



**Ana Catarina
Fernandes Araújo**

**Modelação de interacções intermoleculares em
hidratos de ácidos**

**Molecular modeling of intermolecular interactions in
acid hydrates**



Universidade de Aveiro Departamento de Química
2012

**Ana Catarina
Fernandes Araújo**

**Modelação de interacções intermoleculares em
hidratos de ácidos**

**Molecular modeling of intermolecular interactions in
acid hydrates**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Química, realizada sob a orientação científica do Doutor Paulo Jorge de Almeida Ribeiro-Claro, Professor Associado c/ agregação do Departamento de Química da Universidade de Aveiro

If you wish to understand the fragrance of the rose, or the tenacity of the oak; if you are not satisfied until you know the secret paths by which the sunshine and the air achieve these wonders; if you wish to see the pattern which underlies one large field of human experience and human measurement, then take up chemistry.

— C. A. Coulson, 1973

Dedico este trabalho ao meu Mestre, que magnificamente me guiou nesta épica jornada pela química computacional

o júri

Presidente

Prof. Dr. Artur Manuel Soares da Silva
professor catedrático do Departamento de Química da Universidade de Aveiro

Prof. Dr. Pedro Miguel Duarte Vaz
investigador auxiliar do Departamento de Química e Bioquímica da Faculdade de Ciências da
Universidade de Lisboa

Prof. Dr. Paulo Jorge de Almeida Ribeiro Claro
professor associado com agregação do Departamento de Química da Universidade de Aveiro

palavras-chave

Dissociação, hidratos de ácidos, ácido trifluoroacético, microsolvatação, cálculo de pKa, ciclos termodinâmicos, cluster-contínuo

resumo

Esta dissertação explora o mundo nanoscópico de pequenos agregados onde as pontes de hidrogénio têm um papel preponderantes usando métodos quânticos *ab-initio*. No capítulo introdutório, a área da química computacional é apresentada e algumas noções teóricas referentes aos métodos *ab-initio*, discutidas. No Capítulo 2, o desempenho de vários níveis de teoria é avaliado através do estudo de pequenos agregados de água. O capítulo 3 discute a influência dos critérios de otimização no resultado deste processo, alertando para erros comuns. No Capítulo 4, hidratos gasosos de ácido trifluoroacético (TFA), nas formas dissociada e não-dissociada, são apresentados. Um número mínimo de 4 moléculas de água é necessário para induzir a transferência do protão do TFA para a rede de moléculas de água adjacente. No entanto, 5 moléculas de água são necessárias para que o agregado dissociado se torne mais estável que o seu análogo não dissociado. O Capítulo 5 propõe um novo esquema para o cálculo *ab-initio* de valores de pKa. Este esquema serve-se de hidratos de ácido microsolvatado, nas formas dissociada e não dissociada, em modelo de solvatação contínuo, para calcular a energia livre de dissociação em solução. Para o conjunto de espécies testadas, incluindo 10 ácidos carboxílicos, 1 amina e 2 aminoácidos, o erro médio absoluto é 1.11, o declive experimental 1.2 e o coeficiente de correlação 0.92, o que indica um nível de exactidão aceitável.

Keywords

Dissociation, acid hydrates, trifluoroacetic acid, microsolvation, pKa calculation, thermodynamic cycles, cluster-continuum

abstract

This dissertation concerns the study of small hydrogen bonded systems through the use of quantum mechanical *ab-initio* methods. In the introductory chapter, the field of computational chemistry is presented and some basic theoretical notions concerning *ab-initio* methods are discussed. In Chapter 2, the performance of various levels of theory is assessed through the study of small water clusters. Chapter 3 discusses the influence of optimization criteria in the outcome of the optimization procedure, warning against common pitfalls. In Chapter 4, gas-phase hydrates of trifluoroacetic acid (TFA), in both dissociated and undissociated forms, are presented. A minimum of 4 water molecules is necessary to induce proton transfer from TFA to the neighboring water molecule network. However, 5 water molecules are needed to render the dissociated hydrate more stable than its undissociated counterpart. Chapter 5 proposes a new scheme for the *ab-initio* calculation of pKa values. It uses microsolvated acid hydrates, in both dissociated and undissociated forms, within a continuum solvation model, to calculate the dissociation free energy in solution. For the data set used, including 10 carboxylic acids, 1 amine and 2 aminoacids, the mean unsigned error (MUE) of calculated pKa values is 1.11, the experimental slope 1.2 and the correlation 0.92, which denotes a reasonable level of accuracy.

Index

Chapter 1.....	1
1. Computational chemistry – A Brave New World	2
1.1 What is computational chemistry useful for?	5
1.1.1 Spectroscopy	6
1.1.2 Drug design	7
1.1.3 Nanotechnology.....	7
1.1.4 Proteins in solution.....	8
1.2 The workings of Computational Chemistry.....	9
1.2.1 Foundations of ab-initio methods.....	12
1.2.1.1 The Schrödinger equation.....	12
1.2.1.2 Solving the Schrödinger equation for the H atom	15
1.2.2 Calculation methods for polyelectronic systems.....	17
1.2.2.1 The Born-Oppenheimer approximation.....	17
1.2.2.2 The independent electron approximation	18
1.2.2.3 The LCAO approach.....	19
1.2.2.4 The Variational Principle.....	20
1.2.2.5 The HF-Self Consistent Field procedure	21
1.2.3 Basis sets	21
1.2.3.1 Single, double and triple zeta basis sets.....	25
1.2.3.2 Split-valence basis sets	25
1.2.3.3 Polarized basis sets	26
1.2.3.4 Diffuse basis sets	26
1.2.3.5 Approaching the HF limit.....	27
1.2.4 Weaknesses of HF theory.....	27

1.2.5 Density Functional Theory	28
1.2.6 Møller-Plesset Perturbation Theory	30
1.3 Intermolecular Interactions and their importance	31
1.3.1 The challenge of modeling intermolecular interactions	32
1.3.2 The elusive nature of the hydrogen bond	34
1.4. Tools of the trade	38
Chapter 2.....	39
2. Mini-project: Exploring <i>ab-initio</i> methods through water clusters	40
2.1 Introduction	40
2.2 Computational methods.....	40
2.3 Results and discussion.....	42
Chapter 3.....	47
“GIGO” - Garbage In, Garbage Out: the case of linear water chains	48
Chapter 4.....	52
4. Stepwise solvation of trifluoroacetic acid: chasing the onset of proton dissociation	53
4.1 Introduction	53
4.2 Computational methods.....	57
4.3 Results and discussion.....	58
4.3.1 What is the minimum number of water molecules required for TFA dissociation?	58
4.3.2 Geometries of TFA hydrates	60
4.3.3 Evolution of energy of interaction with cluster size.....	64
4.3.4 Hydrogen bond energy	67
4.3.5 Chemical Hardness and stability	69
4.3.6 Eigen vs. Zundel structures.....	70
4.3.7 Bridging the gap between micro- and bulk solvation.....	73
4.4 Conclusion.....	74

Chapter 5.....	76
5. Calculation of the pKa of organic acids using a hybrid explicit-implicit scheme	77
5.1 Introduction	77
5.1.1 Equations and definitions	77
5.1.2 Thermodynamic cycles or how to obtain $\Delta G^*_{(aq)}$	78
5.1.3 Comment on the nature of pKa.....	83
5.2 Computational methods.....	84
5.3 Results and discussion.....	86
5.3.1 Global performance assessment	86
5.3.2 On the optimum number of explicit water molecules	92
5.3.3 Conformation affects the pKa estimate: the case of acetic acid	96
5.3.4 A lesson from glycine: the importance of explicitly solvating all hydrogen bonding sites.....	98
5.3.5 Calculating “pKa” using the electronic energy of dissociation instead of the Gibbs free energy of dissociation.....	99
5.4 Conclusion.....	102
6. References.....	104

Chapter 1

Computational chemistry – A Brave New World

1. Computational chemistry – A Brave New World

Up until the early XXth century, chemistry was dominated by a macroscopic, “bottom down” approach vision of research. As experiments produced new data, existing theories were adapted or reformulated in order to fit in the novel information. Following experiments were designed according to the rules expressed in those theories. This approach works reasonably well when dealing with bulk phenomena. However, when scientists started looking into ever smaller realms of matter those same theories neatly constructed upon experiment began to crumble like a house of cards.

The ultraviolet catastrophe in the second half of the XIXth century hinted at a major flaw in classical mechanics: the assumption that energy is continuous. Max Planck proposed in 1900 that a blackbody emits radiation in packets of energy, called quanta, with discrete values. In accordance with Planck, Einstein suggested in 1905 that light was formed by quantum particles, later named photons. In 1913 Bohr explained the spectral lines of the hydrogen atom as arising from the existence of quantized energy levels for the electron. In 1924, Louis deBroglie claimed that, just as light also behaves as particles, so do particles exhibit wave behavior, such as electron beams which are able to create interference patterns similar to those produced by waves.

In the light of these facts, a new atomic model accounting for the wave-like nature of electrons was needed and delivered by two equivalent although conceptually different theories: Schrödinger’s wave mechanics and Heisenberg and Born’s matrix mechanics. The former will be discussed in greater detail in the methods section. Finally, it is important to mention the uncertainty principle formulated by Heisenberg in 1927 which states that the position and momentum of a particle cannot be simultaneously known, since the observer inevitably alters the system through the act of measurement.

The revolution of quantum mechanics is not only one of science but of consciousness itself. An honest and inquiring mind, when confronted with such facts as

the uncertainty principle and quantum entanglement, starts an irreversible process of deconstructing its rigid view of a solid and compartmentalized reality.

“Some physicists would prefer to come back to the idea of an objective real world whose smallest parts exist objectively in the same sense as stones or trees exist independently of whether we observe them. This, however, is impossible.”
— Werner Heisenberg

The task that follows is as daunting as exciting: to build a new vision in which the reality accessible by the senses is the result of wavefunction collapse, a mere possibility lost in a sea of possibilities, a cosmic raffle at the same time witnessed and influenced by the observer. In the face of such an exquisite behavior of matter, old dogmas and prejudices are being dissolved at lightspeed. For example, the idea of parallel universes, once relegated to the realms of science fiction, is now discussed in scientific fora ^[1].

Other than revolutionizing the way science is *thought*, quantum mechanics also led to dramatic changes in the way science is *made* ^[2]. Schrödinger’s pioneering work ^[3] provided a mathematical description for the behavior of matter that, if only it could be solved exactly for atoms with more than one electron, would provide all the properties and predict the future behavior of any given physical system without the need for experiments. Well then, why is most of scientific progress still made through experimental work?

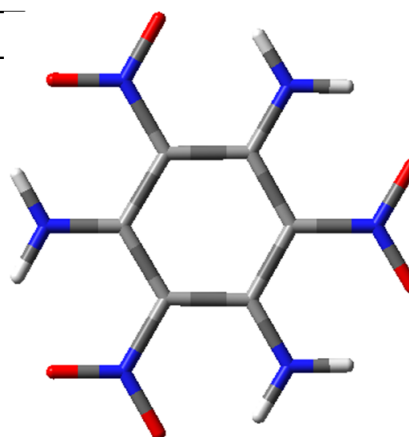
Considering chemistry, specifically, what is the need to study the behavior of molecules in a lab when one has in hand such a neat equation capable of predicting said behavior? Common sense says that if anything sounds too good to be true, it generally is and such an adage fits like a glove here. The first problem is that the Schrödinger’s equation can not be solved analytically for other than hydrogenoid atoms. The alternative is to solve it numerically, a strategy which involves such a large number of calculation steps that it can only be efficiently performed by computers. And here we find the first hurdle in the way of quantum chemistry’s growth: in the 30’s there just wasn’t enough computing power to perform those calculations in a timely manner. As Dirac famously put it:

“The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these equations leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation ^[4].”

As will be discussed in more detail in section 2, several “approximate practical methods”, or models, were indeed developed over the years allowing faster calculations albeit with less accuracy ^[5]. But perhaps the greatest contribution to the advancement of applied quantum chemistry came from an external source: the advent of digital computers and the exponential rise in their capacity. Today, it is possible to study systems with more than 100 atoms by means of quantum mechanical (QM) methods while in the 70’s researchers were still confined to small molecules in the gas phase. Table 1 shows how the evolution in computing power exponentially (and the increasing use of parallel computing) decreased the time necessary to calculate the energy of a molecule with 24 atoms – an impossible task back in 1967 is now trivial.

Table 1 The CPU time required to perform a single point energy calculation of a small molecule (represented on the right) has decreased exponentially in the last decades, as processors grew more powerful and software more efficient.

<i>Software</i>	<i>Processor</i>	<i>CPU time</i>
<i>Polyatom 67</i>	<i>CDC 1604</i>	<i>200 years*</i>
<i>Gaussian 80</i>	<i>VAX 11/780</i>	<i>1 week</i>
<i>Gaussian 88</i>	<i>Cray Y-MP</i>	<i>1 hour</i>
<i>Gaussian 92</i>	<i>Cray Y-MP</i>	<i>9 minutes</i>
	<i>486 DX2/50 MHz</i>	<i>20 hours</i>
<i>Gaussian 94</i>	<i>Pentium 90 MHz</i>	<i>2,6 hours</i>
<i>Gaussian 98</i>	<i>AMD 800 MHz</i>	<i>6 minutes</i>
<i>Gaussian 03</i>	<i>Pentium 4 3,4 GHz</i>	<i>62 seconds</i>



*Single-point energy calculation, HF/6-31G**
(sym-triaminotrinitrobenzene, 24 atoms, 132 electrons, 300 basis functions)*

** estimate*

In summary, the theory was all ready long ago but lacking both the necessary hardware and software for its widespread application. If quantum mechanics were a human being it would be called both a visionary, for presenting a view of reality ahead of its time, and a late-bloomer, since it only found generalized utility in the world of science many decades later. Although computational research is not yet as extensively used as its experimental counterpart and may never surpass it in importance, the former complements the latter. The best proof of this potential is the rising number of experimentalists interested in using computational tools in their research. Such endeavor is nowadays facilitated by the abundance of low-cost computers capable of doing the job and highly intuitive software packages with a graphical user interface. This “democratization” of computational tools, although conceptually stimulating, carries the risk of leading astray those users lacking the basic know-how. GIGO, a popular acronym used as a word of caution against this pitfall, stands for Garbage In Garbage Out, thus warning the software user that computers cannot tell the difference between good and nonsensical input data. Nevertheless, except if blatant errors cause the calculation to abort, the program will always produce output, albeit one devoid of meaningful information. A minimum amount of homework is necessary if one wants to ask the right questions and be able to interpret the answers.

The importance of computation in chemistry has grown so large that computational chemistry became a field in its own right. According to IUPAC ^[6], “Computational chemistry is a discipline using mathematical methods for the calculation of molecular properties or for the simulation of molecular behavior. It also includes, *e.g.*, synthesis planning, database searching, combinatorial library manipulation.” It must be stated here that, although the history of computational chemistry has its roots on quantum mechanics, the branches have grown long and tall to embrace other methods of calculation which are radically different in their nature.

1.1 What is computational chemistry useful for?

Before delving into the many ways of doing computational chemistry it is time to ask what can it do for us? Computational chemistry tools permit the

determination of the electronic structure and potential energy surface of any given system. These may be thought of as the DNA of the system, a code from which all chemically relevant information may be retrieved, once you find the means to translate it. The first question computational chemists ask is usually: what is the lowest energy structure of this molecule? The various conformations corresponding to local and global energy minima, as well as to transition states, are found by performing geometry optimizations. Other commonly calculated observables include vibrational and rotational spectra, kinetic and thermodynamic properties, dipole moments and atomic charge distributions.

But what are all those calculations useful for, other than pure scientific interest? In short, understanding data at an atomistic level and using the derived knowledge to design experiments in more efficient and rational ways. This approach is now applied in a variety of fields such as genetics, analytical chemistry, drug design, materials research, nanotechnology, synthesis, biochemistry... and the list goes on ^[7]. The following lines provide a few examples of the vast applicability of computational chemistry.

1.1.1 Spectroscopy

One of the great successes of computational chemistry has been its alliance with spectroscopy. By using QM methods, one can predict with great accuracy the vibrations of a chemical system and actually observe the corresponding molecular movements using visualization software. Presently, any serious study involving FTIR or Raman spectroscopy compares experimental with calculated data. In most cases, the calculated vibrations are used to validate or elucidate the experimental spectra. Vibrational analysis is also important in geometry optimizations since it permits to distinguish between equilibrium and transition structures. QM also allows the study of chemical systems under the influence of an external magnetic field. Thus, chemical shifts may be computed and NMR spectra predicted. The study of excited states is possible as well and, consequently, the possibility for an electronic transition to take place may be estimated, a valuable tool for electronic spectroscopy.

1.1.2 Drug design

The pharmaceutical industry saves a great deal of money and time by screening and scoring potential pharmacological agents through computational means before testing those *in vivo*. This preliminary stage performed by computers is now so crucial that it was named *in silico*. An example shall illustrate the concept better: suppose that a new drug is being designed and one wishes it to bind to a known receptor X^[8]. A key requirement for a candidate drug is, other than exhibiting bioactivity of course, that it fits into receptor X's binding cavity. The first step is then to screen a database of potential drugs and select those whose active sites have the right size and shape to fit into the receptor's binding cavity. The top rated compounds are thus selected and, in a subsequent phase, the strength of their interaction with the receptor is evaluated. At this stage shape is no longer the determining factor but rather the strength of intermolecular interactions established between the ligand and the binding site. The stronger this interaction is the higher score the pharmacological compound will receive and only those with the highest scores will then be tested in animal models. Thus, large sums of money, laboratory work hours and animal lives can be saved, benefiting all. Some even believe that computational methods are the most reliable replacement for animal experimentation and are developing software tools to fasten the process.

1.1.3 Nanotechnology

Nanotechnology and computational chemistry are growing together, and for good reason^[9]. The behavior of matter changes dramatically when one goes from the bulk to the nano dimension and when designing materials at the nano-level a deep understanding of how and why such changes occur is crucial. When it comes, for example, to self-assembly, a process governed by intermolecular interactions, computational chemistry stands out as the most appropriate tool, capable of offering more meaningful answers than existing characterization techniques. Furthermore, the size of the systems interesting to nanotechnology is inherently small, posing no computer capacity problems for most calculations.

