pallesen, K. T. Tran, A. Bach, *Journal of Medicinal Chemistry* **2018**, *61*, 8088–8103.

[11] A. Minecka, E. Kaminska, M. Tarnacka, I. Grudzka-Flak, M. Bartoszek, K. Wolnica, M. Dulski, K. Kaminski, M. Paluch, *Molecular Pharmaceutics* **2018**, *15*, 4764–4776.

[12] J.-P. Correa-Baena, M. Saliba, T. Buonassisi, M. Grätzel, A. Abate, W. Tress, A. Hagfeldt, *Science* **2017**, *358*, 739–744.

[13] A. J. Neel, M. J. Hilton, M. S. Sigman, F. D. Toste, *Nature* **2017**, *543*, 637–646.

[14] F. D. Toste, M. S. Sigman, S. J. Miller, *Accounts of Chemical Research* **2017**, *50*, 609–615.

[15] S. E. Wheeler, T. J. Seguin, Y. Guan, A. C. Doney, *Accounts of Chemical Research* **2016**, *49*, 1061–1069.

[16] S. A. Kholodar, A. K. Ghosh, K. Świderek, V. Moliner, A. Kohen, *Proceedings of the National Academy of Sciences* **2018**, *115*, 10311–10314.

[17] M. Yáñez, P. Sanz, O. Mó, I. Alkorta, J. Elguero, *Journal of Chemical Theory and Computation* **2009**, *5*, 2763–2771.

[18] J. S. Murray, P. Lane, P. Politzer, *Journal of Molecular Modeling* **2008**, *15*, 723–729.

[19] J. S. Murray, P. Lane, P. Politzer, *International Journal of Quantum Chemistry* **2007**, *107*, 2286–2292.

[20] J. S. Murray, P. Lane, T. Clark, P. Politzer, *Journal of Molecular Modeling* **2007**, *13*, 1033–1038.

[21] J. Y. Lim, P. D. Beer, *Chem* **2018**, *4*, 731–783.

[22] M. H. Kolář, P. Hobza, *Chemical Reviews* **2016**, *116*, 5155–5187.

[23] H. Wang, W. Wang, W. J. Jin, *Chemical Reviews* **2016**, *116*, 5072–5104.

[24] L. C. Gilday, S. W. Robinson, T. A. Barendt, M. J. Langton, B. R. Mullaney, P. D. Beer, *Chemical Reviews* **2015**, *115*, 7118–7195.

[25] G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati, G. Terraneo, *Chemical Reviews* **2016**, *116*, 2478–2601.

[26] P. Politzer, J. S. Murray, T. Clark, *Physical Chemistry Chemical Physics* **2013**, *15*, 11178.

[27] P. Politzer, J. S. Murray, *ChemPhysChem* **2020**, *21*, 579–588.

[28] L. P. Wolters, P. Schyman, M. J. Pavan, W. L. Jorgensen, F. M. Bickelhaupt, S. Kozuch, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*, 523–540.

[29] P. J. Costa, *Physical Sciences Reviews* **2017**, *2*.

[30] L. P. Wolters, N. W. G. Smits, C. F. Guerra, *Physical Chemistry Chemical Physics* **2015**, *17*, 1585–1592.

[31] M. Novák, C. Foroutan-Nejad, R. Marek, *Physical Chemistry Chemical Physics* **2015**, *17*, 6440–6450.

[32] P. L. Bora, M. Novák, J. Novotný, C. Foroutan-Nejad, R. Marek, *Chemistry - A European Journal* **2017**, *23*, 7315–7323.

[33] J. Thirman, E. Engelage, S. M. Huber, M. Head-Gordon, *Physical Chemistry Chemical Physics* **2018**, *20*, 905–915.

[34] M. A. Blanco, Á. Martín Pendás, E. Francisco, *Journal of Chemical Theory and Computation* **2005**, *1*, 1096–1109.

[35] E. Francisco, Á. Martín Pendás, M. Á. Blanco, *Journal of Chemical Theory and Computation* **2005**, *2*, 90–102.

[36] O. A. Syzgantseva, V. Tognetti, L. Joubert, *The Journal of Physical Chemistry A* **2013**, *117*, 8969–8980.

[37] M. Yahia-Ouahmed, V. Tognetti, L. Joubert, *Computational and Theoretical Chemistry* **2015**, *1053*, 254–262.

[38] M. Yahia-Ouahmed, V. Tognetti, L. Joubert, *Theoretical Chemistry Accounts* **2016**, *135*.

[39] K. Eskandari, M. Lesani, *Chemistry - A European Journal* **2015**, *21*, 4739–4746.

[40] H. J. Jahromi, K. Eskandari, A. Alizadeh, *Journal of Molecular Modeling* **2015**, *21*.

[41] G. Buralli, A. Petelski, N. Peruchena, G. Sosa, D. Duarte, *Molecules* **2017**, *22*, 2034.

[42] J. M. Guevara-Vela, D. Ochoa-Resendiz, A. Costales, R. Hernández-Lamoneda, Á. Martín Pendás, *ChemPhysChem* **2018**, *19*, 2512–2517.

[43] N. Orangi, K. Eskandari, J. C. R. Thacker, P. L. A. Popelier, *ChemPhysChem* **2019**, *20*, 1922–1930.

[44] I. Alkorta, A. F. Silva, P. L. A. Popelier, *Molecules* **2020**, *25*, 2674.

[45] T. Sakurai, M. Sundaralingam, G. A. Jeffrey, *Acta Crystallographica* **1963**, *16*, 354–363.

[46] G. R. Desiraju, R. Parthasarathy, *Journal of the American Chemical Society* **1989**, *111*, 8725–8726.

[47] T. Bui, S. Dahaoui, C. Lecomte, G. Desiraju, E. Espinosa, *Angewandte Chemie International Edition* **2009**, *48*, 3838–3841.

[48] L. P. Wolters, F. M. Bickelhaupt, *ChemistryOpen* **2012**, *1*, 96–105.

[49] C. Foroutan-Nejad, Z. Badri, R. Marek, *Physical Chemistry Chemical Physics* **2015**, *17*, 30670–30679.

[50] C. Foroutan-Nejad, *Angewandte Chemie International Edition* **2020**, *59*, 20900–20903.

[51] P. L. A. Popelier in *Intermolecular Forces and Clusters I*, Springer Berlin Heidelberg, **2005**, pp. 1–56.

[52] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory (International Series of Monographs on Chemistry)*, Clarendon Press, **1994**.

[53] X. Fradera, M. A. Austen, R. F. W. Bader, *The Journal of Physical Chemistry A* **1999**, *103*, 304–314.

[54] E. Francisco, J. L. Casals-Sainz, T. Rocha-Rinza, Á. Martín Pendás, *Theoretical Chemistry Accounts* **2016**, *135*.

[55] T. Clark, M. Hennemann, J. S. Murray, P. Politzer, *Journal of Molecular Modeling* **2007**, *13*, 291–296.

[56] T. Clark, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *3*, 13–20.

[57] A. D. Becke, K. E. Edgecombe, *The Journal of Chemical Physics* **1990**, *92*, 5397–5403.

[58] K. Eskandari, H. Zariny, *Chemical Physics Letters* **2010**, *492*, 9–13.

[59] D. J. R. Duarte, G. L. Sosa, N. M. Peruchena, *Journal of Molecular Modeling* **2012**, *19*, 2035–2041.

[60] V. Tognetti, L. Joubert, *Theoretical Chemistry Accounts* **2015**, *134*.

[61] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, , D. J. Fox, *Gaussian09 Revision D.01*, Gaussian Inc., Wallingford CT, 2013.

[62] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *The Journal of Chemical Physics* **2010**, *132*, 154104.

[63] Y. Zhao, D. G. Truhlar, *Theoretical Chemistry Accounts* **2007**, *120*, 215–241.

[64] A. Bauzá, I. Alkorta, A. Frontera, J. Elguero, *Journal of Chemical Theory and Computation* **2013**, *9*, 5201–5210.

[65] N. Mardirossian, M. Head-Gordon, *Journal of Chemical Theory and Computation* **2016**, *12*, 4303–4325.

[66] Y. Wang, P. Verma, X. Jin, D. G. Truhlar, X. He, *Proceedings of the National Academy of Sciences* **2018**, *115*, 10257–10262.

[67] A. S. Mikherdov, A. S. Novikov, V. P. Boyarskiy, V. Y. Kukushkin, *Nature Communications* **2020**, *11*, 2921.

[68] A. S. Ostras', D. M. Ivanov, A. S. Novikov, P. M. Tolstoy, *Molecules* **2020**, *25*, 1406.

[69] M. A. Kryukova, D. M. Ivanov, M. A. Kinzhalov, A. S. Novikov, A. S. Smirnov, N. A. Bokach, V. Y. Kukushkin, *Chemistry – A European Journal* **2019**, *25*, 13671–13675.

[70] M. V. Kashina, M. A. Kinzhalov, A. S. Smirnov, D. M. Ivanov, A. S. Novikov, V. Y. Kukushkin, *Chemistry – An Asian Journal* **2019**, *14*, 3915–3920.

[71] M. A. Kinzhalov, M. V. Kashina, A. S. Mikherdov, E. A. Mozheeva, A. S. Novikov, A. S. Smirnov, D. M. Ivanov, M. A. Kryukova, A. Y. Ivanov, S. N. Smirnov, V. Y. Kukushkin, K. V. Luzyanin, *Angewandte Chemie International Edition* **2018**, *57*, 12785–12789.

[72] W. Liu, D. Peng, *The Journal of Chemical Physics* **2009**, *131*, 031104.

[73] D. M. Ivanov, S. V. Baykov, A. S. Novikov, G. Romanenko, N. A. Bokach, R. A. Evarestov, V. Y. Kukushkin, *Crystal Growth & Design* **2020**, *20*, 3417–3428.

[74] D. S. Bolotin, M. V. Il'in, V. V. Suslonov, A. S. Novikov, *Symmetry* **2020**, *12*, 637.

[75] A. S. Mikherdov, S. A. Katkova, A. S. Novikov, M. M. Efremova, E. Y. Reutskaya, M. A. Kinzhalov, *CrystEngComm* **2020**, *22*, 1154–1159.

[76] D. S. Bolotin, N. S. Soldatova, M. Y. Demakova, A. S. Novikov, D. M. Ivanov, I. S. Aliyarova, A. Sapegin, M. Krasavin, *Inorganica Chimica Acta* **2020**, *504*, 119453.

[77] T. Nakajima, K. Hirao, *Chemical Physics Letters* **1999**, *302*, 383–391.

[78] F. Lovas, *Diatomic Spectral Database, NIST Standard Reference Database 114*, **2002**. http://www.nist.gov/pml/data/msd-di/index.cfm.

[79] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, G. K.-L. Chan, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1340.

[80] Á. Martín Pendás, E. Francisco, *Promolden. A QTAIM/IQA code (unpublished)*.

[81] *AIMAll (Version 17.11.14), Todd A. Keith, TK Gristmill Software, Overland Park KS, USA, 2017 (aim.tkgristmill.com).*

[82] *Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/.*

[83] J. D. Hunter, *Computing in Science & Engineering* **2007**, *9*, 90–95.

[84] G. C. Pimentel, *The Journal of Chemical Physics* **1951**, *19*, 446–448.

[85] R. E. Rundle, *The Journal of Physical Chemistry* **1957**, *61*, 45–50.

[86] Á. Martín Pendás, E. Francisco, *ChemPhysChem* **2019**, *20*, 2722–2741.

[87] V. Oliveira, E. Kraka, D. Cremer, *Physical Chemistry Chemical Physics* **2016**, *18*, 33031–33046.

[88] T. Clark, J. S. Murray, P. Politzer, *ChemPhysChem* **2018**, *19*, 3044–3049.

[89] J. S. Murray, P. Politzer, *Crystals* **2020**, *10*, 76.

[90] E. Francisco, A. Martín Pendás in *Non-Covalent Interactions in Quantum Chemistry and Physics*, A. Otero de la Roza, G. A. DiLabio (Eds.), Elsevier, **2017**, pp. 27 – 64.

Journal Name

Electron-pair bonding in real space. Is the charge-shift family supported?

J. Luis Casals-Sainz, F. Jiménez-Grávalos, E. Francisco, A. Martín Pendás[*†]

Charge-shift bonding (CSB) has been introduced as a distinct third family of electron-pair links that adds to the covalent and ionic tradition. However, the full battery of orbital invariant tools provided by modern real space artillery shows that it is difficult to find CSB signatures outside the original valence-bond framework in which CSB was developed. The CSB concept should therefore be further investigated.

More than a hundred years after Lewis seminal paper,[1] the electron-pair bond may still be considered the most central concept in Chemistry. After its inception, its initial theoretical development by Heitler-London[2] and its incorporation to mainstream computational chemistry,[3] the Lewis pair lies at the core of chemical thinking. Through Pauling's[4] tour de force, the shared electron pair gave rise to the two major covalent (or polar-covalent) and ionic bonding families. These emerge as the degree of atomic sharing of the pair deviates from equality, and have carved the chemists' way of thinking. Subjected to reasonable generalizations, the covalent-ionic dichotomy has resisted exceedingly well the test of time.

Fundamental as they are, neither the Lewis pair nor the electron-pair bonding families are linked to quantum mechanical observables, and their significance rests on how a provided quantum mechanical wavefunction Ψ is interpreted. Usually, this relies on what theoretical framework has been used to generate Ψ, and although as the level of theory is improved all frameworks converge, interpretations do not necessarily do so, *vanishing into thin air*.[5] To understand how an electron pair is shared between two atoms, we have to deal with at least (i) the indistinguishability of electrons and electron pairs and (ii) the absence of isolatable atoms in the quantum mechanical description of two interacting systems. How this is done depends on the theoretical paradigm used, e.g. the valence bond (VB) or the molecular orbital (MO) viewpoints.