1.1.4 Proteins in solution

Last but not least comes the study of proteins in solution ^[10], in particular the riddle of protein folding. It is not difficult, after decoding the DNA, to predict the aminoacid sequences that form different proteins. The real puzzle is finding a way to

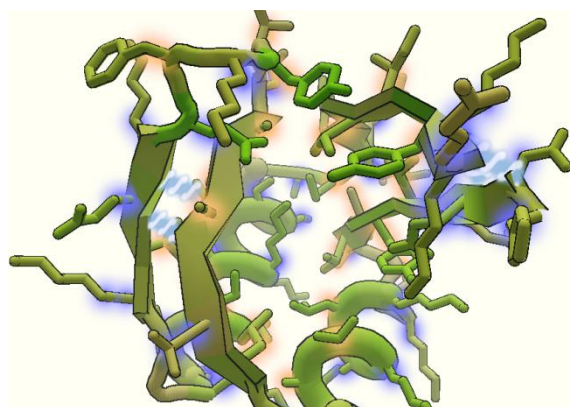


Fig. 1 3D representation of a protein. Taken from <http://fold.it/portal/info/science>

fold that long string of aminoacids into a 3D shape that minimizes the energy of the

macromolecule. The possible configurations are immense, yet a freshly synthesized protein “chooses” a correct folding pattern in tens of μ s. In order to simulate the entire folding pathway, even a supercomputer would take years! What is normally done is crystallizing the protein, obtaining the structural data through X-Ray diffraction and then “relaxing” its structure by simulating a water bath containing the protein. The lowest energy conformation thus found is much more likely to resemble the protein in solution than the crystallized structure. Currently, methods that do not require previous knowledge of the crystal structure, but solely of the aminoacid sequence, are being developed.

In summary, what advantages has computational chemistry to offer?

- Validation/interpretation of experimental results
- Better experimental design
- Money and time-saving (in some cases)
- Environmentally-friendly
- Determination of properties that can't be measured experimentally and/or lack atomistic detail
- Visualization of ultrafast dynamic processes
- Ability to study any system, even if it can not be synthesized with current technology

- A safe way to investigate toxic, mutagenic or otherwise dangerous compounds and/or conditions

- An introspective attitude that requires deep knowledge of chemical nature and is eager to question its own assumptions

1.2 The workings of Computational Chemistry

The planning of any study in computational chemistry must begin by choosing the appropriate method of calculation to be employed. In general, the more detailed and accurate the method, the more computer resources will be needed. This limitation sometimes leads to the sacrifice of accuracy in benefit of time-saving. For example biological systems, such as lipid bilayers, containing thousands of atoms, are so computationally demanding that the calculation method must necessarily be one of the cheapest: non-quantum, classical mechanics force field methods.

Molecular mechanics (MM), Molecular dynamics (MD) and Monte Carlo simulations all employ force fields which describe how the energy of a system changes as bond lengths, bond angles and torsional angles deviate from their equilibrium values, which are determined experimentally. In this approach, molecules are treated as balls on springs whose vibrations may be described as those of an harmonic oscillator. The electrons are not explicitly included in the calculations which, while saving time, impedes the simulation of bond breaking although Car-Parrinello Molecular Dynamics (CPMD) allows the study of electronic transitions to excited states.

While molecular mechanics enables the calculation of the energy of a molecule and to find its equilibrium structure through geometry optimization, molecular dynamics is able to describe how the system changes through time by assuming that atoms move according to Newtonian mechanics. Monte Carlo simulations achieve the same as MD through a different procedure. Instead of integrating equations of motion, random configurations are chosen and their energies measured. If a certain configuration has a higher energy than its predecessor, it is rejected, and the contrary if the energy is lower.

Quantum mechanical (QM) methods treat the electrons explicitly, that is, both the nuclear and electronic coordinates are calculated in each iteration. In fact, while force-field methods require the assumption that molecules are formed by atoms interconnected by covalent bonds, QM merely “sees” atomic nuclei floating in a sea of electrons. The available QM methods differ in the amount of approximations made. As more approximations are introduced the calculation time decreases, but so does accuracy and, consequently, agreement with experimental results. The most accurate methods are those termed *ab initio*, meaning “from the beginning” since these only require the electronic and nuclear coordinates, as well as the total charge and spin multiplicity, in order to perform the calculations. In contrast, semi-empirical methods employ parameters derived from experimental data, that is, apply correction factors which promote the convergence between calculated and experimental results for a set of molecular properties of extensively studied molecular systems. Naturally the *ab initio* methods are preferred whenever possible since they are independent from experiment, however in most cases it is not feasible to study systems with more than 100 atoms from first principles while semi-empirical methods are adequate for a few hundreds of atoms. Molecular mechanics is computationally cheaper than either of the former, being applicable to systems with thousands of atoms. A visual summary of the distinctive features of molecular mechanics, semi-empirical and *ab initio* methods is provided in Fig. 2.

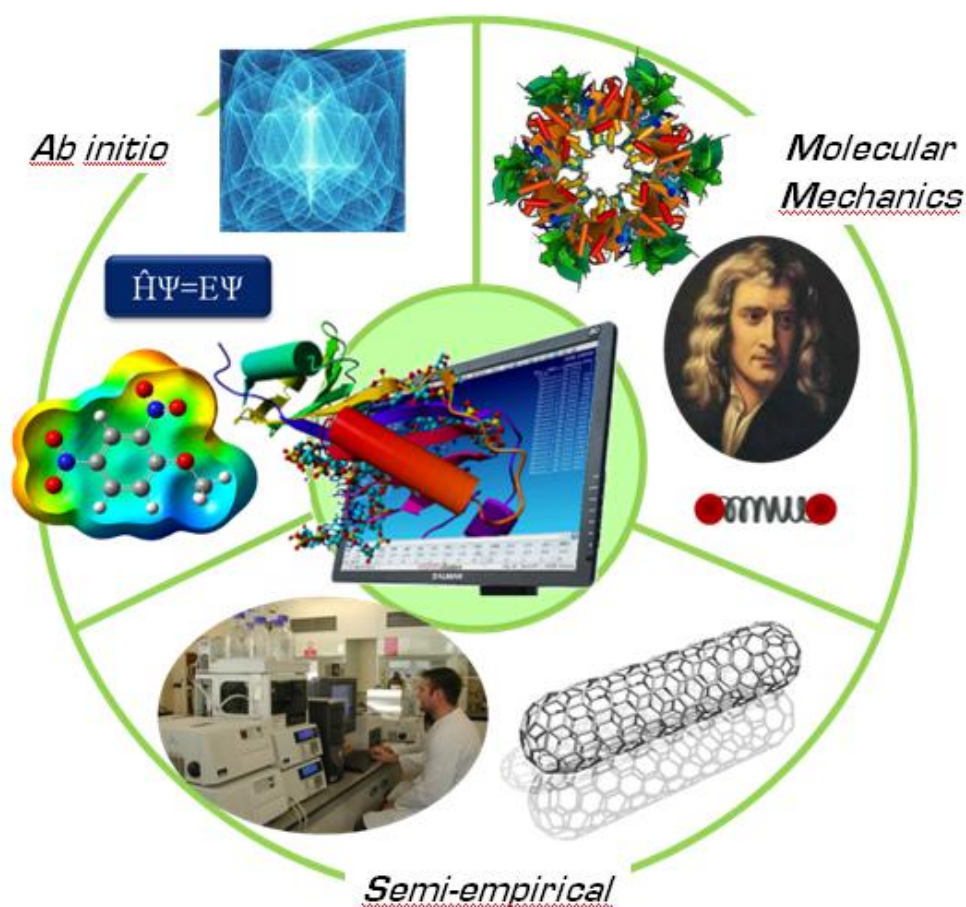


Fig. 2 Conceptual diagram representing the main characteristics of theoretical models. *Ab initio* methods, in most cases, are wavefunction-based and aim to provide approximate solutions to the Schrödinger equation; due to their high cost these methods are only applicable to small molecules. Semi-empirical methods employ parameters derived from experimental data or high-level predictions and are adequate for systems with up to a few hundreds of atoms. Molecular mechanics relies on classical physics and assumes atoms in molecules to behave as balls on springs; being the cheapest method, molecular mechanics allows the study of very large systems, with thousands of atoms.

It is also possible to combine quantum and molecular mechanics methods, a hybrid methodology which is particularly appealing for biological systems. When studying enzymes, for example, it is necessary to employ QM methods in order to assess bond breaking and formation at the active site. However, treating a whole enzyme plus the surrounding water molecules from first principles is not feasible at present.

QM/MM methods ^[11] offer an elegant solution by allowing the QM treatment of a small region located on the active site of the enzyme, where interesting events happen, while applying MM to describe the surrounding protein and solvent, thereby tremendously reducing the computational cost while still producing meaningful results.

Since the calculation methods used along the course of this project belong to the *ab-initio* category, the following section will attempt to describe how these work, starting by analyzing the equation that sits at the root of every one of them. The following discussion is, for the sake of brevity, largely incomplete and superficial. The interested reader should consult the following references for further information ^[12-17]

1.2.1 Foundations of ab-initio methods

1.2.1.1 The Schrödinger equation

Intrigued by the mystery of wave-particle duality, the Austrian physicist Erwin Schrödinger took refuge in the swiss alps, using pearls as earplugs, with his mind set on finding a mathematical description for the wave nature of electrons. And so wave-mechanics was born ^[3].

The fundamental postulate of wave mechanics is embodied in Schrödinger's time-independent equation

$$\hat{H} \Psi = E \Psi$$

which may be translated as “ a wavefunction Ψ exists for every chemical system so that when an Hamiltonian operator \hat{H} acts upon Ψ , the result is the product of Ψ and the energy of the system, E.”

An operator \hat{A} is a mathematical tool that, when applied to a function $f(x)$ gives a new function $g(x)$. In some exceptional cases, $g(x)$ is proportional to $f(x)$, that is $\hat{A}f(x) = a f(x)$ where a is a proportionality constant. $f(x)$ is thus an eigenfunction of \hat{A} , and the constant a is its eigenvalue. Likewise, Ψ is an eigenfunction which describes the quantum state of a particle as a function of space and time and E is the eigenvalue of Ψ when operated by \hat{H} . The Hamiltonian operator may be written as

$$\hat{H} = - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k} \nabla_k^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i < j} \frac{e^2}{r_{ij}} + \sum_{k < l} \frac{e^2 Z_k Z_l}{r_{kl}}$$

Where i and j run over all electrons, k and l run over nuclei, \hbar is Planck's constant divided by 2π , m_e is the mass of the electron, m_k is the mass of nucleus k , ∇^2 is the Laplacian operator, e is the charge of the electron, Z is the atomic number, and r_{ij} is the distance between particles i and j . The Laplacian operator may be written as

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$$

and allows the application of Schrödinger's equation over the 3 spatial dimensions. Thus, in order to solve the equation, one needs to find a function which, when differentiated twice with respect to the position vectors, returns to its original form. For example, exponential, sine and cosine functions exhibit this behavior.

The first and second terms of the Hamiltonian concern the kinetic energy of electrons and nuclei, respectively, while the remaining terms account for the potential energy arising from Coulombic interactions. The third term accounts for the attraction between electrons and nuclei, the fourth adds interelectronic repulsion and the last one corresponds to internuclear repulsion.

The most baffling aspect of Schrödinger's discovery is that his equation is not derived from any other but rather "inspired": an intuitive guess which proved to be accurate. But what is Ψ ? It is difficult to pin-point the physical meaning of Ψ since its units are complex numbers, but it represents the *probability amplitude* of finding a particle in any spatial point, at a given time. Christopher Cramer^[12] compares the wavefunction with "an oracle - when queried with questions by an operator, it returns

answers.” It may seem, so far, that the only answer the wavefunction is able to provide concerns the energy of the system but that applies only when the Hamiltonian operator is used. By employing different operators one calculates different properties. In principle, all information pertaining a particle may be retrieved from its wavefunction.

While Ψ is devoid of concrete physical meaning, at least at our current level of understanding, the modulus of its square, $|\Psi^2|$, has real units, thus being more amenable to interpretation, and represents the probability of finding a particle between x and $x+dx$ at time t . Hence, if $|\Psi^2|$ is integrated over all space the result must be 1, since the particle has to exist somewhere. This probabilistic view is in stark contrast with Bohr’s deterministic picture of electrons orbiting around the nucleus in well-defined trajectories. In quantum mechanics, orbits give way to orbitals, which may be thought of as probability clouds, that is, regions in space where the probability of finding the electron is high. For the sake of simplicity, orbitals are usually represented by isosurfaces, that is, solid contour maps whose spatial coordinates delimit a volume within which there is x probability (usually 90%) of finding the electron. Fig. 3a) displays 1s orbital isosurface for the hydrogen atom while Fig. 3b) shows the corresponding electron probability distribution, where a higher density of dots denotes a region with higher probability of finding the electron.

The size, shape and orientation of atomic orbitals are thus determined by the solutions to Schrödinger’s equation, that is, the wavefunctions. Since only discrete, quantized, energy states are allowed for the electron, only a correspondent number of wavefunctions are acceptable solutions to the Schrödinger equation. In order to be an acceptable solution, the wavefunction must approach zero when reaching infinity (since the probability of finding an electron infinitely distant from the nucleus is close, although not equal, to zero) and both the wavefunction and its first derivative must be continuous.

1.2.1.2 Solving the Schrödinger equation for the H atom

As said earlier, it is possible to find analytical solutions to Schrödinger's eigenvalue problem as long as the system at hand has only one nucleus and one electron. In this case, neither the internuclear nor the interelectronic repulsion terms exist in the Hamiltonian, which is left with the kinetic energy terms and the potential energy term arising from the attraction between the nucleus and the electron. It is mainly due to the interelectronic repulsion term, which accounts for the interactions of each electron with every one of the remainder, that the same analytical approach is not possible for polyelectronic systems. In the latter case, only approximate solutions can be obtained and these result from the product of one-electron wavefunctions. This method, named Hartree-Fock in homage to its developers, will be explained in greater detail later. Before delving into those matters, a few points regarding the analytical resolution of Schrödinger's equation need to be addressed.

The first step to solve this "eigen riddle" is to switch from Cartesian (x,y,z) to polar coordinates (r, θ, φ) where r is the radial distance between the nucleus and the electron, while θ and φ define the angular orientation of the latter. The use of polar coordinates simplifies the calculation since it takes advantage of the hydrogen atom's spherical symmetry, that is, since the electrostatic potential arising from the electron's attraction to the nucleus is spherical, the properties of the system will only depend on the electron's radial and angular motion with respect to the nucleus.

Furthermore, the radial and angular contributions may then be treated separately, so that instead of attempting to solve a complex equation which depends on three variables one may solve three simpler equations, each dependent on only one variable. The resolution of each of these equations gives rise to the three quantum numbers n (from r), l (from θ) and m (from φ), which define, respectively, the size, shape and orientation of the orbital.

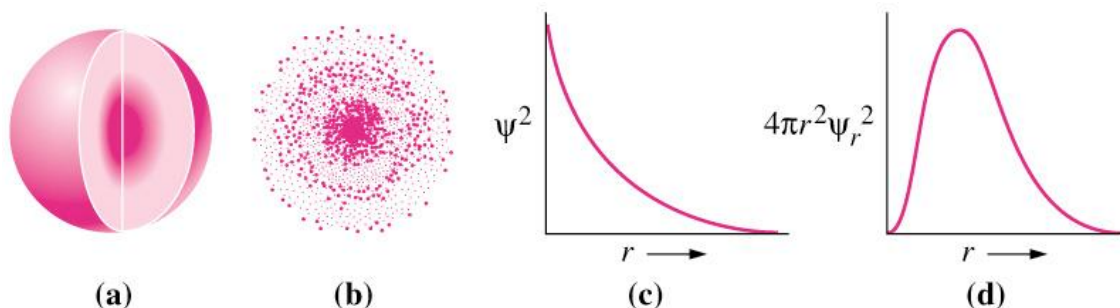


Fig. 3 Different forms of describing the hydrogenic 1s orbital are shown. (a) isosurface of the 1s orbital, where the darker shaded areas correspond to the regions where the probability of finding an electron is highest (b) conveys the same information as the isosurface (a) by using dots instead of a color gradient (c) radial probability distribution (d) radial probability density. Image taken from <http://www.kentchemistry.com/links/AtomicStructure/Schrödinger.htm>

The s orbitals have no angular momentum, that is, their corresponding wavefunctions only depend on the radial distance since the electron cloud is spherically symmetric. The radial wavefunctions for hydrogenic orbitals are mathematically represented by associated Laguerre polynomials. In the case of the 1s orbital, the radial wavefunction assumes the form

$$R_{1,0}(r) = 2 \left(\frac{Z}{a_0} \right)^{3/2} e^{-\frac{Zr}{a_0}}, \text{ where } r \text{ stands for the distance between}$$

electron and nucleus, Z is the nuclear charge and a_0 the Bohr radius.

In Fig. 3 c) the radial probability function for the 1s orbital is shown. It can be observed that the highest probability of finding the electron is near the nucleus and such probability decays exponentially.

In p, d and f orbitals the electron tries to avoid the nucleus and the probability of finding it at $r=0$ is also 0. Such happens because when electrons inhabit an orbital with $l > 0$, a centrifugal force acts on them, pushing them away from the nucleus, near which the repulsive effect surpasses the coulombic attraction. S orbitals, having no angular momentum, are subject to the coulombic attraction only, so that their preferred site is close to the nucleus.

Fig. 3 d) shows the probability density of finding an electron enclosed within a volume defined by two concentric spheres of radii r and $r+\delta r$. Given the behavior of the radial probability distribution it may seem odd that the probability density is 0 near the nucleus. This arises because the squared wavefunction is multiplied by the sphere

volume which, at the nucleus, is 0. But what does it really mean, in a physical sense? The radial probability distribution provides information on which points in space does the electron pass by more often. But the probability density measures the probability of finding an electron in an infinitesimal volume with a certain radius or, in other words, the most probable distance of the electron from the nucleus. Remarkably, the most probable radial distance as predicted by wave mechanics coincides with the atomic radius Bohr had previously calculated.