The existence of two main pair-electron bonding families has been challenged by the addition of a new *charge-shift bonding* (CSB) category. This is introduced in non-orthogonal VB (NOVB)

Departamento de Química Física y Analítica, Universidad de Oviedo, Oviedo. Spain.
E-mail: ampendas@uniovi.es
† Electronic Supplementary Information (ESI) available: basic real space chemical bonding, the two-state model, high level resuls, computational details and raw data. See DOI: 10.1039/b000000x/

as a distinct bonding class in which the bonding energy does come neither from the so-called spin-pairing of covalent bonds nor from the electrostatic stabilization of ionic links, but from the large resonance energy (RE) between the VB covalent and ionic structures. Although the paradigm of a CSB system is the F_2 molecule, whose VB covalent structure is unbound in marked contrast to a *normal* covalent bond like the one in H_2, many other cases have been found over the years, and a number of informative presentations can be found in the literature.[6]

Since according to Carl Sagan's standard, *extraordinary claims require extraordinary evidence*, the proposers of the CSB category have tried to offer that evidence by linking CSB to MO theory; to real space descriptors of chemical bonding like density differences, the laplacian of the electron density or the electron localization function (ELF); and to experimental behavior. Charge-shift bonds were soon associated to systems affected by the lone pair bond weakening effect (LPBWE) described by Sanderson,[7] so efforts have also been put on rationalizing their physical origin.[8] This was found to lie in very large kinetic energies coming from two-center three-electron Pauli repulsions between bonding electrons and compact lone pairs.

A decisive test of any new concept should be its independency from the theory used to *discover* or define it. In chemical bonding a framework invariant paradigm is available. It is based on (i) partitioning the space into 3D regions associated to atoms and (ii) on computing quantum mechanical observables within these regions. Although several partitioning strategies exist, the one provided by the quantum theory of atoms in molecules (QTAIM) is widely used.[9] These techniques need only a proper wavefunction, are orbital invariant by construction, and provide indices and energetic quantities which are directly related to Lewis' concepts.

In this Communication we show that charge-shift bonds display all the characteristics of a standard covalent interaction at larger

than expected distances. The latter are forced by the LPBWE, but the nature of the interaction has no specific peculiarity. This is found by critically examining the CSB family in the light of real space reasoning.

Electron-pair bonding in real space

Real space theories of the chemical bond have acquired a considerable level of maturity and notoriety.[10] Once an atomic partition of space is defined appropriately (through the QTAIM, for instance), one has access to both electron counting (through electron distribution functions, EDFs[11]) and energetic (through the interacting quantum atoms scheme, IQA[12]) descriptors. EDFs, that provide the probability of finding a given distribution of the N electrons of a system into atoms, behave as Pauling's resonance structures. Their knowledge gives also the average atomic electron population (thus the atomic net charge, Q^A and the ionic bond order,[13] $\iota^{AB} = -Q^A Q^B$) and all their further statistical moments. The interatomic covariance, or an equivalent form known as the delocalization index δ^{AB}, measures the number of shared electron pairs between atoms and provides the covalent bond order.[14] Similarly, the binding energy of a molecule becomes a sum of atomic deformation energies, E_{def}^A, and interatomic interactions, E_{int}^{AB}. The former basically correspond to traditional promotion costs, and the latter determine *in situ* bond energies, separated into covalent (cov) and electrostatic (ion) terms. For each atomic pair, the covalent energy is proportional to the number of shared pairs, while the electrostatic one is dominated by the product of the atomic charges. Ionicity and covalency are thus uniquely and invariantly defined. A brief introduction to this formalism is found in the supplementay information (ESI).

For two-center (A,B) electron-pair bonds, only three electron distributions are possible. Two zwitterionic, with the two electrons residing on either of the atoms —(2,0) and (0,2)—, and one similar to the VB-covalent structure, (1,1). Denoting the probability that one electron lies in A as p and using a $-1 \leq f \leq 1$ correlation factor between the two electrons, all 2c,2e links can be mapped. It is easy to show (see the ESI) that $\iota + \delta = 1$ when $f = 0$, so that ionicity excludes covalency and vice versa for electron-pair links (not in more complex situations[15]), and that $\delta = 4p(1-p)(1-f)$, this meaning also that covalency comes from electron delocalization and that it implies a non-vanishing population of the real space ionic resonance structures. A full classification for correlated situations is also known.[16]

Charge-shift bonding

CSB in NOVB appears when much (or all) of the binding energy of a system comes from the resonance energy between the VB-covalent, Ψ_{cov}, and the VB-ionic, Ψ_{ion}, structures, i.e. when neither $\langle \Psi_{cov}|\hat{H}|\Psi_{cov}\rangle$ nor $\langle \Psi_{ion}|\hat{H}|\Psi_{ion}\rangle$ is appropriately bound and it is $\langle \Psi_{cov}|\hat{H}|\Psi_{ion}\rangle$ that determines the bond energetics. Being consubstantial to Quantum Mechanics via the superposition principle, resonance is representation dependent: a change of basis changes its value. Even more, two non-orthogonal states coupled by a large resonance integral may become uncoupled after orthogonalization, and vice versa. *The resonance energy condition is therefore not invariant, and cannot serve our purpose to uncover invariant features of CSB*. Arguments based on two-configuration self-consistent field calculations,[17] that have been offered as a proof that CSB also appears under MO theory, simply mimic the NOVB wavefunction, and do not change the above conclusion. In the form of a two-electron two-orbital configuration mixing (CI(2,2)), this oversimplified description captures the essential elements of CSB, allowing us to examine it also under a real space perspective. To that end, we construct a model homoatomic two-electron wavefunction with strictly localized functions a and b with overlap $\langle a|b\rangle = S$, so that the g and u orthogonal MOs are written as $g, u = N_{g,u}(a \pm b)$. The spatial parts of the Heitler-London VB singlet covalent and ionic functions are the textbook $\Psi_{cov} \equiv (ab + ba)$ and $\Psi_{ion} \equiv (aa + bb)$, so that the NOVB function is $\Psi = c\Psi_{cov} + i\Psi_{ion}$, while its complete active space (CAS) analogue becomes $\Psi = \lambda|g\bar{g}\rangle + \mu|u\bar{u}\rangle$, with $\lambda^2 + \mu^2 = 1$. A one-to-one map $(c,i) \leftrightarrow (\lambda,\mu)$ exists. Full details are found in the ESI. The Hartree-Fock mean-field solution corresponds to $\lambda = 1$. In H_2, Ψ_{cov} provides a rather accurate binding curve, and mixing it with Ψ_{ion} introduces a slight resonance stabilization with a final small i contribution. In F_2 none of Ψ_{cov}, Ψ_{ion} or Ψ_{HF} are bound, but the CAS(2,2) Ψ is. Such a calculation within the $2p_z$ manifold with a 6-31G* basis provides $R_e = 1.48$ Å and $D_e \approx 16$ kcal mol^{-1}, to be compared with the experimental 1.40 Å and 39 kcal mol^{-1} values, respectively. The essential features of CSB are thus captured by this naïve model. If orbitals are not allowed to relax, then the Ψ_{cov} and Ψ_{ion} functions correspond to a $(\lambda_c, -\mu_c)$, (λ_c, μ_c) pair, with $\lambda_c, \mu_c > 0$, respectively. Ψ_{HF} lies at $\mu = 0$, so examining the $-\lambda_c \leq \mu \leq \lambda_c$ window provides a continuous map that visits all the four $\Psi_{cov}, \Psi_{CI}, \Psi_{HF}$, and Ψ_{ion} states. We use CAS orbitals in the following. It is relevant (ESI) that the first and second order densities of the 2e subsystem have very simple expressions in the λ, μ space: $\rho(r;r') = 2\lambda^2 g(r)g(r') + 2\mu^2 u(r)u(r')$, so any one-electron operator property varies linearly with λ^2; $\rho_2(r_1, r_2) = 2\lambda^2 g^2(r_1)g^2(r_2) + 2\mu^2 u^2(r_1)u^2(r_2) + 2\lambda\mu\, g(r_1)u(r_1)g(r_2)u(r_2)$. The latter expression shows that the $gg - uu$ resonance lies only in the two-electron $2\lambda\mu\, g(r_1)u(r_1)g(r_2)u(r_2)$ term if orbitals are not relaxed. This points to electron-electron repulsion, and not to the large kinetic energy of the covalent term (as usually admitted in the CSB literature) as the driving force for CSB.

Plenty of real space signatures of CSB have been reported so far.[18] For instance, negative deformation electron densities between the nuclei, indicative of density retraction in the bonding region. However, as recognized,[18] $\Delta\rho$ is reference dependent, and a change from spherically-averaged to valence-prepared atomic references may change the sign of $\Delta\rho$, *which is not invariant*. The sign of the laplacian of the density at the QTAIM bond critical point (bcp) has also been advocated as a CSB feature. A homonuclear *covalent* system usually displays $\nabla^2\rho(r_{bcp}) < 0$ values, but these become positive in F_2 and many other CSB cases. Recall, however, that for a simple $\rho(r) = Ne^{-\zeta r}$ exponential decay $\nabla^2\rho(r) = Ne^{-\zeta r}(\zeta^2 - 2\zeta/r)$, that turns positive after a critical distance. For instance, when stretching a normal H_2 molecule, its laplacian becomes positive at the bcp at about $R_{HH} = 3.2$ au, so that a positive laplacian just indicates a too long distance. Moreover, since in a homodiatomic $u(r_{bcp}) = 0$, it can be shown (ESI) that $\nabla^2\rho(r_{bcp})$ is positive if $\mu = 1$, so a large g, u mixing makes

Table 1 6-31G* data for F_2 with CAS(2,2) orbitals at $R_e = 1.49$ Å. gg and uu refer to MO determinants built from fixed CAS orbitals. The CAS contributions have been separated into those of the 2e active system and those of the remaining 16e electrons. The latter are also classified into π and σ symmetries. All data in au obtained at the bcp.

Ψ	ρ_{bcp}	$\nabla^2\rho$	ELF	CAS	ρ_{bcp}	$\nabla^2\rho$
gg	0.22	0.46	0.57	18e	0.21	0.66
CAS	0.21	0.66	0.43	16e	0.06	0.90
uu	0.06	2.68	0.02	2e	0.15	-0.23
				π	0.00	0.42
				σ	0.06	0.48

Table 2 Total energy components for several 6-31G* wavefunction of F_2 with fixed CAS(2,2) orbitals at $R_e = 1.49$ Å. All data in au (198 and 533 au have been added to E and V_{ne} respectively, and 198, 108, and 20 au subtracted from T, V_{ee}, and V_{ee}^{ca}, respectively). Energy at the dissociation limit $E = -198.724$ au. $\lambda_{cov} \approx 0.83$ estimated from orbital overlaps (ESI). V_{ee}^a and V_{ee}^{ca} are the interelectron repulsion among the active electrons and among the frozen core and the active electrons, respectively.

Ψ	E	T	V_{ne}	V_{ee}	V_{ee}^{ca}	V_{ee}^a
gg	-0.656	-0.034	-0.373	0.025	0.264	0.660
CAS	-0.751	0.290	-0.811	0.044	0.444	0.499
cov	-0.641	1.229	-2.078	0.482	0.967	0.414
ion	-0.075	1.229	-2.078	1.048	2.001	0.980
uu	0.173	3.475	-5.110	2.083	2.220	0.762

Table 3 Relevant IQA data in 6-31G* for different two-state wavefunctions of H_2 and F_2. $\lambda_{cov} \approx 0.982$ in dihydrogen. All data in au and all orbitals and distances fixed to those in the CAS.

H_2	E_{def}^A	T^A	V_{ee}^{AA}	E_{int}^{AB}	E_{cov}^{AB}	V_{ee}^{AB}	δ^{AB}
cov	0.005	0.601	0.136	-0.153	-0.194	0.321	0.715
CAS	0.013	0.578	0.165	-0.189	-0.222	0.296	0.833
HF	0.043	0.567	0.198	-0.230	-0.262	0.269	1.000
ion	0.132	0.601	0.263	-0.280	-0.321	0.195	1.284

F_2	E_{def}^A	T^A	V_{ee}^{AA}	E_{int}^{AB}	E_{cov}^{AB}	V_{ee}^{AB}	δ^{AB}
cov	0.092	99.614	40.435	-0.084	-0.122	27.611	0.397
CAS	0.077	99.145	40.237	-0.164	-0.199	27.569	0.713
HF	0.173	98.983	40.270	-0.263	-0.297	27.484	1.206
ion	0.526	99.614	40.870	-0.387	-0.425	27.381	2.040

the laplacian tend to a positive value necessarily, whatever the distance. Similarly, g,u mixing decreases $\rho(\boldsymbol{r}_{bcp})$ from $g^2(\boldsymbol{r}_{bcp})$ to 0 as λ goes from one to zero (ESI). Since all these properties depend on the one-matrix, and we have just shown that resonance is a two-electron phenomenon in an orthogonal framework, neither the behavior of the density nor of the laplacian can be related to covalent-ionic resonance in CSB.

The results of this model can be tested on real calculations. Table 1 shows how scalar fields at the bcp behave in F_2. Interestingly, ρ is dominated by the 2e active subsystem, but $\nabla^2\rho$ is not. Surprisingly, $\nabla^2\rho$ of the 2e subsystem is negative, like in H_2, and it is the effect of the *non-bonding* electrons that makes it overall positive. A similar criticism can be cast on arguments based on the ELF function, which is commonly obtained from the 1-matrix. CSB has been related to small ELF bonding domains with low electron count and large population variances (or fluctuations). However, the standard ELF function is difficult to generalize outside the single-determinant (SD) framework[19] (needed to deal with CSB), shedding doubts on its meaning. Moreover, variances obtained from SD expressions are in gross error when applied to correlated (or DFT) descriptions. Sometimes, even the ELF disynaptic bonding domain is absent in high accuracy wavefunctions (like in coupled-cluster calculations of F_2[18]) and the variance argument can simply not be applied. All this is compatible with an old proposition that would consider F_2 and other systems in a proto-bonding situation,[20] but is hardly related with covalent-ionic resonance.