Schrödinger tested the validity of his equation by calculating the energy levels for the hydrogen atom and verifying a total agreement with experimental spectral lines. This initial success stirred so much interest from the scientific community that soon after its publication approximate methods were devised for the treatment of bigger atoms and such development continues to this day, with no signs of slowing down.

1.2.2 Calculation methods for polyelectronic systems

In the impossibility of finding analytical solutions to the Schrödinger equation for polyelectronic systems, one is reduced to find approximate solutions through numerical methods, that is, use a simpler function that mimics the behavior of the real wavefunction as closely as the available computational power and system size allow. The various *ab initio* calculation methods employ different sets of approximations but there is at least one that lays at the foundation of all: the Born-Oppenheimer approximation.

1.2.2.1 The Born-Oppenheimer approximation

The fundamental idea of this approximation is that the electronic and nuclear contributions to the wave function may be separated. Protons are 1836 times heavier and consequently much slower than electrons. From the viewpoint of a fast travelling

electron, the nucleus hardly moves. It is thus assumed that nuclear motion is essentially dependent on electronic motion so that nuclear coordinates remain fixed in each calculation step and only the motion of electrons is explicitly taken into account. After determining the electron distribution, the nuclei adjust to the electronic cloud by slightly changing their positions.

The Born-Oppenheimer approximation allows the neglect of nuclear kinetic energy, the substitution of the nuclear-nuclear repulsive term by a constant value for a given geometry and the elimination of correlation in the electron-nuclear attraction contribution.

The most important aspect of Born-Oppenheimer's approximation is a conceptual one. If we attempted to describe a molecular system using a wavefunction which is dependent on both nuclear and electronic positions, as well as spin states, all the familiar concepts of bonds, rotations, vibrations and potential energy surfaces would have to be redefined in quantum mechanical terms of probability clouds. Although more truthful, the pure quantum mechanical model is still far too impractical and esoteric for our classically trained brains to use it in an everyday basis. As for the numerical utility, this approximation does indeed simplify the calculation while introducing a minor error, but still the much more troublesome many-electron problem remains.

The motion of an electron is affected by the instantaneous repulsion caused by each and every one of the remainder electrons, a phenomenon named correlation. The calculation of the other terms in the Hamiltonian is straightforward, but the interelectronic repulsion term requires accounting for each interacting electron pair. Hence, every electron added to the calculation brings along the burden of 3 additional degrees of freedom, rendering the analytical resolution not merely impractical but, so far, impossible. However incorrectly, further simplification requires the assumption that electrons behave independently of each other.

1.2.2.2 The independent electron approximation

If we assume that the probability of finding an electron in the orbital i at the position r_i is independent of the position of the other electrons, then the electronic wavefunction may be written as

$$\Psi(r_1, r_2, \dots, r_n) = \prod_{i=1}^n \psi_i(r_i)$$

Accordingly, the electronic Hamiltonian (and consequently the total electronic energy) results from the summation of the one-electron Hamiltonians and the energies of the individually occupied molecular orbitals, respectively. The product of one-electron wavefunctions written above is known as the Hartree product. Within this model, each electron feels the attraction exerted by the nucleus and an effective uniform potential dependent on the surrounding electronic density, an attempt to take into account interelectronic repulsion in an average way. However elegant it may be, this solution does not respect the Pauli Exclusion Principle, since it fails to account for spin. Electrons have half-integer spin, hence according to this principle the wavefunction must be antisymmetric, that is, if two electrons are interchanged the the wavefunction changes sign. By expressing the wavefunction as a Slater determinant of spin-orbital products, the Pauli Exclusion principle is respected.

1.2.2.3 The LCAO approach

The Linear Combination of Atomic Orbitals (LCAO) approach states that, just as many-electron wavefunctions may be obtained as the product of one-electron wavefunctions, so may the molecular wavefunctions Ψ be constructed as linear combinations of atomic orbitals ϕ centered on individual atoms, where the coefficients c_i indicate how much an individual one-electron wavefunction contributes to the final wavefunction.

$$\Psi = \sum_i (c_i \phi_i)$$

Hence, if we wish to determine the molecular wavefunctions for the H_2 molecule, at least two atomic 1s wavefunctions must be combined yielding the bonding and anti-bonding molecular wavefunctions. Although, from the chemical point of view, adding two 1s orbitals seems to suffice, we must remember that what we are actually doing is adding up mathematical functions in order to obtain a new function which closely reproduces chemical behavior, and the more functions are combined the better

our approximate wavefunction will fit the real wavefunction. This matter will be further addressed in the “Basis sets” section.

At this point dramatic approximations have been introduced in the calculation scheme, so much so that the exact and tidy Schrödinger equation starts to seem more and more like a mirage. How can we be sure that the guess wavefunctions constructed through the LCAO approach have any physical meaning, that is, resemble the real wavefunctions enough to correctly describe chemical behavior? The next section shall bring the answer.

1.2.2.4 The Variational Principle

The variational principle is of great importance in computational chemistry. It fundamentally states that any approximate wavefunction resulting from a LCAO is always higher in energy than the exact ground state of the system.

$$E = \frac{\int \Psi^* \hat{H} \Psi. d\tau}{\int \Psi^* \Psi. d\tau} = \frac{\int (\sum_i c_i \phi_i) \hat{H} (\sum_j c_j \phi_j) dr}{\int (\sum_i c_i \phi_i) (\sum_j c_j \phi_j) dr} \geq E_{exact}$$

Where $\Psi^*\Psi$ is the product of the trial wavefunction and its complex conjugate, \hat{H} is the one-electron Hamiltonian and $d\tau$ integrates the former over all space.

Hence, there is a simple criteria to judge the quality of a trial wavefunction: the lower its energy, the better. A trial wavefunction may be constructed in any way we please as long as its energy can be minimized. The energy of a HF wavefunction depends on the coefficients employed in the LCAO. Consequently, minimizing the wavefunction energy requires finding an optimal set of coefficients. From calculus we know that the minima of any function is found where its first derivative is zero. In this case we write

$$\frac{\delta E}{\delta a_k} = 0$$

The various minima of E correspond to the allowed energy levels of the trial wavefunction. Thus, for the ground state, the global energy minimum is of interest.

Once this energy is quantified, its associated coefficients can be retrieved and, with these, an optimal wavefunction is built.

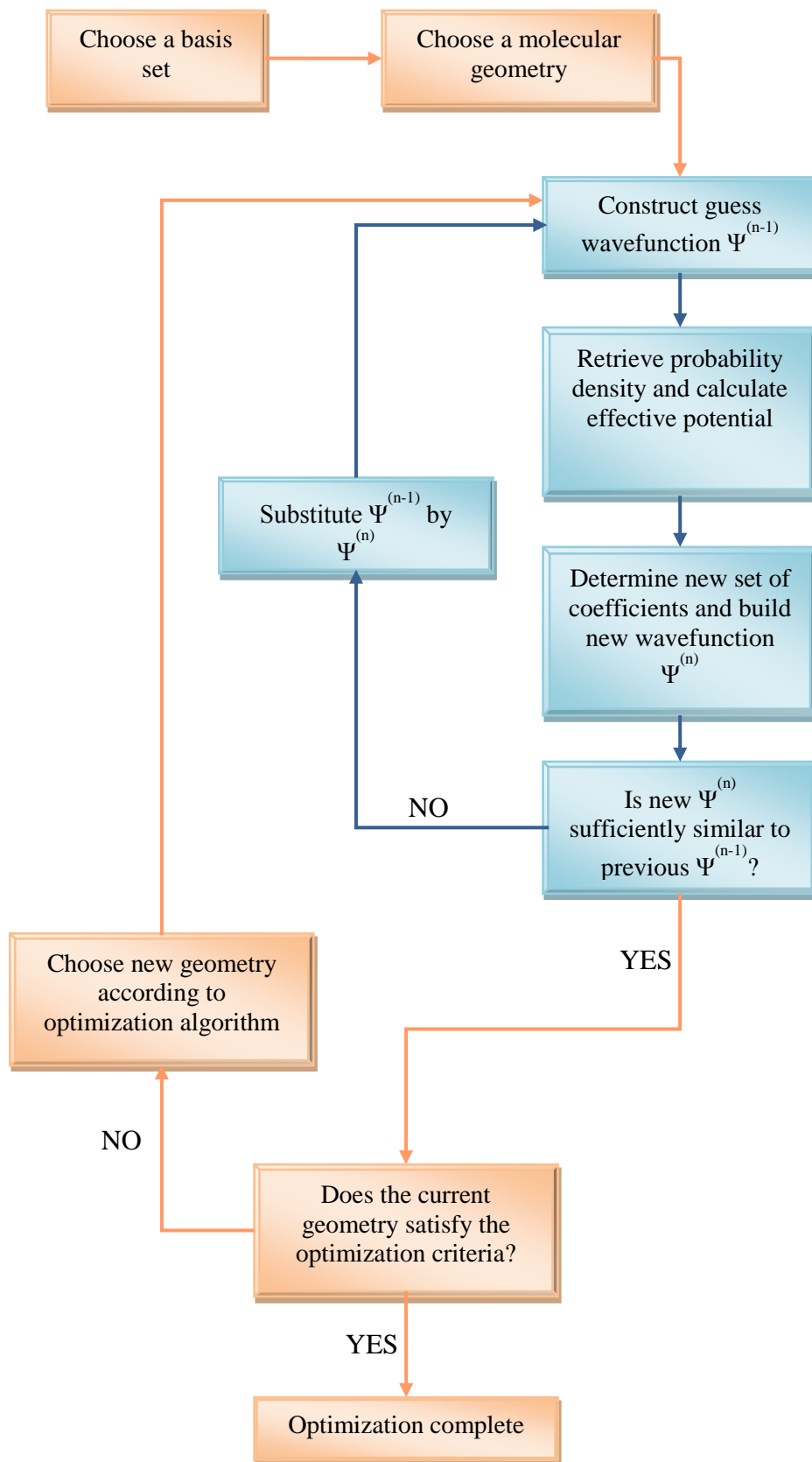
1.2.2.5 The HF-Self Consistent Field procedure

The HF method is inspired in the independent electron approximation and, as such, the interelectronic term in the Hamiltonian is substituted by an average interaction potential – effective potential - which depends on the local probability density. Thus, each electron “feels” a uniform electric field produced by all the other electrons. The obvious problem herein is that we need the probability density distribution in order to calculate the effective potential and then determine the orbital coefficients. But without knowing the coefficients *a priori*, the probability density can not be retrieved. Hence an iterative process is followed, starting with the construction of a guess wavefunction, from which the probability density distribution is calculated, now allowing the determination of the effective potential. New orbital coefficients which minimize the wavefunction energy are generated, a new wavefunction constructed and the process repeats itself until there is no significant change in the energies and density distributions of two consecutive wavefunctions.

At this point we say the calculation has converged. Scheme 1 displays a flow chart outlining the HF-SCF cycle (in blue) and the geometry optimization process (in salmon).

1.2.3 Basis sets

According to the HF-SCF method, the guess wavefunction may be constructed with any set of appropriate mathematical functions, also called basis functions.



Scheme 1 Flow-chart depicting the Self-Consistent Field cycle (in blue) and the geometry optimization procedure (in salmon)

Which characteristics should a mathematical function have in order to be an eligible basis function?

- It must behave in a way that has chemical meaning, that is, should have large amplitude in regions of space where the probability of finding the electron is high and vice-versa;

- Must allow for a computationally efficient calculation of the integrals appearing in HF equations.

In a world with unlimited computer power, only the first condition listed above would be observed and hydrogenic wavefunctions would be used since, for one electron atoms, these are exact. In the real world, a small degree of accuracy must be sacrificed for a great increase in speed. Thus hydrogenic wavefunctions, which are very computationally demanding, were put aside in favour of Slater-Type orbitals, whose behavior closely resembles that of the former while being much easier to compute.

Slater-Type Orbitals have the following form:

$$\phi^{STO} \alpha, n, l, m(r, \theta, \varphi) = NY_{lm}(\theta, \varphi)r^{n-1}e^{-\alpha r}$$

Where N is a normalization constant, Y_{lm} are the spherical harmonics which determine the shape and orientation of the STO, depending on the quantum numbers l and m , α is an exponent that depends on the atomic number (since it defines the orbital size) and r is the radial distance.

Further improvement in efficiency may be achieved by using Gaussian-Type Orbitals (GTOs), introduced by Boys as cheaper alternatives to STOs. GTOs can be written as:

$$\phi^{GTO} \zeta, n_x, l_y, m_z(x, y, z) = Nx^{l_x}y^{l_y}z^{l_z}e^{-\zeta r^2}$$

Here ζ determines the size of the orbital (as α in STOs) while the exponents l_x , l_y and l_z determine the type (s, p, d...). GTOs are easier to integrate than STOs, since their radial decay is dictated by e^{-r^2} and the integrals formed from these functions have

relatively simple analytical solutions, while those formed from STOs, which depend on e^{-r} require a much more fastidious numerical resolution.

Unfortunately, GTOs do not mimic hydrogenic radial wavefunctions as well as STOs, especially near and far from the nucleus. The accuracy offered by STOs and efficiency characteristic of GTOs may be simultaneously achieved by using linear combinations of GTOs. The more GTOs are linearly combined, the better the resulting basis function mimics the STO behavior while still allowing the analytical resolution of the associated integrals, which nevertheless become increasingly more complicated as more GTOs are added. An ideal balance between computational cost and accuracy is found when 3 Gaussians, also called primitives, are linearly combined giving rise to a contracted basis function. This constitutes the minimal basis set, STO-3G, which stands for “a Slater-Type Orbital approximated by 3 Gaussians”. Fig. 4 shows how the radial behavior of STO-3G closely resembles that of an STO.

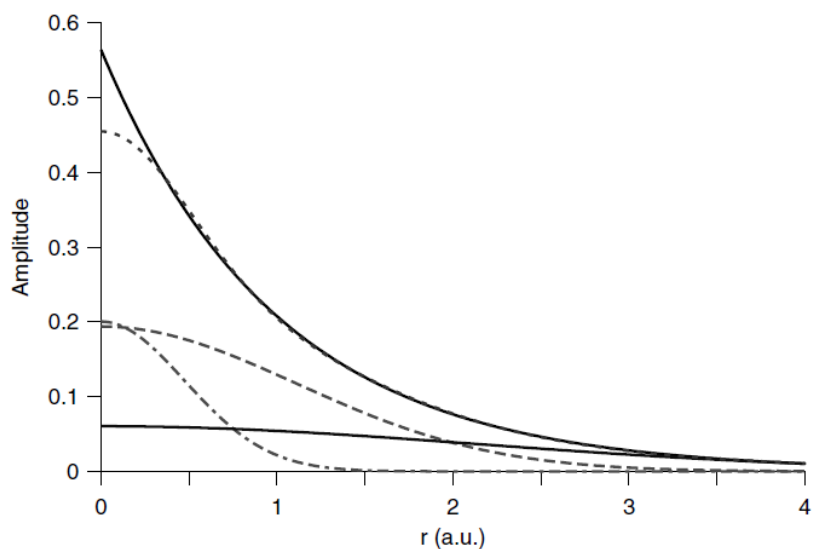


Fig. 4 Radial behaviour of different basis functions for the hydrogenic 1s orbital in atom-centered coordinates. The bold solid line is a STO while the dashed lines are contracted GTOs. The GTO found nearest to the STO is STO-3G (from Cramer).

1.2.3.1 Single, double and triple zeta basis sets

STO-3G is said to be minimal since it is the cheapest basis set able to provide meaningful results. It is also classified as single-zeta, since there is only one basis function describing each type of orbital, from core to valence. For example, when describing Li at least 5 orbitals are needed: 1s, 2s, 2p_x, 2p_y, 2p_z. If a STO-3G basis set is used, then 5 contracted Gaussians will be used in the LCAO, one for each of the atomic orbitals in Li.

Double and triple zeta basis sets represent the next step in accuracy. These basis sets use, respectively, two and three contracted Gaussian functions, with different zeta values, to describe each atomic orbital. The use of GTOs with different ζ s is analogous to combining two atomic orbitals of different sizes, thus allowing more flexibility for the description of the molecular orbital.

1.2.3.2 Split-valence basis sets

Double and triple zeta basis sets are especially relevant within the split-valence framework. The rationale behind this approach is that when two atoms come together to form a covalent bond, their core orbitals are not very much affected, as opposed to the valence orbitals which undergo major changes. Hence, there is no need to describe core orbitals with the same thoroughness and flexibility as valence orbitals require. Split-valence basis sets use a single contracted basis function to describe the core orbitals while more than one contracted GTO is used to describe each valence orbital. For example, the 6-311G basis set describes core orbitals using solely one contracted GTO, resulting from the linear combination of 6 primitives while the valence orbitals are described using three different GTOs (since there are three numbers after the hifen), one of which is the contraction of 3 primitives while the remainder arise from a single primitive.

1.2.3.3 Polarized basis sets

Even split-valence basis sets fail to account for polarization effects. A 6-311G basis would describe an oxygen atom in a given molecule by including in the LCAO only the basis functions representing the 1s, 2s and 2p orbitals. However, when there is an unequal electron distribution between two atoms due to polarization effects, these basis sets lack the ability to properly describe it. This can be achieved by adding to the basis set orbitals with a higher angular momentum than the occupied orbitals. Therefore, *p* functions are added to hydrogen atoms, *d* functions to second-row atoms, *f* functions to transition metals and so on. Polarized basis sets are identified by the use of an asterisk * or the letter d, p, f in parentheses. For instance, the 6-311G* (or 6-311G(d)) basis set employs polarization functions on non-hydrogen atoms while 6-311G** (or 6-311G(d,p)) also includes polarization functions on hydrogen atoms.

1.2.3.4 Diffuse basis sets

Normally electron density is significant only at distances which are relatively close to the nucleus. Basis functions plot the electron distribution from the nucleus to infinity. Clearly, it is not advantageous to extend the calculation to regions far from the nucleus, so a cut-off in the radial distance is used.