Turning to cohesion, CSB has been linked to the kinetic energy (T), in line with Kutzelnigg and Ruedenberg classical arguments,[21] although the latter have been criticized for they ignore other equally important energy components in favor of T.[22] According to this view, in standard covalent bonding the virial theorem is restored as a bond is being formed through an increase in T driven by orbital contraction. When lone pairs give rise to large T's due to 2c-3e repulsions, a new restoration mechanism sets in and heavy mixing with the ionic configurations, which largely decrease T, serve this purpose. As a result, *large fluctuations of the pair density are thought to appear*.[18] However, in a fixed orbital framework which allows comparison across VB or MO paradigms, T, a one-electron property, changes linearly (with λ^2) between its gg (lower) and uu (higher) limits. Were it not for the V_{ee} electron repulsion, there would be no gg,uu mixing at all. Moreover, T in the covalent and ionic mixtures (ESI) are identical with fixed orbitals. Table 2 shows actual data. For fixed orbitals, the HF determinant always displays the smallest T, but its density is not

compact enough and the total one-electron energy ($T + V_{ne}$) is more stabilizing for the uu state than for the gg one (contrarily to what happens in H_2, see the ESI). From the VB point of view, the covalent structure has too large T, and the system responds by c,i mixing so that both T and the electron repulsion decrease.

A clue about the effect of c,i or λ,μ mixing is found on separating V_{ee} into core-core (c), core-active (ca) and active-active (a) pieces. The former is constant, the latter is smallest in the Ψ_{cov} calculation (as in H_2 with no lone-pairs). It is thus the ca repulsion that dominates the global V_{ee} behavior, being too large in the VB covalent structure. *The absence of lone pairs leads to no V_{ee}^{ca}, to a one-electron energy decreasing form uu to gg, and to small c,i mixing that decreases V_{ee} of the active orbitals. The presence of many lone pairs leads to the contrary behavior: one-electron energy decreasing towards uu, dominating ca repulsion, and large c,i mixing that tries to decrease its impact.* As shown in the ESI, the first behavior is found from H_2 up to N_2, the second in O_2 and F_2.

Is then CSB a distinct family of pair-electron bonding? Real space analysis clarifies this by answering the following questions: Is covalency changed in any substantial way in CSB? Is CSB characterized by a larger than usual fluctuation of the electron pair? IQA data for the model 6-31G* H_2 and F_2 systems are found in Table 3. Notice than in homodiatomics, all real space partitions coincide with that of the QTAIM. As seen, the covalent interaction in both CAS models as well as their delocalization indices are quite similar. In H_2, as expected, atomic deformations grow from the covalent to the CAS to the ionic solutions as we increase the contribution of the deformed hydride-like terms. The optimal CAS mixing is achieved close to Ψ_{cov}: the slight increase in the CAS E_{def} is compensated through E_{int} via E_{cov}. In F_2, $E_{def}(\Psi_{cov})$ is almost 20 times larger than in H_2, with a rather smaller E_{int}

(due to a larger distance and a more compact electron distribution). Mixing increases the interaction, as in H_2, but *decreases* the deformation. Going further, *this is clearly due to a considerable decrease in the intra-atomic electron repulsion, which is minimum in the CAS function.* As shown in the ESI, the larger than expected distance is a direct effect of this intra-atomic effect, which grows very quickly on approaching the two atoms. Aside from this, covalency displays the standard behavior overall. Invariant real space arguments uncover intra-atomic electron repulsion as the driving force behind CSB anomalies. *These are carried over with the atom, in agreement with Sanderson's LPBWE insights.* As fluctuations are regarded, the EDF shows only one delocalized electron pair in F_2, so only the $(9,9)$ and $(10,8) \equiv (8,10)$ structures have non-negligible probabilities. In the CAS solution of H_2, $p(1,1) = 0.583$, $p(2,0) = 0.208$. At the CAS solution of F_2, no anomalous fluctuation is found, with $p(9,9) = 0.684$ and $p(10,8) = 0.151$, the latter being even smaller than $p(2,0)$ in H_2, in line with the similar structure VB weights for H_2 and F_2 (ESI). This is compatible with a proto-bond with hindered delocalization. *At complete variance with CSB arguments, the electron-pair fluctuation is smaller in F_2 than in a purportedly normal covalent bond.*

All these model results, which we think provide useful insights, are fully supported by heat-bath multiconfigurational correlated calculations that match well the experimental bond distances and dissociation energies, both in F_2 and in a number of representative CSB systems. Details are found in the ESI. In F_2, for instance, a $(23,14)$ AVAS CASSCF provides a much compacted $R_e = 1.41$ Å, but a fluorine deformation energy of 0.078 au very close to that in the model, a stronger covalent energy of -0.259 au, and a considerably smaller electrostatic repulsion that justifies the increased binding. In these higher levels of theory the intra-atomic repulsion is lower and permits a decrease in R_e together with a more extensive delocalization. δ increases to 0.886, although its main σ eigen-component is still low (0.613), and $p(9,9) = 0.609$, $p(10,8) = 0.188$ approach the values of a correlated 2c-2e covalent bond with $f = 0.21$. Equivalent insights are obtained after examining the $Z = 6, 7, 8$ second period diatomics, N_2H_4, H_2O_2, heavier representative diatomics like S_2, Cl_2, Br_2, and some CSB transition metal systems like Cu_2, Ag_2, Au_2 or Hg_2^{2+} (ESI). Finally, polar CSBs, like those in H-F, C-F, or Si-F bonds have also been analyzed. In CH_3F, for instance, the CH_3-F -0.393 au ionic bond energy is considerably larger than the -0.261 au covalent one, with $\delta = 0.896$. Analysis of the CH_3-F EDF gives $p(n_{CH_3} = 8, n_F = 10) = 0.586$, $p(9,9) = 0.300$, $p(10,8) = 0.076$ which describes a polar interaction with small $f = 0.22$, a value very close to that in F_2. In SiH_3F, the ionic energy escalates to -0.715 au, covalency decreases with $\delta^{Si,F} = 0.591$, and $f \approx 0.60$. No sign of larger than normal ($f < 0$) fluctuations is found. Much on the contrary, in this case the link has clear proto-bonding ($f \gg 0$) signatures. In propellane, the deformation energy of the bridging carbons (b), E_{def}^b, is even smaller than that for non-bridge C's, E_{def}^{nb}, and a bridge proto-bond is found with $E_{cov}^b = -0.109$ au, $\delta^b = 0.410$.

Summarizing, although CSB stands out clearly in NOVB theory, it is difficult to find specific invariant signatures of this purported third electron-pair bond family, which in our opinion is still better described in terms of a classical Sanderson's intra-atomic LPBWE leading to proto-bonding.

References

1 G. N. Lewis, *J. Am. Chem. Soc.*, 1916, **38**, 762–786.

2 W. Heitler and F. London, *Z. Physik*, 1927, **44**, 455–472.

3 B. M. Gimarc, *Molecular structure and bonding. The qualitative molecular orbital approach*, Academic Press, New York, 1979.