In certain cases, such as anions and excited states, the former approach is not able to reproduce chemical reality, since the probability of finding the electron far from the nucleus is considerably higher. In order to account for this effect, diffuse basis functions are added to the basis set. These functions have small ζ exponents which causes them to decay slowly with distance, thus still retaining a significant electron density at the cut-off distance. Diffuse functions are indicated by a plus (+) sign in the basis set nomenclature. In an analogy to polarized functions, a single + denotes the addition of diffuse orbitals for non-hydrogen atoms and ++ for all atoms.

As a final note on basis set nomenclature, the notation HF/6-311G++(d,p)//HF/6-311G+(d) means that a geometry optimization was performed with the smaller basis set and the energy (or other observable) of the optimized geometry was measured with the larger basis set. Although time-saving, this strategy is unreliable,

since the minimum energy structures may not be exactly the same for different basis sets.

1.2.3.5 Approaching the HF limit

The limit of accuracy a HF calculation is able to provide may be approached by using larger and larger basis sets. As computational power increases, some of the smaller basis sets become obsolete (STO-3G is an example) while larger ones grow increasingly common. Examples of the latter are the aug-cc-pVXZ (X=D,T,Q) basis sets of Dunning and co-workers. Here “aug” stands for augmented, referring to the use of diffuse functions on all atoms, “cc” is correlation-consistent, meaning that the exponents and contraction coefficients were optimized for calculation methods which account for electron correlation and pVXZ stands for polarized valence double-, triple- or quadrupole-zeta.

Large basis sets such as those of the Dunning family were created with the aim of extrapolating their results to the HF-limit using a curve fitted over the values of some property computed with increasingly larger basis sets.

1.2.4 Weaknesses of HF theory

The central weakness of HF theory stems from its roots: even at the HF-limit the calculated energy will always be higher than the exact energy because HF fails to account for the total energetic contribution of electron correlation. As a result, bond lengths estimated at the HF level are shorter than reality. Consequently, vibrational frequencies are ~10% larger than experimental figures.

By obeying the Pauli Exclusion Principle and thus forbidding two electrons of the same spin to occupy the same point in space, HF does account for exchange interaction, a component of the correlation energy. However, the instantaneous effects each electron has on the motion of all others were neglected by the independent electron approximation.

The explicit treatment of electron correlation requires the use of post-HF methods. Among these are Configuration Interaction (CI), Møller-Plesset Perturbation Theory (MPPT) and the Coupled Cluster (CC) method.

Another hurdle HF presents is that depending on the wavefunction to provide information about the system requires previously building that same function, which depends on $4N$ coordinates (N being the number of electrons). It would be simpler to use an observable depending on only 3 coordinates and able to provide the same information as the wavefunction. The electron density fits the former criteria and this rationale is the basis of Density Functional Theory (DFT). The obvious advantage of DFT is that while wavefunctions become increasingly more complex for larger system sizes, the electron density will always depend on 3 coordinates. Thus DFT is cheaper than HF and consequently applicable to large systems.

For the sake of brevity and relevance to the present work, only MPPT and DFT shall be addressed, and even those, in a rather brief manner.

1.2.5 Density Functional Theory

The great departure between HF and DFT is that the first optimizes wavefunctions while the latter optimizes electron probability densities. The framework of both theories is otherwise similar and proceeds through the iterative SCF cycle. HF and DFT also differ in the way they treat electron exchange and correlation. While HF employs an effective potential to approximate correlation effects and calculates exchange exactly, DFT uses an exchange-correlation functional. A functional is a function of another function, and the exchange-correlation energy (E_{XC}) is a function of the probability density (ρ). There are many available DFT functionals and three main functional “architectures”: LDA, GGA and Hybrid functionals.

The Local Density Approximation (LDA) assumes that the exchange-correlation energy is a function of the local value of ρ . It further states that the E_{XC} density of a molecular system, at some point in space where the probability density is ρ , equals the E_{XC} of a uniform electron-gas with the same value of ρ . Even for a uniform electron gas, the correlation component of E_{XC} does not have analytical solution, hence

LDA functionals are approximate expressions designed to reproduce a large set of theoretical and experimental data. Because of this parametrization, DFT is categorized by some as a semi-empirical method. Still, LDA is not a good approximation since it assumes the local density to be uniform which, in a molecule, is not true.

The Generalized Gradient Approximation (GGA) takes one step further by adding to LDA a gradient-correction term which causes the E_{XC} to depend both on the local density and its gradient, that is, its first derivative.

The most popular GGA exchange energy functional is the one developed by Becke, usually abbreviated “B”, followed by that of Perdew, Burke and Wang, PW. Among correlation functionals, those finding more widespread use are P86 (Perdew 1986), PW91 (Perdew, Wang 1992) and LYP (Lee, Young and Parr 1988). The chosen combination of exchange and correlation functionals is indicated in the literature by joining the respective abbreviations, as in BLYP.

Hybrid density functionals result from merging DFT and HF methods in order to promote the cancelation of errors. For example, barrier heights to chemical reactions are underestimated by DFT calculations using GGA functionals and overestimated by HF methods. These differences arise due to the different way of treating exchange and correlation. Hence, hybrid functionals include an exchange energy term which is calculated exactly, using HF, and added to a LDA or GGA functional. The extent to which the HF term contributes to the final E_{XC} is determined using a set of parameters, which are optimized to promote the best fit between calculated and experimental data (alternatively, the reference values may be those calculated with high-level methods such as CCSD(T)). Hybrid functionals usually perform better than gradient-corrected or LDA functionals.

The total number of available functionals, arising from all possible combinations, is immense. Finding the most suitable one requires previous knowledge of its performance for the system of interest. B3LYP, the hybrid version of BLYP, performs quite decently in a wide range of systems and is presently the most popular functional among the three categories. However, it is not universally suitable. For example, when studying systems where London dispersion interactions play an

important role, B3LYP is a poor choice since it fails to account for them. In these cases, specific functionals with dispersion corrections should be used instead.

DFT presently has major importance in computational chemistry. By offering an attractive combination of speed (faster than HF) and accuracy (sometimes as accurate as MP2), DFT permits studying large systems on a quantum level and is the method of choice whenever quantum calculations are performed on biomolecules.

1.2.6 Møller-Plesset Perturbation Theory

Both HF and DFT methods follow a variational approach to find the correct description of the system. An alternative to variational methods is offered by perturbation theory. The latter states that the real Hamiltonian may be approximated through a simpler Hamiltonian, for which the solution to Schrödinger's equation is known, plus a perturbing factor. In Møller-Plesset theory the electron-electron repulsion is treated as a perturbation on the one-electron Hamiltonian, formally

$$H = H^{(0)} + \lambda V$$

where $H^{(0)}$ is the one-electron Hamiltonian, V is the perturbation operator and λ is a constant assuming values between 0 and 1 which determines the amount of perturbation induced. The goal here consists in determining Ψ using the simple $H^{(0)}$. When the perturbation is small ($\lambda \ll 1$), Ψ may be written as a power series

$$\Psi = \Psi^{(0)} + \lambda \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \dots$$

where $\Psi^{(0)}$ is the uncorrected wavefunction. By adding the first order correction, $\Psi^{(1)}$, the Hartree-Fock solution is obtained. Applying higher-order corrections yields wavefunctions which incorporate electron correlation with increasing accuracy, but the complexity of calculations also grows with each correction order. The most usual MP methods are MP2, a second-order correction which is not absurdly expensive and provides reasonably accurate results, and MP4, the fourth order correction, which represents a significant step up in accuracy but still finds less applicability than MP2 due to its prohibitive cost even for medium systems.

1.3 Intermolecular Interactions and their importance

Chemistry, in the last century, was primarily concerned with covalent interactions, given that their formation and breaking determine how chemical reactions occur. In the past few decades, chemists realized that many important phenomena can only be properly rationalized if intermolecular interactions, also deemed noncovalent interactions, are accounted for^[18]. The broad term “intermolecular interactions” includes hydrogen bonding, π - π stacking, electrostatic and van der Waals interactions.

Life as we know it would not be possible without the concerted action of intermolecular interactions. Liquids, for example, would not exist without them. Biological systems rely largely on the “intelligent” design of intermolecular interaction networks, in which hydrogen bonds and π - π stacking play especially important roles^[19]. Base-pairing in DNA, protein folding, receptor-ligand binding and membrane permeation are but a few in a myriad of processes governed by intermolecular interactions. As mentioned earlier, drug design relies heavily on the optimization of noncovalent interactions between the pharmacological compound and its target receptor. Understanding the nature of the interactions at play is crucial for successful planning. For example, one may take advantage of the highly directional nature of hydrogen bonds, which become weaker as their geometry deviates from linearity, in order to design highly selective pharmacophores.

In the emergent field of nanotechnology noncovalent interactions also play a pivotal role, especially in self-assembly, a process during which molecular sub-units “recognize” one another and come into contact in order to maximize intermolecular interactions, thus building supramolecular structures. This strategy is especially useful for the mass production of molecular electronic devices^[20].

1.3.1 The challenge of modeling intermolecular interactions

Molecular mechanics offers a fast although limited prescription to describe intermolecular interactions: empirical potential energy functions, such as the Lennard-Jones potential. One of the many limitations of this approach is that atomic charges are assigned at the beginning of the calculation and remain constant. Thus, charge-transfer interactions cannot be properly studied without *ab initio* methods, a task which still faces some challenges^[21-22].

Intermolecular interactions are one or two orders of magnitude weaker than covalent bonds thus the correct quantification of the former requires the use of methods of higher accuracy and diffuse basis sets. Accurate results (with error < 1 kcal/mol) can only be provided by coupled-cluster methods, such as CCSD(T), with very large basis sets (or even extrapolated to the CBS limit) . Presently, such calculations are only applicable to systems of medium dimension (max. 30-50 atoms). The majority of interesting systems, such as those referred in the previous section, largely exceed this limit. Hence, cheaper methods must be employed in their study. MP2 has been a very popular choice since it is not outrageously expensive while still accounting for electron correlation, although it overestimates London dispersion. Still, MP2 is prohibitive for systems with more than 100 atoms. DFT is obviously alluring for this purpose, and shows a satisfactory performance in the description of hydrogen bonds and charge transfer. The Achilles heel of most “cheap” methods lies in their inability to describe dispersion interactions, which are predominant in π - π stacking and hydrophobic interactions. Dispersion interactions arise from the attractive component of electron correlation. All the DFT functionals described so far calculate an approximate electron correlation energy based on the local properties of the density. However, the very nature of dispersion interactions require a non-local treatment of correlation. The simplest way of minimizing this fault is to apply an empirical dispersion terms to common functionals. This method, named DFT-D, performs better than MP2 when dispersion interactions are dominant. Another approach is to parametrize functionals in way that

accounts for dispersion interactions, as is the case with the M05 and M06 series, developed by Zhao and Truhlar.^[23-24]

Another big puzzle for computational chemists is how to account for solvation effects while modeling intermolecular interactions. Clearly, the solvent medium is radically different from the gas-phase and a host-guest interaction may be quite favorable in vacuum but not as favorable as their individual interactions with the solvent which would, were this the case, keep host and guest apart. Quantifying the interaction energy for a system in solution requires estimating solvation free energies. This may be accomplished by explicitly including the solvent, the accurate-but-expensive approach or, including it implicitly by considering the solvent to be a continuum with a characteristic dielectric constant.

Still regarding the interaction energy, an additional problem arises due to incomplete basis sets – the basis set superposition error (BSSE)^[25]. The interaction energy between two systems A and B may be calculated as the energy of the AB complex minus the individual energies of A and B. The error arises since the fragment A in the AB dimer uses its own basis functions along with those of B in order to describe its molecular orbitals, and the same happens for B. The energy of the dimer is thus artificially lowered with respect to the individual monomers, which had a smaller set of available basis functions. Two strategies are possible to correct for BSSE:

- Using larger basis sets significantly reduces this error, which vanishes at the basis set limit. Computational cost is a downside.

- The counterpoise correction (CP), proposed by Boys and Bernardi^[26], attempts to mitigate BSSE by calculating the energies of the monomers using the total set of basis functions available to the dimer.

The total interaction energy may be decomposed into its components^[27] in order to determine the nature of the intermolecular interactions at play. Sometimes, this task is not straightforward since there are two or more components which contribute similarly to the final energy. The nature of the hydrogen bond, for example, is still a subject for debate.

1.3.2 The elusive nature of the hydrogen bond

The current definition of “hydrogen bond”, as recommended by the IUPAC^[28], is “an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation.” An hydrogen bond is depicted as X-H...Y-Z where the dots indicate the bond. X is the hydrogen bond donor and Y the acceptor. A few criteria must be met before evidence of bond formation is confirmed, among them:

- The linear bond angle corresponds to the maximum strength and shortest bond length
- The hydrogen bond must be a sum of electrostatic, charge-transfer and dispersion interactions.

Contrary to the contemporary view, the hydrogen bond was once regarded as entirely electrostatic in nature. It is so, but only at relatively large intermolecular distances ($> 2,5 \text{ \AA}$). As the hydrogen atom approaches the acceptor, charge transfer and dispersion interactions become stronger, resulting in an overall stronger hydrogen bond^[29-30]. However, the true nature of the hydrogen bond is still a matter of debate among scholars.

The vastness of our ignorance concerning the hydrogen bond becomes clear once we realize how little is known about the most ubiquitous hydrogen-bonded network on earth: water.

Water, contrary to other compounds with similar molecular weight (H_2S , CO , $\text{CH}_4\dots$), is liquid at room temperature. Such anomaly arises due to the strong nature of hydrogen bonds, which hold water molecules tightly together.

In addition, the behavior of water is radically divergent from other liquids: given the low molecular weight of the water molecule the density of liquid water is higher than other liquids in the same weight range; ice is less dense than liquid water;

water diffuses more easily when under high pressure; at very low temperatures, the heat capacity increases dramatically; when in contact with a hydrophobic solute, water forms a cage-like structure (clathrate hydrates) on the solute surface in a geometry that allows for the maximum number of hydrogen bonds to be established.

Due to its mysterious behavior, the structure of water is yet to be completely understood and different models have been proposed for its rationalization ^[31]. While some assume that there are constant bond breaking and formation events between water molecules (mixture models) others assume that the hydrogen bond network is continuous and hydrogen bonds are never broken but rather distorted from their optimum geometries ^[32].

Tapping into the elusive nature of water is one of the motivations for the study of water clusters ^[33]. A water cluster is an aggregate of two or more water molecules. In the limit, bulk water is simply a gigantic cluster. Thus, scientists hope smaller clusters will provide some cues about the structure of the bulk ^[34].

Numerous studies on water clusters, in some cases containing up to 280 water molecules but most of them concerning small clusters, have been performed ^[32, 35-37]. Thus, for $(\text{H}_2\text{O})_n$ with $n < 10$ the minimum energy structures have been determined, binding energies calculated and vibrational analysis performed, at various levels of theory and different basis sets. The minimum energy structures for the first 10 clusters, optimized at the MP2/6-31G(d) level ^[38], are shown in Fig. 5.

Up to the pentamer, the cyclic conformation is preferred and the water molecules are arranged in the (D-A) pattern, that is, each of them is simultaneously a donor and an acceptor. The hexamer is a special case for which several low-lying isomers, with small energy differences, have been identified. The hexamer marks the transition from 2D to 3D clusters, and the latter dominate for $n > 6$.

The (D-A) pattern of hydrogen bonds is not the only possibility, since each water molecule can be a double donor, a double acceptor or different combinations of these. The D-A configuration is advantageous because it maximizes cooperativity effects ^[39].

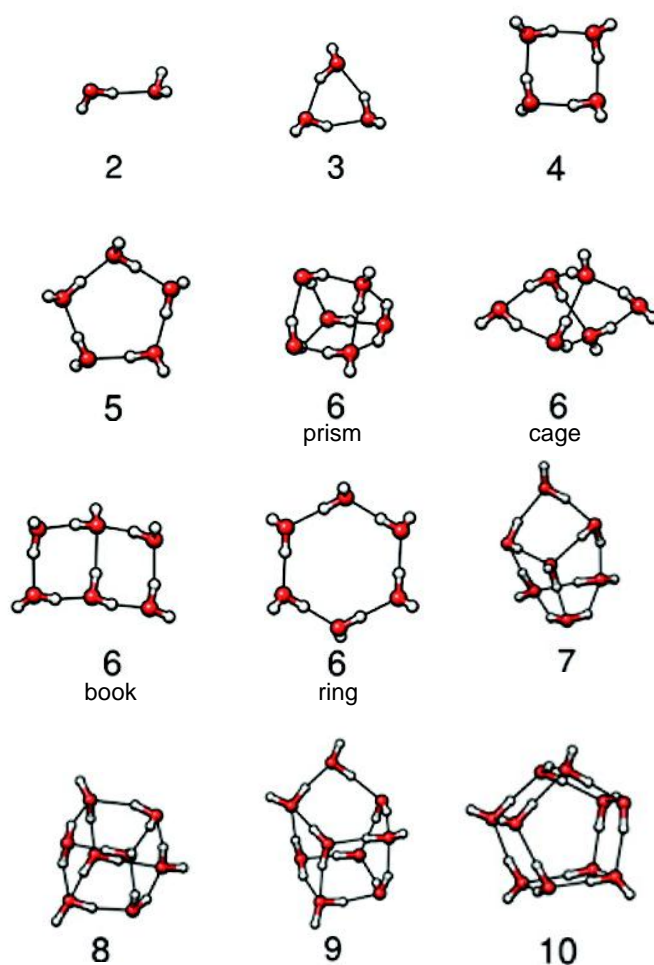


Fig. 5 Minimum energy structures of small $(\text{H}_2\text{O})_n$ clusters with $n=1-10$, optimized at the MP2/6-31G(d) level. Four low-lying isomers are shown for the water hexamer, the prism being the lowest energy structure at the calculated level.

Cooperativity is another curious feature of hydrogen bonds: when a bond is established, the hydrogen-bond acceptor becomes polarized and, consequently, its strength as a donor increases. Hence, the difference in energy between the water trimer and 3 isolated water molecules is greater than the sum of the 3 newly established hydrogen bonds. A further stabilization in energy is due to cooperativity and occurs through progressive charge transfer between H and the acceptor sites ^[32].

When water clusters are confined in apolar environments^[40-41], such as nanotubes or protein cavities, their minimum energy conformations are quite different, as are those of ionic water clusters^[42-43].

The issue of proton dissociation is intimately connected with hydrogen bonds. The pKa of an hydrogen bond donor (the acid) depends largely on the strength of the hydrogen bond formed with its acceptor (in this case, water). Most computational studies on acid dissociation focus on the determination of dissociation constants through the use of thermodynamic cycles ^[44], an approach which necessarily concerns bulk systems. A smaller number, however, has ventured into the realms of hydrated acid clusters ^[45-46]. The basic question these studies aimed to answer is “How many water molecules are necessary to cause proton dissociation?” A digest of results concerning some simple acids is provided in Table 2.

Table 2 Calculated number of water molecules needed to ionize a few simple acids. n_L and n_G denote the number of water molecules present in the ionized clusters. n_L corresponds to local minima and n_G to global minima. The references from which the data present in this table was taken can be found by consulting the review by Leopold ^[45]

Acid	n_L	n_G
HF	4	>7
HF		>10
HCl	4	4 ^b
HBr	3	4
HBr	3	4
HI	3	3 ^c
HI	3	3
H ₂ SO ₄	3 ^d	5 ^d
HNO ₃	4	
HNO ₃	4	5
HClO ₄	3	3
HSO ₃ X (X = F, Cl, Br)	3	3
CH ₃ SO ₂ OH	4	4

In general, 3 to 5 water molecules are needed for proton dissociation to occur among the studied systems. For the weak acid trifluoroacetate (TFA) the “magic number” of hydration water molecules necessary to abstract a proton has not yet been determined neither by theory nor experiment. Thus, the hydrated TFA clusters are a case study of the present project, discussed in Chapter 4.

The estimation of pKa values through *ab-initio* methods has also attracted our attention, even more so given the recent trends of using hydrates of the solute to

explicitly represent its first solvation shell. In Chapter 5, we propose a novel scheme for the estimation of pKa values using the hybrid implicit-explicit approach.

The aforementioned “case studies”, which explore the theme of acid dissociation in microsolvated environments from two different perspectives, constitute the bulk of the present work.

Additionally, two mini-projects are presented in Chapters 2 and 3.

In Chapter 2 the performance of different levels of theory on the calculation of the global minimum structures of small water clusters is assessed.

Chapter 3 discusses a popular article regarding linear water chains. One of the structures presented as a local minimum is in fact an artifact, unsuspected by the author. The discussion serves as a cautionary tale against the all too common pitfall of overlooking optimization criteria, leading to false minima which appear as legitimate.

1.4. Tools of the trade

Just as an experimentalist does nothing without laboratory equipment, so a computational chemist needs his software to perform calculations. For the present work, the Gaussian09^[47] program will be used. It is a powerful and complete commercial software package, widely used for *ab initio* calculations.

At the end of a calculation, Gaussian randomly selects a quote, which may be wise, funny or downright demotivational but always seems to fit the occasion. Since the inspiration necessary to write a meaningful, one-paragraph conclusion to this long introduction is lacking, I relied on software for help and ran a random job hoping to get a fitting quote. Once again, Gaussian does not disappoint and simply points out

“I could have done it in a much more complicated way” – said the Red Queen, immensely proud. ’ Lewis Carroll, *Alice in Wonderland*

Chapter 2

Mini-project: exploring ab-initio methods through water clusters

2. Mini-project: Exploring *ab-initio* methods through water clusters

2.1 Introduction

Water clusters were among the first cluster systems to be studied through *ab-initio* methods. In the 90's, Sotiris Xantheas published a number of studies in the *Journal of Chemical Physics* ^[37, 48-50] regarding small water clusters which became almost mandatory literature items and constitute an excellent first case study for those who are only beginning to learn the art of computational chemistry. My first practical task as an apprentice was indeed to replicate the results of Xantheas using different levels of theory. This project allowed me to get acquainted with the Gaussian09 software, use GaussView, prepare input files, learn where to find information in output files and get a basic intuitive sense of how well each level of theory performs and how much computational resources it requires.

2.2 Computational methods

The starting geometries of $(\text{H}_2\text{O})_{n=2-4}$ clusters were built according to the topologies reported in the literature to be the global minima for each cluster size. These were further optimized at the following levels of theory and indicated basis-sets:

HF/6-31G

HF/6-31++G(d,p)

B3LYP/6-311++G(d,p)

X3LYP/6-311++G(d,p)

MP2/aug-cc-pvdz

Vibrational analysis was performed to ensure that all optimized structures correspond to true minima on the PES, as indicated by the absence of imaginary frequencies.

The binding energy of each cluster was calculated by subtracting the electronic energy of the individual water molecules, in their isolated geometries, from the electronic energy of the complex.

The geometries optimized at the MP2/aug-cc-pvdz were further corrected for the BSSE using the Counterpoise (CP) scheme proposed by Boys and Bernardi ^[26].

As mentioned in the introduction, the calculation of the interaction energy with modest basis sets is subject to the basis set superposition error, that is, an aggregate of molecules is artificially stabilized as each monomer utilizes the nearby basis functions of its neighbors in order to better describe its own wavefunction.

The energy differences resulting from the BSSE error for a dimer AB can be written as

$$\begin{aligned}\Delta E_{BSSE}(A) &= E_{AB}^{MCBS}(A) - E_{AB}^{DCBS}(A) \\ \Delta E_{BSSE}(B) &= E_{AB}^{MCBS}(B) - E_{AB}^{DCBS}(B)\end{aligned}$$

Where the superscript denotes the basis used, the subscript the geometry of the monomer and the monomer is indicated inside parentheses. This may be read as ‘‘The BSSE arising from monomer A is calculated by subtracting the energy of A, calculated with a dimer centered basis set (DCBS) in the dimer geometry, from the energy of A, calculated with a monomer centered basis set (MCBS) in the dimer geometry (AB)’’.

Since $E^{DCBS} < E^{MCBS}$ (due to the discussed artificial stabilization effect) then $\Delta E_{BSSE} > 0$ and the counterpoise corrected energy of the complex is given by

$$E^{CP}(AB) = E_{AB}^{DCBS}(AB) + \Delta E_{BSSE}(A) + \Delta E_{BSSE}(B)$$

The calculation of the counterpoise corrected binding energy is a more complex matter than its uncorrected counterpart and requires further discussion. Let the uncorrected binding energy be written as

$$E_{bind}(AB) = E_{AB}^{DCBS}(AB) - E_A^{MCBS}(A) - E_B^{MCBS}(B)$$

The corrected binding energy is computed by adding the BSSE to $E_{bind}(AB)$ yielding

$$\Delta E_{bind}^{CP} = E_{AB}^{DCBS}(AB) - E_A^{MCBS}(A) - E_B^{MCBS}(B) + [E_{AB}^{MCBS}(A) - E_{AB}^{DCBS}(A)] + [E_{AB}^{MCBS}(B) - E_{AB}^{DCBS}(B)]$$

Note that the corrected binding energy is a sum of the interaction energy,

$$\Delta E_{int}^{CP} = E_{AB}^{DCBS}(AB) - E_{AB}^{DCBS}(A) - E_{AB}^{DCBS}(B)$$

and the monomer deformation energies,

$$E_{AB}^{MCBS}(A) - E_A^{MCBS}(A) \text{ and } E_{AB}^{MCBS}(B) - E_B^{MCBS}(B).$$

2.3 Results and discussion

The optimized geometries of (H₂O)₂₋₄ clusters, corresponding to global minima on their respective PES, are displayed in Fig. 6. Their features are well known and have been amply discussed, therefore we shall skip the classic detailed geometric analysis and proceed with the method performance comparison.

Some relevant geometric parameters, namely oxygen-oxygen distance ($R_{(O...O)}$), hydrogen bond length ($r_{(O...H-O)}$), O-H bond lengths ($R_{(O-Hd)}$ and $R_{(O-Hf)}$ where d denotes donor hydrogen and f indicates a free hydrogen), dihedral angles ($\delta_{(O...H-O)}$) and bond angles ($\Phi_{(H-O-H)}$) of all water clusters have been measured and their binding energies calculated. The average value for each descriptor was plotted against the number of water molecules in the cluster. The resulting Plot 1(I-VII) allow a visual performance assessment of the various levels of theory. The data reported by Xantheas^[48] concerning the same group of clusters and computed at the MP2/aug-cc-pvdz level is also included.

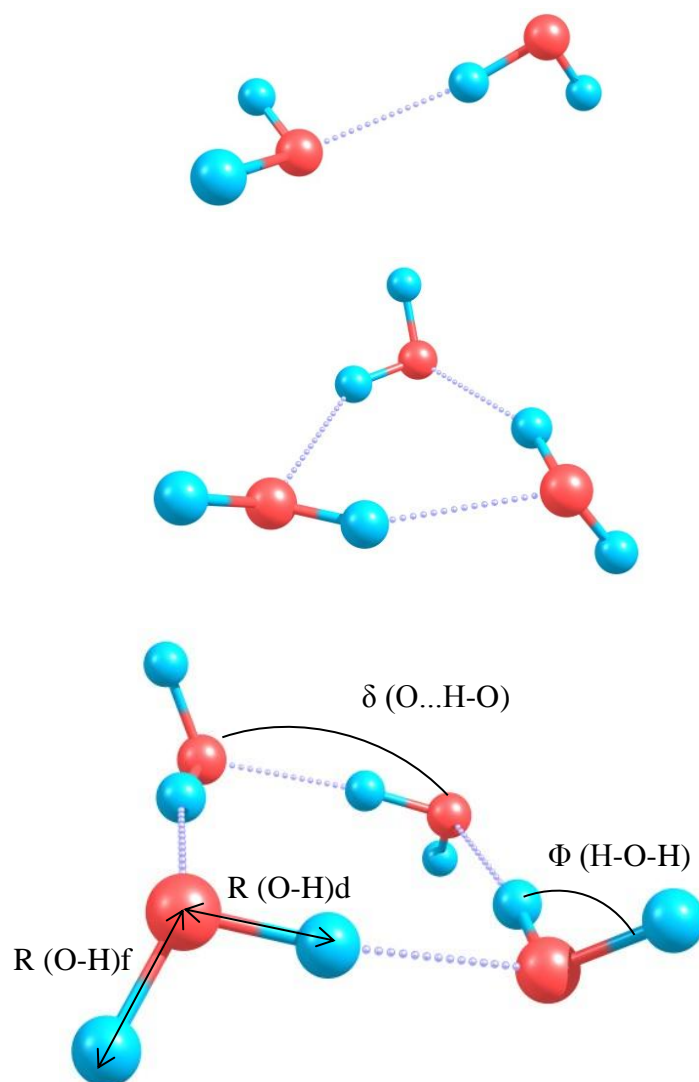


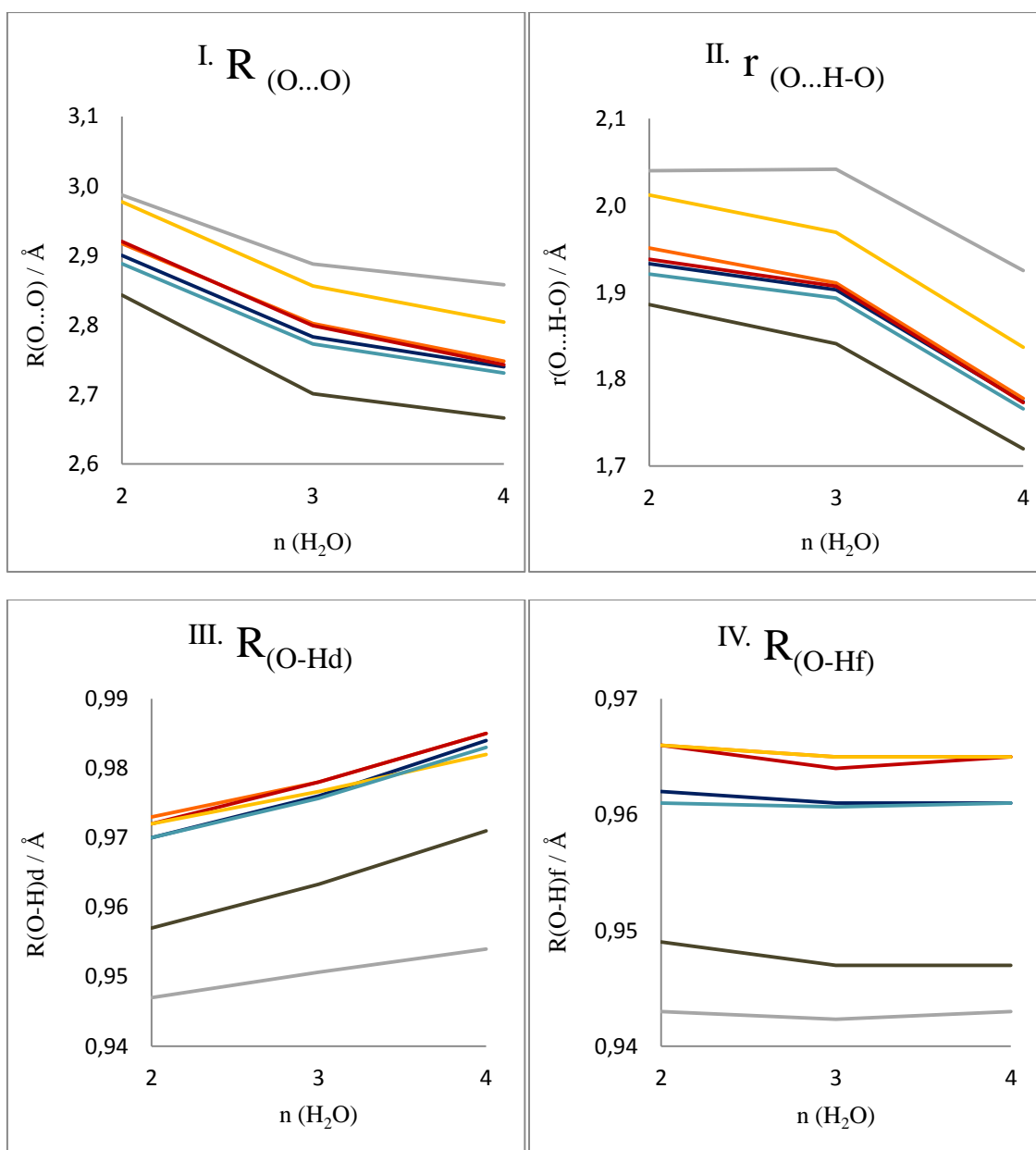
Fig. 6 Minimum energy structures of the water dimer, trimer and tetramer.

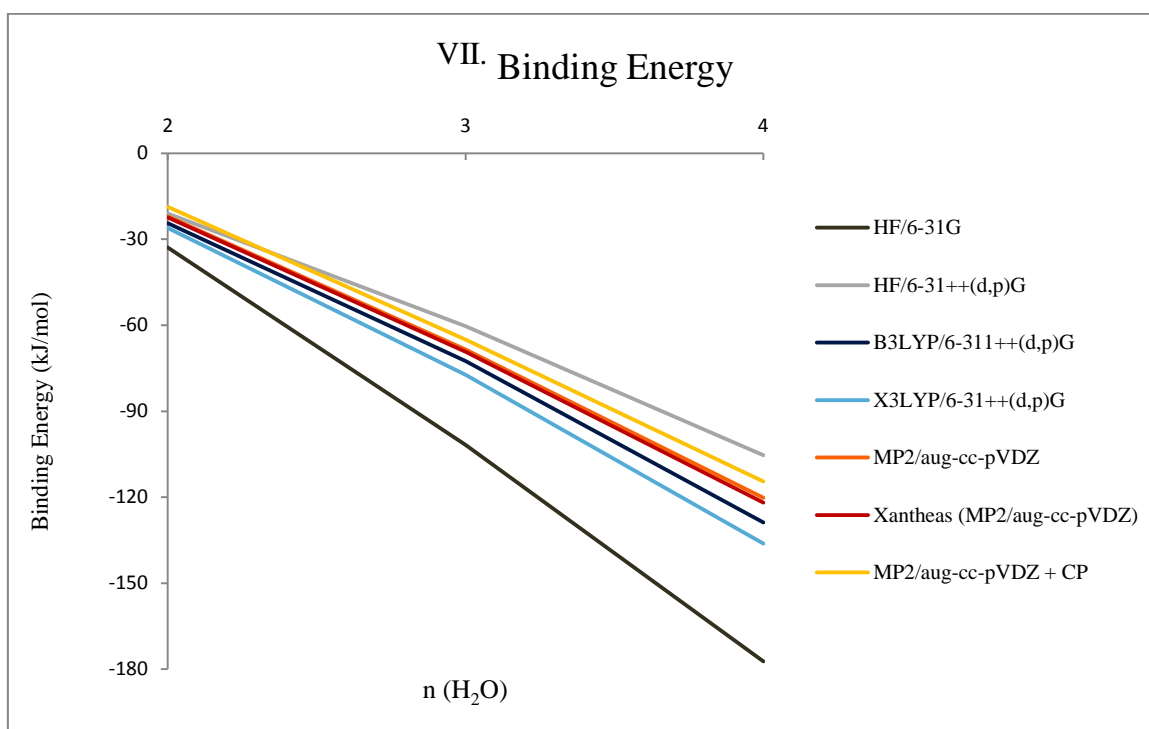
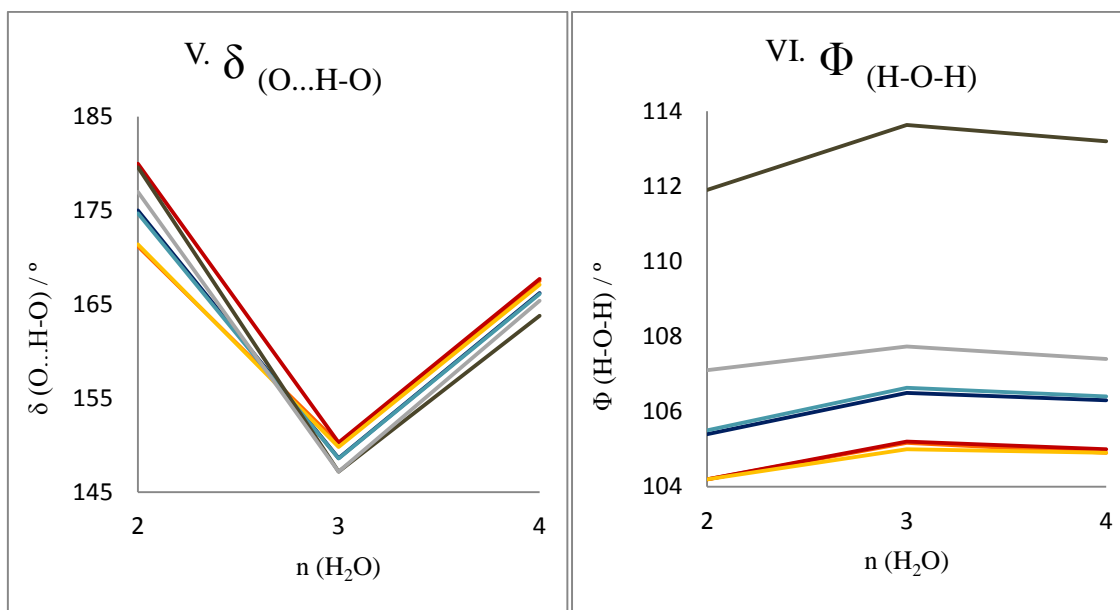
The main conclusions one draws from this visual assessment are the following:

- Our results obtained at the MP2/aug-cc-pvdz level successfully replicate those of Xantheas. Since this was the most accurate method and largest basis set used in the current study, it serves as benchmark for assessing the performance of the other methods;
- The HF method delivers the worst results, even using a basis set with diffuse and polarization functions. As discussed in the introduction, this bad performance is due to the inability of HF methods to properly account for electron correlation.
- B3LYP and X3LYP functionals perform almost identically;
- B3LYP results are very similar to those obtained at the MP2 level. Interatomic distances and dihedral angles computed at these two levels of theory are practically identical, while B3LYP consistently overestimates bond angles by 1° and slightly overestimates the binding energy.
- As expected, the counterpoise correction lengthens the intermolecular distances while not affecting neither covalent bonds nor angles. Accordingly, the counterpoise corrected binding energy is lower, denoting a weaker complex than that predicted without correcting for BSSE.