4 L. Pauling, *The Nature of the Chemical Bond*, Cornell Univ. Press., Ithaca, N. Y., third Ed., 1960.

5 R. S. Mulliken, *J. Chem. Phys.*, 1965, **43**, S2–S11.

6 S. Shaik, D. Danovich, W. Wu and P. C. Hiberty, *Nat. Chem.*, 2009, **1**, 443–449.

7 R. T. Sanderson, *Polar Covalence*, Academic Press, New York, 1983.

8 P. C. Hiberty, R. Ramozzi, L. Song, W. Wu and S. Shaik, *Faraday Discuss.*, 2007, **135**, 261–272.

9 R. F. W. Bader, *Atoms in Molecules*, Oxford University Press, Oxford, 1990.

10 P. L. A. Popelier, *Structure and Bonding, Vol 115.*, Springer, Berlin, 2005, ch. Quantum Chemical Topology: On Bonds and Potentials, pp. 1–56.

11 E. Francisco, A. Martín Pendás and M. A. Blanco, *J. Chem. Phys.*, 2007, **126**, 094102–1–094102–13.

12 M. A. Blanco, A. Martín Pendás and E. Francisco, *J. Chem. Theory Comput.*, 2005, **1**, 1096–1109.

13 A. Martín Pendás and E. Francisco, *Phys. Chem. Chem. Phys.*, 2018, **20**, 16231–16237.

14 C. Outeiral, M. A. Vincent, A. Martín Pendás and P. L. A. Popelier, *Chem. Sci.*, 2018, **9**, 5517–5529.

15 J. L. McDonagh, A. F. Silva, M. A. Vincent and P. L. A. Popelier, *J. Phys. Chem. Lett.*, 2017, **8**, 1937–1942.

16 A. Martín Pendás, E. Francisco and M. A. Blanco, *Phys. Chem. Chem. Phys.*, 2007, **9**, 1087–1092.

17 H. Zhang, D. Danovich, W. Wu, B. Braïda, P. C. Hiberty and S. Shaik, *J. Chem. Theory Comput.*, 2014, **10**, 2410–2418.

18 S. Shaik, D. Danovich, B. Silvi, D. L. Lauvergnat and P. C. Hiberty, *Chem. Eur. J.*, 2005, **11**, 6358–6371.

19 E. Matito, B. Silvi, M. Duran and M. Solà, *J. Chem. Phys.*, 2006, **125**, 024301.

20 B. Silvi and A. Savin, *Nature*, 1994, **371**, 683–686.

21 K. Ruedenberg, *Rev. Mod. Phys.*, 1962, **34**, 326–376.

22 G. B. Bacskay and S. Nordholm, *J. Phys. Chem. A*, 2013, **117**, 7946–7958.

Bibliography

(1) Mikulecky, D. C. In *Worldviews, Science and Us*, 2005, pp 30–52.

(2) Rasmussen, J., *On the structure of knowledge - a morphology of metal models in a man-machine system context*; Risø-M 2192; Risø National Laboratory: 1979.

(3) Frigg, R.; Nguyen, J. In *The Stanford Encyclopedia of Philosophy*, Zalta, E. N., Ed.; Metaphysics Research Lab, Stanford University: 2020.

(4) McComas, W. F. *Sch. Sci. Math.* **1996**, *96*, 10–16.

(5) Williams, L. P. In *Encyclopedia Britannica*, Zalta, E. N., Ed., 2018.

(6) Barseghyan, H.; Overgaard, N.; Rupik, G. Introduction to History and Philosophy of Science, 2018.

(7) Gribbin, J., *Historia de la ciencia, 1543-2001*; Crítica, SL: 2003.

(8) Dampier, W. C., *Historia de la ciencia y de sus relaciones con la filosofía y la religión*; Aguilar: 1950.

(9) Schiff, L. I., *Quantum Mechanics*; McGraw-Hill Book Co., Inc.: 1968.

(10) Bohm, D., *Quantum Theory*; Dover Publications, Inc.: 1968.

(11) Ling, S. J.; Sanny, J.; Moebs, W. University Physics Volume 3, 2016.

(12) Richtmyer, F. K., *Introduction to modern physics*; McGraw-Hill Book Co., Inc.: 1928.

(13) Levine, I. N., *Quantum Chemistry*; Prentice Hall: 2000.

(14) Atkins, P.; Friedman, R., *Molecular Quantum Mechanics*; Oxford University Press: 2005.

(15) Szabó, A.; Ostlund, N. S., *Modern quantum chemistry : introduction to advanced electronic structure theory*; Dover Publications, Inc.: 1996.

(16) Helgaker, T.; Jørgensen, P.; Olsen, J., *Molecular Electronic-Structure Theory*; John Wiley & Sons, Ltd: 2000.

(17) Jorner, K.; Tomberg, A.; Bauer, C.; Sköld, C.; Norrby, P.-O. *Nat. Rev. Chem.* **2021**, *5*, 240–255.

(18) Lewars, E. G., *Computational Chemistry. Introduction to the Theory and Applications of Molecular and Quantum Mechanics*; Springer International Publishing Switzerland 2016: 2016.

(19) Jensen, F., *Introduction to Computational Chemistry*; John Wiley & Sons, Inc.: 2006.

(20) McArdle, S.; Endo, S.; Aspuru-Guzik, A.; Benjamin, S. C.; Yuan, X. *Rev. Mod. Phys.* **2020**, *92*.

(21) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. *J. Chem. Phys.* **2007**, *126*, 144111.

(22) Fonseca Guerra, C.; Handgraaf, J.-W.; Baerends, E. J.; Bickelhaupt, F. M. *J. Comput. Chem.* **2004**, *25*, 189–210.

(23)   Rousseau, B.; Peeters, A.; Van Alsenoy, C. *J. Mol. Struct.-THEOCHEM* **2001**, *538*, 235–238.

(24)   AIMAll (Version 17.11.14), Todd A. Keith, TK Gristmill Software, Overland Park KS, USA, 2017 (aim.tkgristmill.com).

(25)   Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 2547–2553.

(26)   Clark, A. E.; Davidson, E. R. *Int. J. Quantum Chem.* **2003**, *93*, 384–394.

(27)   Mayer, I.; Salvador, P. *Chem. Phys. Lett.* **2004**, *383*, 368–375.

(28)   Li, L.; Parr, R. G. *J. Chem. Phys.* **1986**, *84*, 1704–1711.

(29)   Bultinck, P.; Popelier, P. In 2009, pp 215–228.

(30)   Bader, R. F. W., *Atoms in Molecules: A Quantum Theory*; International Series of Monographs on Chemistry; 22; Clarendon Press: 1994.

(31)   Popelier, P. L. A. In *The Chemical Bond II: 100 Years Old and Getting Stronger*, Mingos, D. M. P., Ed.; Structure and Bonding, 170; Springer International Publishing Switzerland 2016: 2016, pp 71–118.