Since the B3LYP/6-311++G(d,p) combination performs very well while being computationally efficient, it was selected as the “working method”, to be used in subsequent projects involving larger systems, such as the acid hydrates.

- HF/6-31G
- HF/6-31++G(d,p)
- B3LYP/6-311++G(d,p)
- X3LYP/6-311++G(d,p)
- MP2/aug-cc-pvdz
- Xantheas (MP2/aug-cc-pvdz)
- MP2/aug-cc-pvdz + CP





Plot 1 Geometric and energetic parameters of the water dimer, trimer and tetramer optimized at different levels of theory. $R_{(O...O)}$ is the oxygen-oxygen distance, $r_{(O...H-O)}$ the hydrogen bond length, $R_{(O-H_d)}$ and $R_{(O-H_f)}$ are the O-H bond lengths where d denotes donor hydrogen and f indicates a free hydrogen, $\delta_{(O...H-O)}$ stands for dihedral angle and $\Phi_{(H-O-H)}$ for bond angle.

Chapter 3

“GIGO” - Garbage In, Garbage Out: the case of linear water chains

“GIGO” - Garbage In, Garbage Out: the case of linear water chains

As in any other field of chemistry, there are numerous pitfalls along the computational chemist's path. One of these pitfalls is the blind reliance on software packages without previous thorough knowledge of their inner workings.

Sometimes, what seems like a minor detail dramatically impacts computational results. With enough luck, erroneous conclusions drawn from faulty output are detected and corrected before reaching publication. However, some stray examples manage to make their way through the long peer-reviewed quality control system, thus inducing the reader in error. In some unfortunate cases, the error is taken as truth and propagates through subsequent work thus wasting precious intellectual and material resources. Having fallen prey to this situation, I've decided to expose the unfortunate mistake that led me astray thus warning fellow researchers against similar traps in their own work.

As discussed in the introduction of the present dissertation, the available literature on water clusters is quite extensive. While in the beginning most efforts were directed at finding the minimum energy structures of small water clusters in the gas-phase, lately the spotlight has been on the arrangement of water molecules in biological media, crystal matrices and nanoporous organic or inorganic hosts. While in the gas phase the most stable structures are invariably planar rings or built of fused rings of water molecules, when trapped in confined environments water molecules tend to adopt radically different topologies, including linear and helical arrangements.

Recently, a few authors endeavored to explore the stability of such 1D water clusters in the gas-phase. Parthasarathi, Elango, Subramanian and Sathyamurthy ^[51] performed geometry optimization of wire-like helical water clusters with 5 to 20 water molecules, in the gas phase, at the HF/6-311++G** and B3LYP/6-311++G** levels of theory without using any constraints.

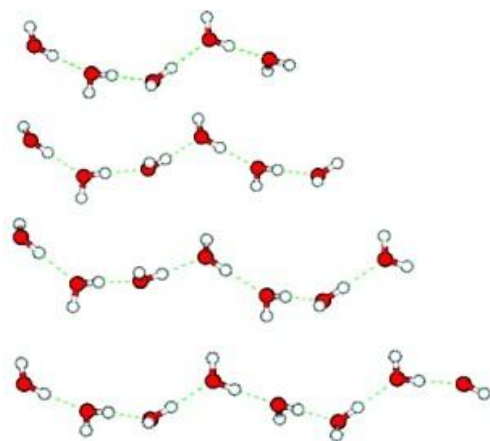


Fig. 7 Linear helical water chains minimum energy structures, as reported in ^[51]

The helical structures displayed in Fig. 7 were reported to be minima in the PES. The authors state that “these water clusters form a linear helical chain motif even in the absence of host molecules/scaffolds” and “cyclization or formation of cage/prism structures was not observed at the HF and B3LYP(DFT)/6-311++G** levels of theory (...) However, MP2/6-311++G** calculations for (H₂O)_{5,7} clusters led to the formation of open faced prism-like structures and a combination of trigonal cyclic and square planar arrangements and also spiro-cyclic assemblies.” This obvious dissonance should have alerted the authors for the possibility of the HF/DFT results being defective. On the contrary, this fact was merely reported without discussion. The first impression I got after reading this article was favorable, as I also did not notice the aforementioned dissonance.

On the contrary, I was inspired to check whether TFA would form stable clusters with these water helices. In order to so, I first built a water helical wire with 5 water molecules, the smallest size reported by Parthasarathi and his team, and performed a geometry optimization at the B3LYP/6-311++G** level of theory. To my dismay, the optimization procedure led the initial helical configuration to morph into a spiro-cyclic ^[52] one, although the optimization did not converge. A second optimization attempt led to the well known planar ring structure.

At this point, I put the TFA hydrate search aside and dedicated my time to solving this mystery. Since the initial structure I used was as similar as possible to the “stable” wire reported by the authors and the level of theory was the same, something else had to

account for the different result. Briefly re-reading the “Computational Details” of the article in discussion sufficed to identify the major difference: the authors used the G98W version of Gaussian while I used the G09W version.

The default parameters of these versions differ significantly. Assuming that the authors have, as I did, ran the program using the default parameters, these differences could lead to diverging results. This hypothesis turned out to be correct.

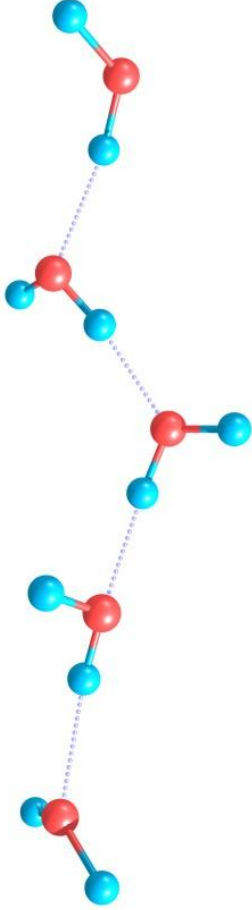
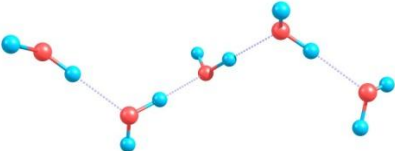
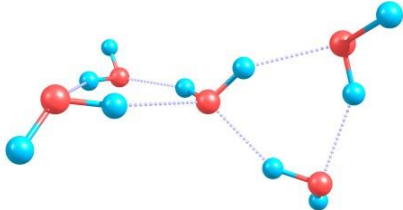
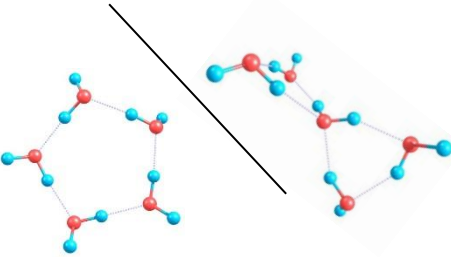
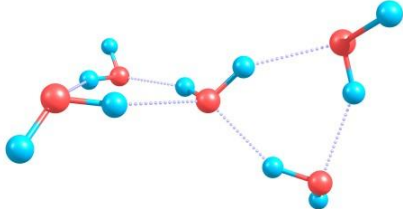
Firstly, I optimized the same initial structure with G03W, which is more similar to G98W than G09W (G98W is no longer available for testing).

As shown in Table 3, the optimization in G03W with default parameters yielded a minimum with a linear helical structure. G09W and G03W differ in the default convergence criteria for the scf procedure, which is stricter in G09W (scf=tight).

A second batch of optimizations were thus ran in both programs with verytight convergence criteria for the scf procedure. As the label indicates, the verytight criteria requests full accuracy for the scf procedure and should be used whenever doubt arises as to the legitimacy of a given minimum energy structure. The output, for both versions of the program, was a fully converged minimum energy structure with a spirocyclic configuration. A third batch of optimizations was yet executed using a finer DFT integration grid (int=ultrafine), yielding the same result as in the previous case. A finer DFT grid calculates more density points and is thus more accurate.

It is then safe to conclude that the linear chain with five water molecules reported by Parthasarathi is an artifact generated by an insufficiently accurate optimization procedure. The other linear helical structures, with 6 and more water molecules, also reported by the author, have also been tested and proved to be true minima within the G09W stricter conditions. It is easy then to understand how easily the author was deceived into thinking the same applied to the linear cluster with 5 water molecules. Let this example be a gentle reminder not to blindly trust our output, especially when using default convergence criteria.

Table 3 Results of the optimization procedures performed with different parameters in G09W and G03W

INPUT	JOB PARAMETERS	OUTPUT
	G03 with default parameters	helical, minimum 
	G03 with scf=verytight	spirocyclic, minimum 
	G03 with int=ultrafine	
	G09 with default parameters	Ring or spirocyclic, not converged 
	G09 with scf=verytight	spirocyclic, minimum 
	G09 with int=ultrafine	

Chapter 4

Stepwise solvation of trifluoroacetic acid: chasing the onset of proton dissociation

4. Stepwise solvation of trifluoroacetic acid: chasing the onset of proton dissociation

4.1 Introduction

“Nature uses only the longest threads to weave her patterns, so that each small piece of her fabric reveals the organization of the entire tapestry.”

— Richard P. Feynman, *The Character of Physical Law* (1965)

As Feynmann beautifully states, the fundamental nature of chemical interactions between species may be probed by studying a small system, which serves as a model for a larger reality. With this view in mind, scientists eagerly study water clusters of various sizes, hoping to gain insight on the interplay of intermolecular interactions governing the behaviour of water, from the gaseous to the condensed state ^[53-54]. Clusters, therefore, do not belong to neither the gaseous nor the condensed realms but, instead, span the intermediate worlds between them.

One of the fundamental processes now being probed through cluster chemistry is the dissociation of acid and basic solutes in aqueous media. Acid-base chemistry sits at the foundation of life as we know it and is important in a wide range of fields. Regardless of its widespread importance, fundamental knowledge of these systems is still in its infancy.

Theoretical investigations on the issue of acid dissociation in microsolvated environments are usually of a quantum mechanical nature and concern gaseous acid hydrates of varying sizes ^[45]. Of great interest are the structures and energetics of these hydrates in both their neutral and zwitterionic forms. For every species, there is a minimum number of hydration water molecules required to provoke dissociation. These may be modeled in the gas phase or in a continuum solvation model.

The first acids to be studied in this fashion were the halogenated species HF, HCl, HBr and HI ^[55], attractive due to their high acidity and small size. Nowadays, the scientific community is leaning towards the solvation and dissociation of organic acids, especially carboxylic acids ^[46, 56-64], which are ubiquitous in biological environments. Besides, carboxylic acids often have an amphiphilic nature, experiencing hydrophilic hydration around the COOH head group and hydrophobic hydration of the alkyl chain.

Trifluoroacetic acid (TFA) is one of the strongest organic acids known so far, with an experimentally measured pK_a of ~0,5. It plays an important role in atmospheric chemistry, being the most abundant halogenated acid in the environment, a product of the oxidation of hydrochlorofluorocarbons. Since TFA does not react with OH· radicals, its primary route of removal from the atmosphere is through wet deposition. TFA hydrates thus serve as nucleation sites for cloud condensation in the rainout process. Computational studies permit the evaluation of how this system evolves (up to a certain limiting size) and the retrieval of spectroscopic data which is of crucial importance for the monitoring of TFA hydrates in the environment.

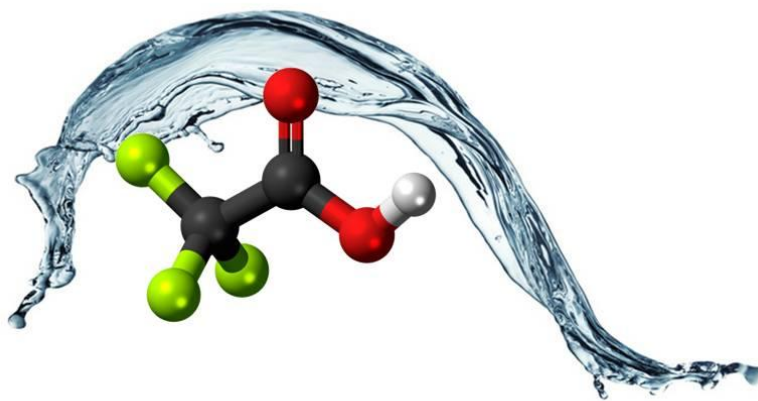


Fig. 8 3D representation of trifluoroacetic acid

Although carboxylic acids, especially acetic acid, have received widespread attention from the theoretical chemistry community, the literature on TFA is relatively

scarce. The following paragraphs provide the most relevant conclusions from studies concerning TFA hydrates and solvation. These stem from an atmospheric chemistry perspective and thus focus in small TFA hydrates, with 1 to 4 water molecules, which are expected to be the dominant forms in the atmosphere.

Ito ^[65] and Ouyang, Starkey and Howard ^[66] performed experimental measurements on small TFA hydrates and compared those with calculated data. Although using different characterization techniques – the former, infrared; the latter, microwave spectroscopy – the authors reached convergent conclusions regarding the structure of TFA clusters.

Ito prepared TFA clusters by mixing TFA and vapor in an Ar matrix and measured their infrared spectra. Using DFT and MP2, he performed geometry optimizations of TFA clusters with 1-4 water molecules and found the corresponding minimum energy structures, shown in Fig. 9, as well as transition states. A striking similarity may be found between the arrangement of water in TFA clusters with up to 4 water molecules and pure water clusters. Like the former, small TFA clusters are the most stable in the ring conformation, in a donor-acceptor pattern that maximizes cooperativity. Hydration water molecules establish side-on hydrogen bonds with the carboxyl moiety, which acts as an hydrogen bond donor through the –OH and an acceptor, through the =O. As the cluster size increases, the distance between the acidic proton and the bonded acceptor decreases, but not to the point of dissociation.

A vibrational analysis of the minimum energy structures was performed and the calculated spectra confirmed the assignments based on experimental results. However, three experimental peaks, between 1720 cm⁻¹ and 1740 cm⁻¹, could not be identified. The authors hypothesize that these peaks either correspond to a dissociated TFA cluster or to a neutral cage-like structure with n>4. Due to the large amount of structural candidates for dissociated clusters, the authors did not attempt to measure the vibrational spectra of dissociated species.

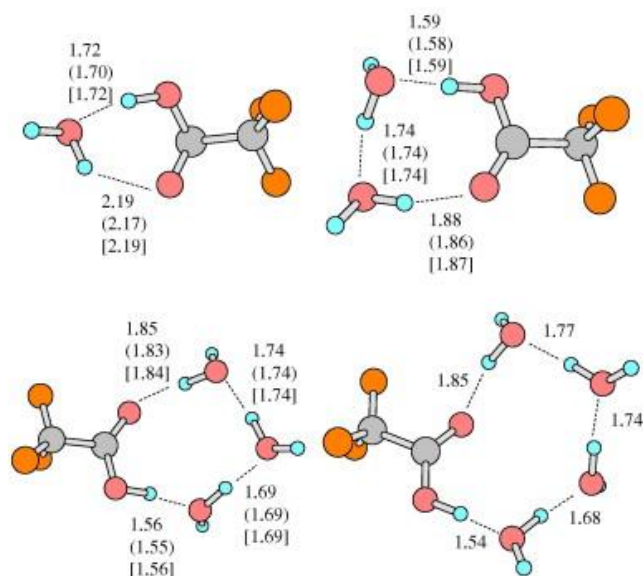


Fig. 9 Minimum energy structures of TFA(H₂O)_n clusters for n=1-4 obtained at the B971/6-311++G(3df,3pd) level. Hydrogen bond distances are expressed in angstrom. Distances within parentheses and brackets were obtained at the MP2/6-311++G(3df,3pd) and B3LYP/6-311++G(3df,3pd) levels, respectively.

Thomas ^[67] and co-workers provided spectroscopic evidence, supported by an *ab-initio* MD study, pointing out that the bulk dissociation process of TFA involves the formation of a contact ion-pair intermediate, where the carboxylate moiety is in direct contact with the hydronium cation, before transitioning into a solvent separated form, in which the hydronium cation is in the second solvation shell. In a diluted aqueous TFA solution both forms were shown to exist in dynamic equilibrium.

Clearly, the uncharted territory of TFA hydration and dissociation is vast. The general goal of this case study is to explore such unventured realms and, hopefully, cast some light on the subject.

The aforementioned studies did not determine the required number of water molecules for TFA dissociation to happen. Such is the task of the present work. We found minimum energy structures of TFA hydrates with 4 to 9 water molecules in their dissociated and undissociated states. The topologies and energetics of these clusters have been analyzed in an attempt to illustrate the dissociation phenomena in growing clusters.

4.2 Computational methods

The systems at study consist of TFA hydrates in the undissociated – HTFA.(H₂O)_n – and zwitterionic, or dissociated state – TFA⁻(H₃O⁺)(H₂O)_{n-1} – with n=[4-9]. The minimum energy structures presented herein were optimized at the B3LYP/6-311++G** level of theory using the Gaussian 09^[47] electronic structure program. Vibrational analysis confirmed the latter to be local minima due to the absence of imaginary frequencies. A re-optimization of all structures, using the VeryTight SCF convergence criterion, UltraFine grid and counterpoise correction was carried out. Transition state structures connecting dissociated and undissociated minima were calculated using the QTS2 scheme. The presence of a single imaginary frequency confirmed these structures are first-order saddle points.

The total interaction energy (*E*_{int}) may be calculated as the difference between the energy of a complex AB in its optimized geometry and the sum of the energies of the individual subsystems A and B with the same geometries as those adopted in the complex AB, which may be written as

$$E_{int} = E(AB)_{AB} - E(A)_{AB} - E(B)_{AB}$$

The interaction energy was corrected for the basis set superposition error (BSSE) using the counterpoise scheme.

The Gibbs free energy of all hydrates was calculated at 298K and 1 atm. For the calculation of the Gibbs free energy Gaussian09 computes assumes the system behaves as an ideal gas and that its first and higher excited states are inaccessible. The software uses standard statistical thermodynamic formulae in order to obtain the electronic, translational, rotational and vibrational contributions to entropy and enthalpy and then retrieves the Gibbs free energy through the equation $G=H-TS$ (for detailed information please consult ref^[68]).

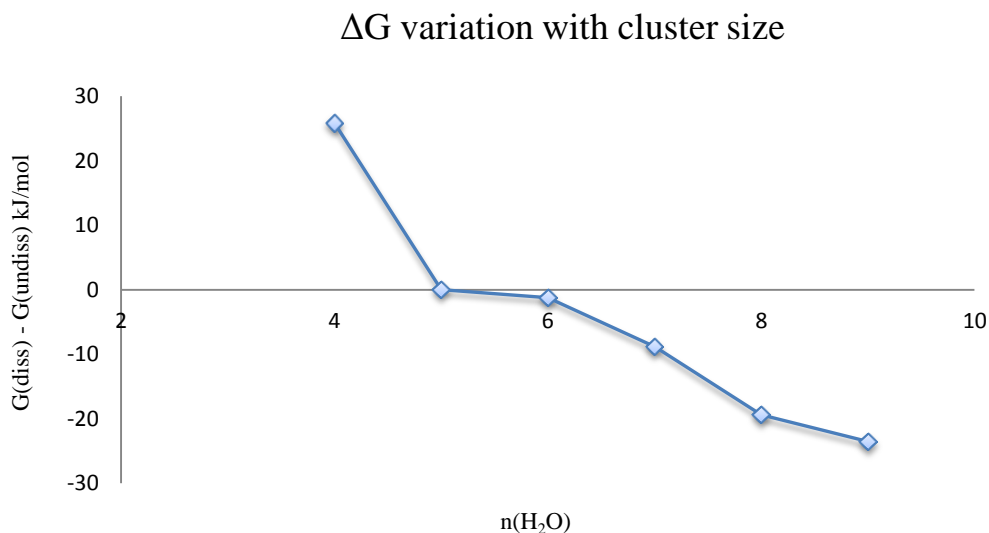
The energy of each individual hydrogen bond was estimated using a correlation function suggested by Wendler and co-workers^[69]. Wendler evaluated the interaction

energies of more than 250 hydrogen bonded dimers and correlated these with physical descriptors, such as hydrogen bond length, by fitting functions. For the present analysis we took the energy-bond length fitting function and re-parametrized it to better describe the particular case of TFA hydrates, yielding the optimized fit displayed in the following equation, where E_{int} is the interaction energy in kJ/mol and r_{Hb} the hydrogen bond length (H.....O) in pm.

$$E_{int} = (1,19 \times 10^{10}) / (r_{Hb}^{3.816})$$

4.3 Results and discussion

4.3.1 What is the minimum number of water molecules required for TFA dissociation?



Plot 2 Variation of the free energy difference, ΔG , between the dissociated and undissociated hydrates with cluster size

Plot 2 displays the relative Gibbs free energy at 298K of TFA hydrates in the dissociated and undissociated forms.

Hydrates with 1,2 and 3 water molecules were not included since, in this size range, no dissociated minima were found.

With 4 water molecules, dissociation becomes possible, although clearly unfavourable, since the dissociated hydrate is ~30 kJ/mol higher in free energy than its undissociated counterpart. Note that the undissociated analog is not the global minimum at $n=4$, which is a planar ring, but rather a tent-like structure similar to the dissociated form. We have decided to compare structural analogues instead of the lowest possible minima in order to understand how proton transfer alone affects the overall energetics of the system by itself. If structures with radically different topologies were compared, we would have to account for larger differences in strain and number of hydrogen bonds between dissociated and undissociated clusters. Within the chosen strategy, the main difference between the former analogs is the placement of the acidic proton.

With 5 water molecules, the undissociated and dissociated clusters are practically isoenergetic, although the dissociated form is marginally more stable. Hence, formally, the minimum number of H_2O molecules required to cause TFA dissociation is 5. In this size range, the interconversion between dissociated and undissociated forms is barrierless, as proven by the inexistence of a transition state linking the two. This barrierless transfer is maintained at $n=6$ although in this case the dissociation process is somewhat more favorable.

As more water molecules are added, dissociation becomes increasingly advantageous. The steepness of the plot decreases from $n=8$ to $n=9$ which may indicate that the chemical environment of this cluster size approaches that of the bulk.

4.3.2 Geometries of TFA hydrates

The optimized geometries of dissociated and undissociated TFA hydrates are displayed in Table 4.

The dissociated local minimum found at $n=4$ displays a tent-like structure, with the hydronium ion sitting on top, forming 3 hydrogen bonds with 3 water molecules, which in turn act as hydrogen bond donors towards the carboxylate headgroup. Tent-like structures have been found for a variety of dissociated acids in clusters with 4 water molecules, including formic acid^[57], the halogenated HX series (X=F, Cl, Br, I)^[55], H_2SO_x and HS^[70]. The data thus suggests that, in order for the aggregate to be stable, even as a local minimum, the hydronium ion must be tri-coordinated.

A dramatic structural transition takes place from $n=4$ to $n=5$ and curved topologies formed by planar fused rings become dominant. These topologies are similar to those found in hydrates of other carboxylic acids. Although our limited PES search does not allow us to determine if these are local or global minima, the striking similarity between the topology of these clusters and those of protonated water clusters^[71], protonated carbonic acid hydrates^[72] and the aqueous hydration shell of carboxylate anions^[73-74] are good indicators that the minimum energy structures found herein for TFA hydrates are, at least, realistic candidates for global minima.

The dissociated hydrates with $n=5$ and $n=6$ have a contact ion-pair formed by the carboxylate unit and the hydronium ion. As the cluster grows, from $n=7$ onwards, topologies where the hydronium ion is located in the second solvation shell of TFA^- become more favourable. From these observations one concludes that the force driving the proton away from the acid and towards the water molecule network grows stronger as the latter grows larger.

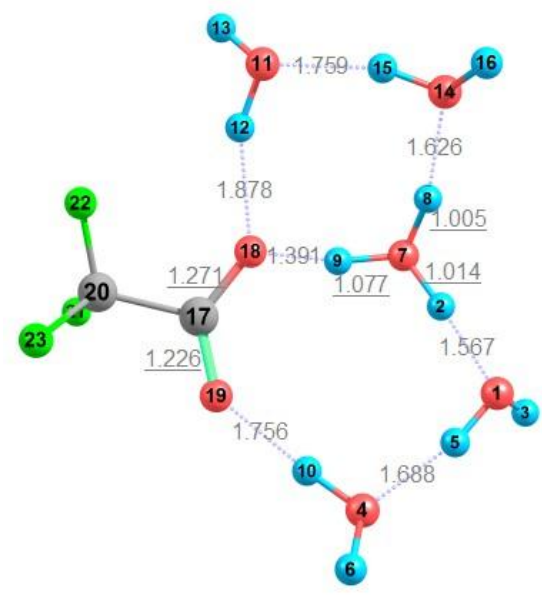
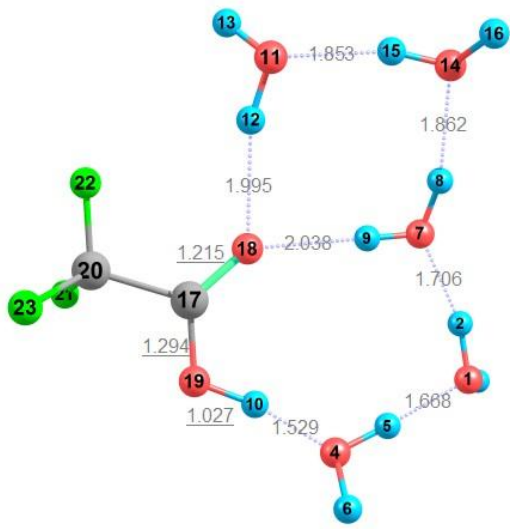
The prevalence of pentameric and, to a lesser extent, tetrameric units, is also found in protonated water clusters of this size range^[71] and this “preference” is usually attributed to the reduced conformational strain of this geometry.

As expected, even in the larger cluster studied herein, the solvation shell is located around the carboxylate/carboxyl moiety while the CF_3 backbone remains desolvated. Such is expected since the carboxyl(ate) end is capable of forming 4 hydrogen bonds with the water molecule network, while the CF_3 group is superhydrophobic, repelling both hydrophilic and hydrophobic molecules.

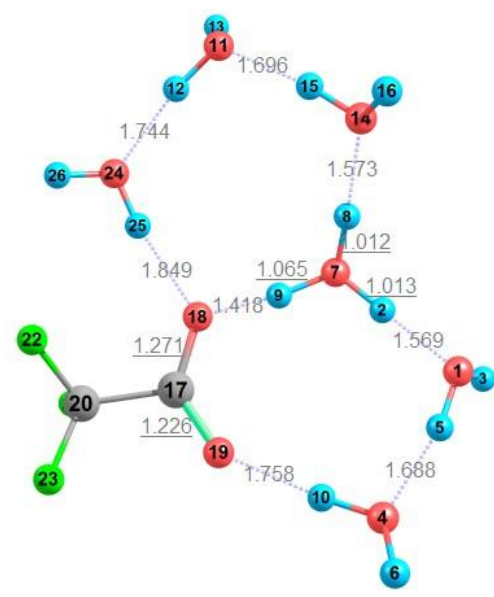
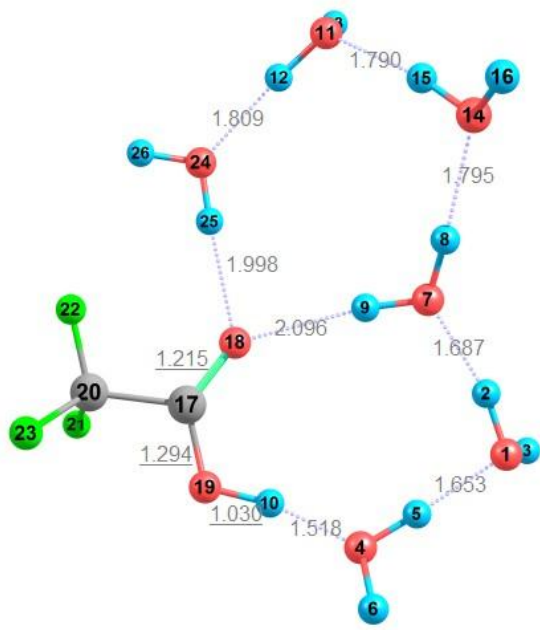
Table 4 Geometries of the dissociated and undissociated TFA hydrates. The indicated values denote hydrogen bond lengths.

n	$\text{HTFA}(\text{H}_2\text{O})_n$	$\text{TFA}^-(\text{H}_3\text{O}^+)(\text{H}_2\text{O})_{n-1}$
4		

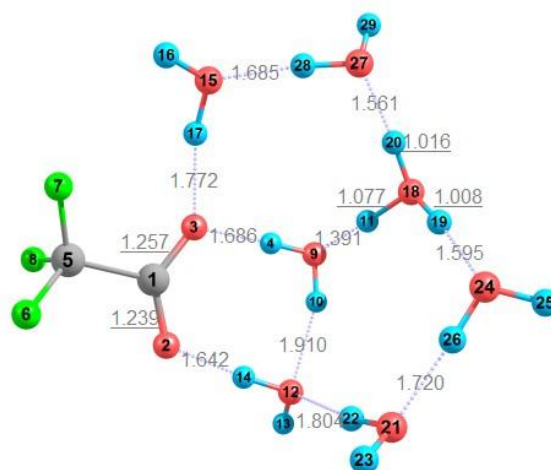
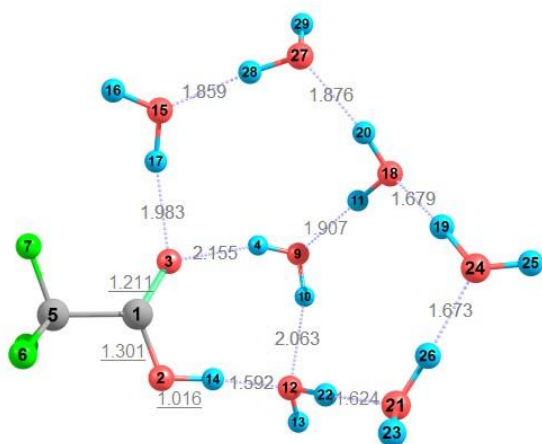
5



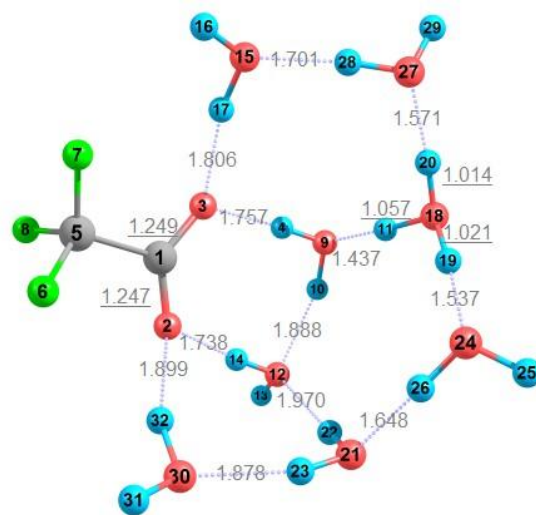
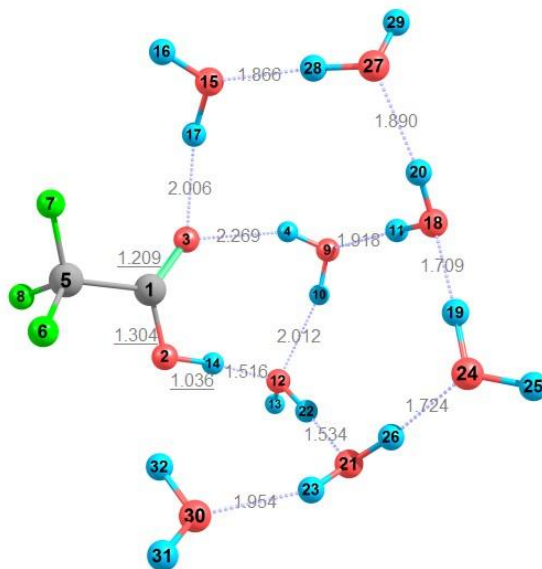
6

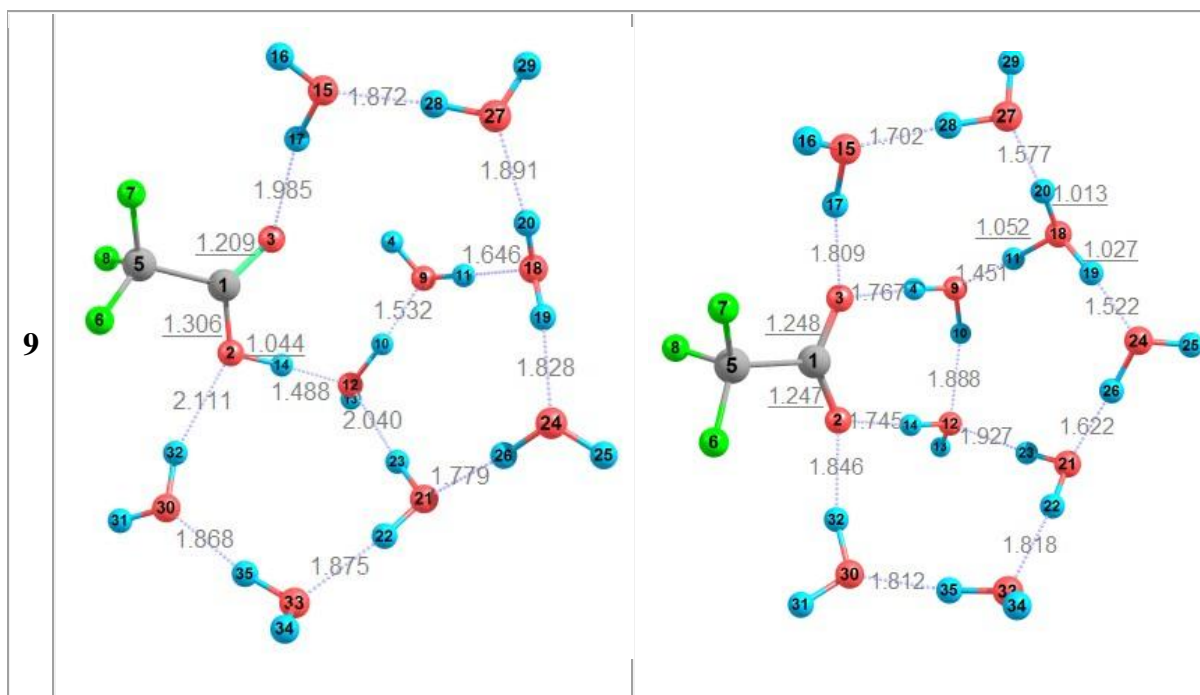


7



8





4.3.3 Evolution of energy of interaction with cluster size

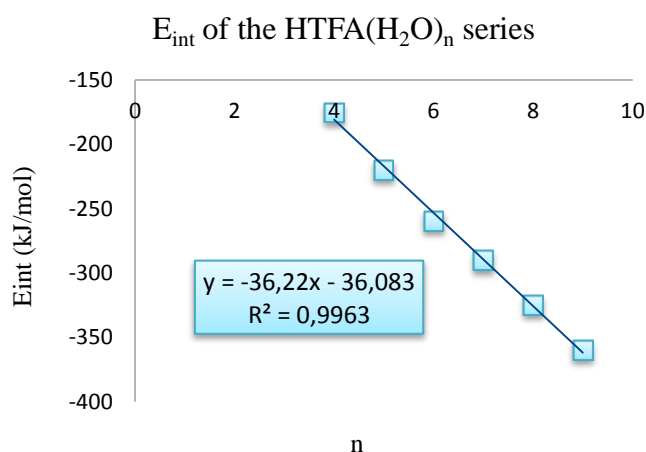
The energy of interaction, E_{int} , of all studied hydrates is negative, ensuring that their formation from the corresponding monomers is favorable.