(32)   Casals-Sainz, J. L.; Martín Pendás, Francisco, E. In *Intermolecular Interactions in Crystals: Fundamentals of Crystal Engineering*, Novoa, J. J., Ed.; The Royal Society of Chemistry: 2018, pp 178–221.

(33)   Stone, A. J., *The Theory of Intermolecular Forces*; Oxford University Press: 2013.

(34)   Hayes, I.; Stone, A. *Mol. Phys.* **1984**, *53*, 83–105.

(35)   Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887–1930.

(36)   Morokuma, K. *J. Chem. Phys.* **1971**, *55*, 1236–1244.

(37)   Bickelhaupt, F. M.; Baerends, E. J. *Rev. Comput. Chem.* **2000**, *15*, 1–86.

(38)   Guevara-Vela, J. M.; Francisco, E.; Rocha-Rinza, T.; Martín Pendás, *Molecules* **2020**, *25*.

(39)   Ohanian, H. C. *Am. J. Phys* **1986**, *54*, 500–505.

(40)   Sebens, C. T. *HIST PHILOS M P* **2019**, *68*, 40–50.

(41)   Landau, L. D.; Lifshitz, E. M., *Quantum Mechanics*; John Wiley & Sons, Ltd: 1974.

(42)   Koch, W.; Holthausen, M. C., *A Chemist's Guide to Density Functional Theory*; John Wiley & Sons, Ltd: 2001.

(43)   Cramer, C. J., *Essectials of Computational Chemistry: Theories and Models*; Pergamon Press: 2002.

(44)   Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871.

(45)   Mardirossian, N.; Head-Gordon, M. *Mol. Phys.* **2017**, *115*, 2315–2372.

(46)   Zhao, Y.; Truhlar, D. G. *Chemical Physics Letters* **2011**, *502*, 1–13.

(47)   Mardirossian, N.; Head-Gordon, M. *Journal of Chemical Theory and Computation* **2016**, *12*, 4303–4325.

(48)   McWeeny, R.; Sutcliffe, B. T., *Methods of Molecular Quantum Mechanics*; Monographs of Theoretical Chemistry, Vol. 2; Academic Press: London, 1969.

(49)   Pendás, A. M.; Blanco, M. A.; Francisco, E. *J. Chem. Phys.* **2004**, *120*, 4581–4592.

(50)   Popelier, P., *Atoms in Molecules. An Introduction*; Pearson Education: United Kingdom, 2000.

(51) Fradera, X.; Austen, M. A.; Bader, R. F. W. *J. Phys. Chem. A* **1999**, *103*, 304–314.

(52) Fradera, X.; Poater, J.; Simon, S.; Duran, M.; Solà, M. *Theor Chem Acc* **2002**, *108*, 214–224.

(53) Outeiral, C.; Vincent, M. A.; Martín Pendás, Popelier, P. L. A. *Chem. Sci.* **2018**, *9*, 5517–5529.

(54) Menéndez, M.; Álvarez Boto, R.; Francisco, E.; Martín Pendás, *J. Comput. Chem.* **2015**, *36*, 833–843.

(55) Ziesche, P. In *Many-Electron Densities and Reduced Density Matrices*, Cioslowski, J., Ed.; Springer US: Boston, MA, 2000, pp 33–56.

(56) Francisco, E.; Martín Pendás, A.; Blanco, M. A. *J. Chem. Phys.* **2007**, *126*, 094102.

(57) Francisco, E; Menéndez, M.; Martín Pendás, Á, *Electron number distribution functions*. To be published.

(58) Blanco, M. A.; Martín Pendás, A.; Francisco, E. *J. Chem. Theory Comput.* **2005**, *1*, 1096–1109.

(59) Francisco, E.; Martín Pendás, A.; Blanco, M. A. *J. Chem. Theory Comput.* **2006**, *2*, 90–102.

(60) Martín Pendás, Á.; Francisco, E. PROMOLDEN. A QTAIM/IQA code (un-published).

(61) Stone, A.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047–1064.

(62) Francisco, E.; Menéndez Crespo, D.; Costales, A.; Martín Pendás, *J. Comput. Chem.* **2017**, *38*, 816–829.

(63) Francisco, E.; Casals-Sainz, J. L.; Rocha-Rinza, T.; Pendás, A. M. *Theor. Chem. Acc.* **2016**, *135*, 170.

(64) Suárez, D.; Díaz, N.; Francisco, E.; Martín Pendás, A. *ChemPhysChem* **2018**, *19*, 973–987.

(65) Jiménez-Grávalos, F.; Casals-Sainz, J. L.; Francisco, E.; Rocha-Rinza, T.; Pendás, Á. M.; Guevara-Vela, J. M. *Theor. Chem. Acc.* **2020**, *139*.

(66) Tomasi, J. *Theor. Chem. Acc.* **2004**, *112*, 184–203.

(67) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3094.

(68) Jiménez-Grávalos, F.; Díaz, N.; Francisco, E.; Martín-Pendás, Suárez, D. *ChemPhysChem* **2018**, *19*, 3425–3435.

(69) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094.

(70) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(71) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

(72) Neese, F. *WIREs Comput Mol Sci* **2012**, *2*, 73–78.

(73) Neese, F. *WIREs Comput Mol Sci* **2018**, *8*, e1327.

(74) Guevara-Vela, J. M.; Chávez-Calvillo, R.; García-Revilla, M.; Hernández-Trujillo, J.; Christiansen, O.; Francisco, E.; Martín Pendás, Rocha-Rinza, T. *Chem. Eur. J.* **2013**, *19*, 14304–14315.

(75) Guevara-Vela, J. M.; Romero-Montalvo, E.; Mora Gómez, V. A.; Chávez-Calvillo, R.; García-Revilla, M.; Francisco, E.; Pendás, M.; Rocha-Rinza, T. *Phys. Chem. Chem. Phys.* **2016**, *18*, 19557–19566.

(76)   Segarra-Martí, J.; Merchán, M.; Roca-Sanjuán, D. *J. Chem. Phys.* **2012**, *136*, 244306.

(77)   Goodman, L.; Gu, H.; Pophristic, V. *J. Phys. Chem. A* **2005**, *109*, 1223–1229.

(78)   Buissonneaud, D. Y.; van Mourik, T.; O'Hagan, D. *Tetrahedron* **2010**, *66*, 2196–2202.

(79)   Martins, F. A.; Freitas, M. P. *Eur. J. Org. Chem.* **2019**, *2019*, 6401–6406.

(80)   Bickelhaupt, F. M.; Baerends, E. J. *Angew. Chem. Int. Ed.* **2003**, *42*, 4183–4188.

(81)   Baranac-Stojanović, M. *RSC Adv.* **2014**, *4*, 43834–43838.

(82)   Thacker, J. C. R.; Popelier, P. L. A. *J. Phys. Chem. A* **2018**, *122*, 1439–1450.

(83)   Díaz, N.; Jiménez-Grávalos, F.; Suárez, D.; Francisco, E.; Martín-Pendás, *Phys. Chem. Chem. Phys.* **2019**, *21*, 25258–25275.

(84)   Ben-Naim, A., *Solvation Thermodynamics*; Springer: Boston, NA, USA, 1987.

(85)   Soteras, I.; Morreale, A.; López, J. M.; Orozco, M.; Luque, F. J. *Braz. J. of Phys.* **2004**, *34*, 48–57.

(86)   Pei, J.; Wang, Q.; Zhou, J.; Lai, L. *Proteins: Struct. Funct. Genet.* **2004**, *57*, 651–664.

(87)   Bernazzani, L.; Duce, C.; Micheli, A.; Mollica, V.; Sperduti, A.; Starita, A.; Tiné, M. R. *J Chem Inf Model* **2006**, *46*, 2030–2042.

(88)   Kang, H.; Choi, H.; Park, H. *J Chem Inf Model* **2007**, *47*, 509–514.

(89)   Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

(90)   Baldridge, K.; Klamt, A. *J. Chem. Phys.* **1997**, *106*, 6622–6633.

(91)   Barca, G. M. J. et al. *J. Chem. Phys.* **2020**, *152*, 154102.

(92)   Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. Minnesota Solvation Database – version 2012, University of Minnesota, Minneapolis, 2012.

(93)   Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 5665–5675.

(94)   Meylan, W. M.; Howard, P. H. *J. Pharm. Sci.* **1995**, *84*, 83–92.

(95)   Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

(96)   Ren, P.; Chun, J.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J.; Baker, N. A. *Q. Rev. Biophys.* **2012**, *45*, 427–491.

(97)   Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. *Chem. Rev.* **2013**, *114*, 779–814.

(98)   Bojarowski, S. A.; Kumar, P.; Dominiak, P. M. *ChemPhysChem* **2016**, *17*, 2455–2460.

(99)   Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *J. Chem. Phys.* **2000**, *112*, 7300–7306.

(100) Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 3330–3342.

(101) Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2014**, *10*, 4480–4487.

(102) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. *J. Chem. Theory Comput.* **2015**, *11*, 2609–2618.

(103) Jiménez-Grávalos, F.; Suárez, D. *J. Chem. Theory Comput.* **2021**, *17*, 4981–4995.

(104) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(105) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(106) Stone, A.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047–1064.

(107) Liu, C.; Piquemal, J.-P.; Ren, P. *J. Chem. Theory Comput.* **2019**, *15*, 4122–4139.

(108) Popelier, P. L. A. *Int. J. Quantum Chem.* **2015**, *115*, 1005–1011.

(109) Hughes, Z. E.; Thacker, J. C. R.; Wilson, A. L.; Popelier, P. L. A. *J. Chem. Theory Comput.* **2019**, *15*, 116–126.

(110) Yáñez, M.; Sanz, P.; Mó, O.; Alkorta, I.; Elguero, J. *J. Chem. Theory Comput.* **2009**, *5*, 2763–2771.

(111) Villanueva, E. F.; Mó, O.; Yáñez, M. *Phys. Chem. Chem. Phys.* **2014**, *16*, 17531–17536.

(112) Casals-Sainz, J. L.; Jiménez-Grávalos, F.; Costales, A.; Francisco, E.; Pendás, M. *J. Phys. Chem. A* **2018**, *122*, 849–858.

(113) Brea, O.; Mó, O.; Yáñez, M.; Alkorta, I.; Elguero, J. *Chem. Commun.* **2016**, *52*, 9656–9659.

(114) Wolters, L. P.; Schyman, P.; Pavan, M. J.; Jorgensen, W. L.; Bickelhaupt, F. M.; Kozuch, S. *WIREs Comput Mol Sci* **2014**, *4*, 523–540.

(115) Kolář, M. H.; Hobza, P. *Chem. Rev.* **2016**, *116*, 5155–5187.

(116) Cavallo, G.; Metrangolo, P.; Milani, R.; Pilati, T.; Priimagi, A.; Resnati, G.; Terraneo, G. *Chem. Rev.* **2016**, *116*, 2478–2601.

(117) Costa, P. J. *Phys. Sci. Rev.* **2017**, *2*.

(118) Wolters, L. P.; Smits, N. W. G.; Guerra, C. F. *Phys. Chem. Chem. Phys.* **2015**, *17*, 1585–1592.

(119) Novák, M.; Foroutan-Nejad, C.; Marek, R. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6440–6450.

(120) Bora, P. L.; Novák, M.; Novotný, J.; Foroutan-Nejad, C.; Marek, R. *Chem. Eur. J.* **2017**, *23*, 7315–7323.

(121) Thirman, J.; Engelage, E.; Huber, S. M.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2018**, *20*, 905–915.

(122) Syzgantseva, O. A.; Tognetti, V.; Joubert, L. *J. Phys. Chem. A* **2013**, *117*, 8969–8980.

(123) Yahia-Ouahmed, M.; Tognetti, V.; Joubert, L. *Comput. Theor. Chem.* **2015**, *1053*, 254–262.

(124) Yahia-Ouahmed, M.; Tognetti, V.; Joubert, L. *Theor. Chem. Acc.* **2016**, *135*.

(125) Eskandari, K.; Lesani, M. *Chem. Eur. J.* **2015**, *21*, 4739–4746.

(126) Jahromi, H. J.; Eskandari, K.; Alizadeh, A. *J. Mol. Model.* **2015**, *21*.

(127) Buralli, G.; Petelski, A.; Peruchena, N.; Sosa, G.; Duarte, D. *Molecules* **2017**, *22*, 2034.

(128) Guevara-Vela, J. M.; Ochoa-Resendiz, D.; Costales, A.; Hernández-Lamoneda, R.; Martín Pendás, Á. *ChemPhysChem* **2018**, *19*, 2512–2517.

(129) Orangi, N.; Eskandari, K.; Thacker, J. C. R.; Popelier, P. L. A. *ChemPhysChem* **2019**, *20*, 1922–1930.

(130) Alkorta, I.; Silva, A. F.; Popelier, P. L. A. *Molecules* **2020**, *25*, 2674.

(131) Jiménez-Grávalos, F.; Gallegos, M.; Martín Pendás, Novikov, A. S. *J. Comput. Chem.* **2021**, *42*, 676–687.

(132) Pimentel, G. C. *J. Chem. Phys.* **1951**, *19*, 446–448.

(133) Rundle, R. E. *J. Phys. Chem.* **1957**, *61*, 45–50.

(134) Wolters, L. P.; Bickelhaupt, F. M. *ChemistryOpen* **2012**, *1*, 96–105.

(135) Shaik, S.; Maitre, P.; Sini, G.; Hiberty, P. C. *J. Amer. Chem. Soc.* **1992**, *114*, 7861–7866.

(136) Shaik, S.; Danovich, D.; Wu, W.; Hiberty, P. C. *Nat. Chem.* **2009**, *1*, 443–449.

(137) Casals-Sainz, J. L.; Jiménez-Grávalos, F.; Francisco, E.; Martín Pendás, A. *Chem. Commun.* **2019**, *55*, 5071–5074.

(138) Shaik, S.; Danovich, D.; Galbraith, J. M.; Braïda, B.; Wu, W.; Hiberty, P. C. *Angewandte Chemie International Edition* **2020**, *59*, 984–1001.