For both dissociated and undissociated hydrates, the E_{int} increases fairly regularly with each added water molecule, as shown by the high correlation coefficients of the respective linear regression lines displayed in Plot 3 and Plot 5. The stabilizing effect of adding one water molecule is approx. 36 kJ/mol for undissociated clusters and 45 kJ/mol for dissociated clusters (based on the slope of linear regressions). The energetic gain is higher in dissociated clusters since with each additional water molecule the network

becomes more efficient at charge delocalization, an effect which is not as dramatic in undissociated clusters due to the absence of charged species in the latter.

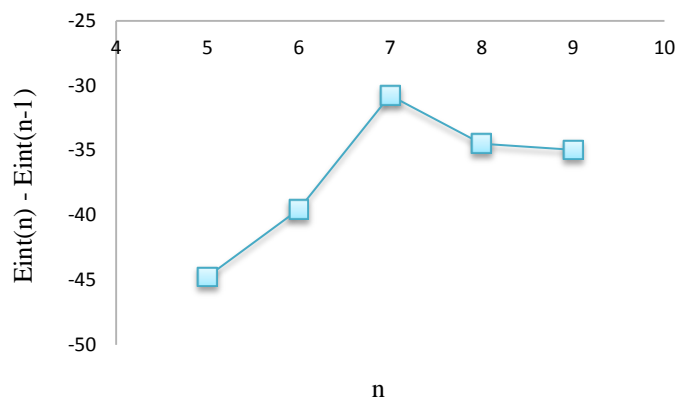
The stepwise interaction energy shown in Plot 4 and Plot 6. The most striking features of these plots concern the first and last transitions. In both dissociated and undissociated hydrates, the highest stepwise interaction energy occurs at the $n=4$ to $n=5$ transition, reiterating that the tent like structure found at $n=4$ is unfavorable, leading to weaker intermolecular interactions when compared with the slightly curved topology adopted thereafter. The plot concerning the $n=7$ to $n=9$ transitions is almost flat, meaning that the increment in E_{int} remains truthfully constant in this size range, hinting that this size range is a “good enough” model for the bulk solvation of TFA and TFA-.

The E_{int} of dissociated clusters is systematically higher than that of the undissociated forms. This occurs not because the dissociated clusters are systematically more stable (Plot 2 shows this is not true until $n=7$) but, instead, because the monomeric units that form dissociated clusters include charged species which, in isolation, are much less stable than neutral species. Consequently, cluster formation is more favorable for an ensemble which includes charged species than for another in which all monomers are neutral. This fact bears no relation to the relative stabilities of dissociated and undissociated clusters.



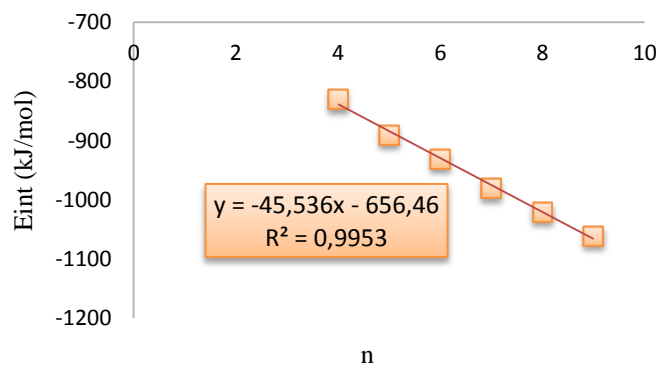
Plot 3 Variation of the energy of interaction E_{int} of undissociated TFA hydrates with cluster size

Stepwise E_{int} of the HTFA(H_2O) $_n$ series

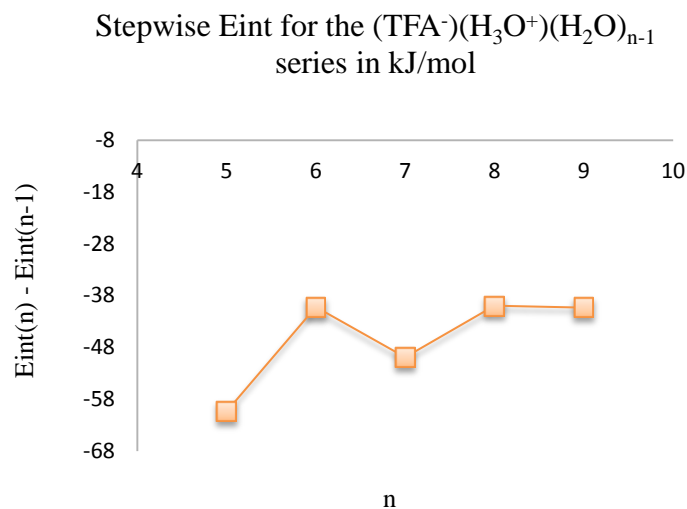


Plot 4 Stepwise increment of the energy of interaction E_{int} of undissociated TFA hydrates with each added water molecule

E_{int} for the $(\text{TFA}^-)(\text{H}_3\text{O}^+)(\text{H}_2\text{O})_{n-1}$ series



Plot 5 Variation of the energy of interaction E_{int} of dissociated TFA hydrates with cluster size



Plot 6 Stepwise increment of the energy of interaction E_{int} of dissociated TFA hydrates with each added water molecule

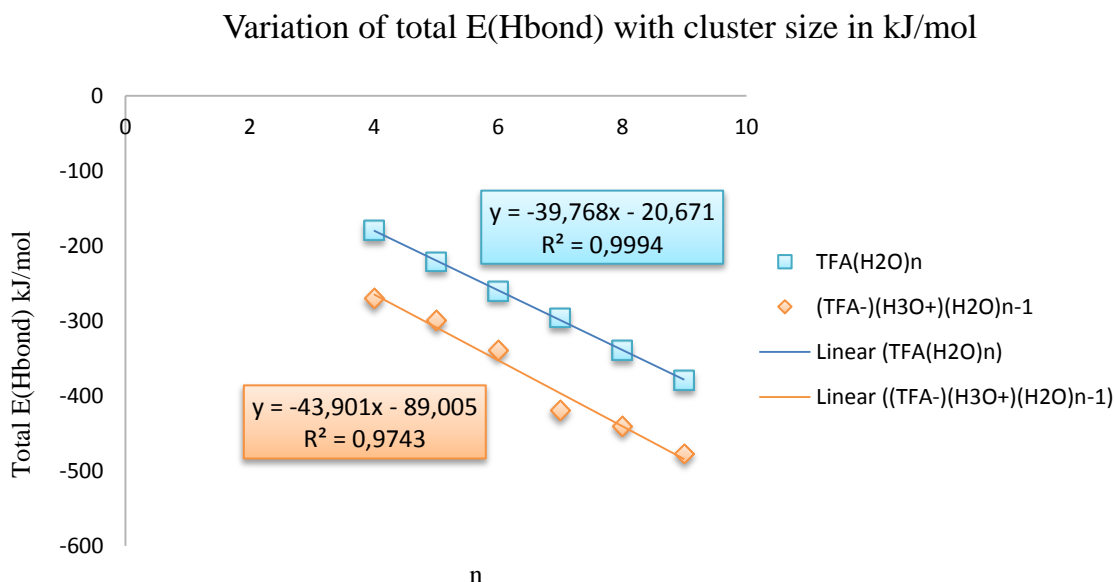
4.3.4 Hydrogen bond energy

The correlation function described above was used to ascribe bond energies to each of the hydrogen bonds present in the studied TFA hydrates. Shorter bonds are stronger, thus higher in energy than longer hydrogen bonds.

The results from the hydrogen bond energetic analysis were quite baffling to us as we expected clusters with overall stronger hydrogen bonding interactions to be the most stable. In fact, as Plot 7 shows, all dissociated clusters have stronger hydrogen bonds than their undissociated counterparts despite the fact that at $n=4$ the undissociated cluster is far more stable. There are two possible reasons for this seeming dissonance:

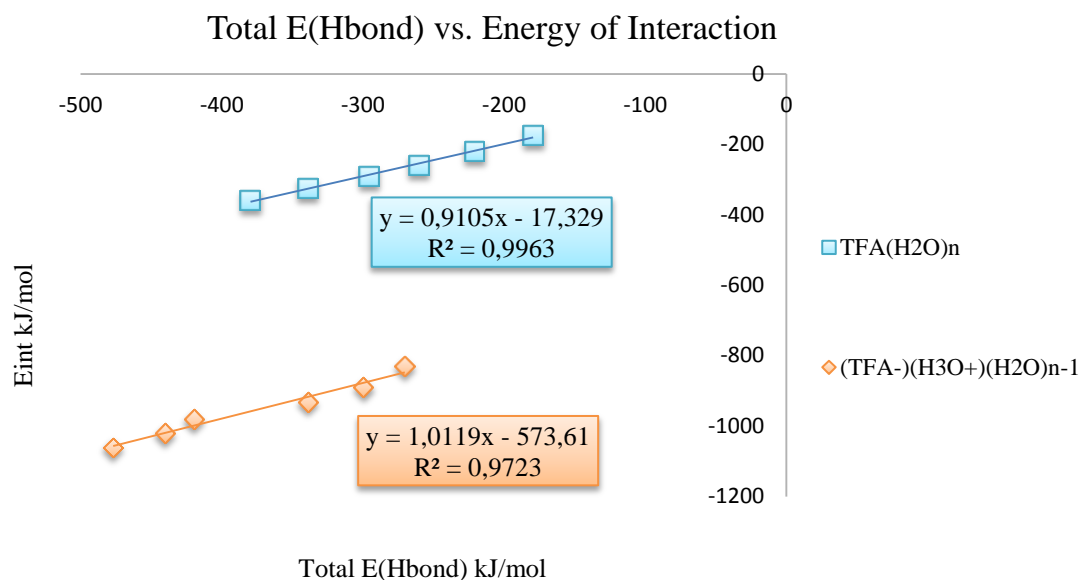
- a) the fitting function employed is improper to describe the system at hand;

b) there are other factors at play – for example, conformational strain – which considerably affect the overall cluster stability but do not reflect on the overall hydrogen bond strength;



Plot 7 Sum of the hydrogen bond energies for each dissociated and undissociated hydrate, calculated according to the aforementioned correlation function, plotted against the number of H₂O in the cluster

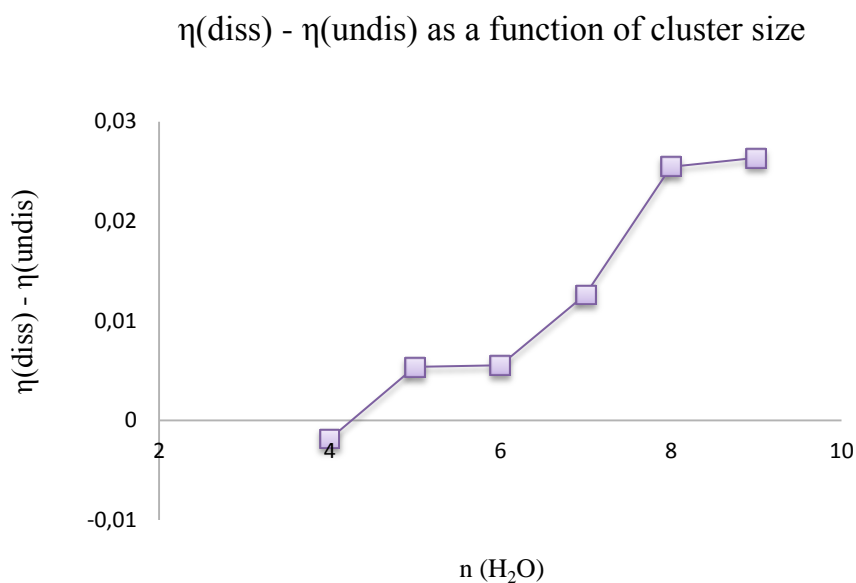
Plot 8 shows the total hydrogen bond energies of the various cluster systems studied herein plotted against the respective interaction energies. The first feature to catch our attention is the large value of the y intercept in the linear regression for dissociated hydrates – almost 575 kJ/mol. The chemical meaning of this constant is the average amount of interaction energy not attributed to hydrogen bonds. In contrast, the y intercept for undissociated hydrates is less than 20 kJ/mol. This small amount is easily attributable to London dispersion energy. In dissociated hydrates, the “excess” energy may be attributable to the isotropic coulombic attraction, that is, the non-directional coulombic attraction which does not strengthen hydrogen bonds. Such a strong contribution from coulombic forces arises from the uneven charge distribution in dissociated clusters.



Plot 8 Sum of the hydrogen bond energies for each dissociated and undissociated hydrate, calculated according to the aforementioned correlation function, plotted against the corresponding energy of interaction E_{int}

4.3.5 Chemical Hardness and stability

Chemical hardness can be estimated as half the HOMO-LUMO gap of a given chemical system. According to the principle of maximum hardness^[75], the greater the gap, the more stable is the system, since it is harder to polarize its electron cloud. Plot 9 displays the difference in chemical hardness of dissociated and undissociated hydrates as a function of cluster size. As more water molecules are added, this difference increases, meaning that dissociated clusters become increasingly more stable than their undissociated counterparts, which is in total accordance with the information provided by the difference of free energies for both systems, displayed in Plot 2.



Plot 9 Variation of the difference in chemical hardness between dissociated and undissociated clusters with cluster size

4.3.6 Eigen vs. Zundel structures

The structure and behavior of the hydrated proton is still a matter of intense debate. The old notion of a H_3O^+ cation weakly solvated by water molecules has been discarded in favor of more realistic representations: the Eigen (H_9O_4^+) and Zundel (H_5O_2^+) cations^[76].

In the Eigen cation, 3 water molecules are strongly hydrogen bonded to a central H_3O^+ moiety while in the Zundel cation a proton is shared by two water molecules, equidistant to each of them, as displayed in Fig. 10.

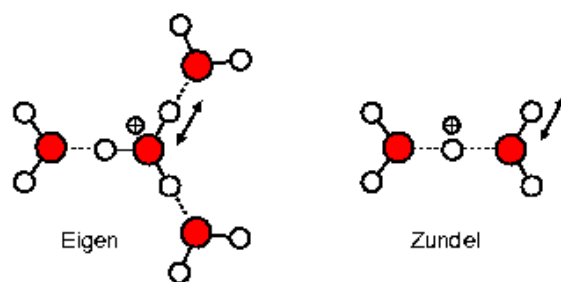


Fig. 10 Topologies of the Eigen and Zundel complexes. (Image taken from <http://www.science.uva.nl/research/molphot/trvs/liq.html>)

In the minimum energy structures of protonated water clusters, as well as acid hydrates, both Eigen and Zundel forms may be found, depending on the size and geometry of the cluster and nature of the solute. The interconversion between Eigen and Zundel cations is thought to be the basis of the Grothuss proton diffusion mechanism, an hypothesis supported by the existence of intermediate structures deemed "deformed Eigen" and "deformed Zundel" cations. The structures and vibrational frequencies of both conformations are well documented ^[77-78].

In the TFA hydrates studied herein, the hydrated cation is of the deformed Eigen type, as illustrated in Fig. 11. An assymetric H_3O^+ core is coordinated with either 3 water molecules ($n=4$ and $n=7-9$) or with 2 water molecules and the carboxylate moiety of trifluoroacetate. The O...O distance between H_3O^+ and the H_2O molecules or COO^- moieties directly hydrogen bonded to it in TFA hydrates are close to those found in the undistorted Eigen complex, which typically vary between 2.50 and 2.55 Å. The H_9O_4^+ cation is highly symmetrical since hydronium is solvated by 3 water molecules. On the contrary, the eigen complex present in TFA hydrates is uneven, since one of its O...O lengths is consistently 0,1 Å shorter than the other two. Unsurprisingly, the shorter distance is that between H_3O^+ and COO^- , in those clusters where a contact ion-pair is formed, or between H_3O^+ and the H_2O that is directly bonded to COO^- . Thus, the electrostatic pull of the carboxyl group affects the eigen complex even when hydronium resides in the second solvation shell.

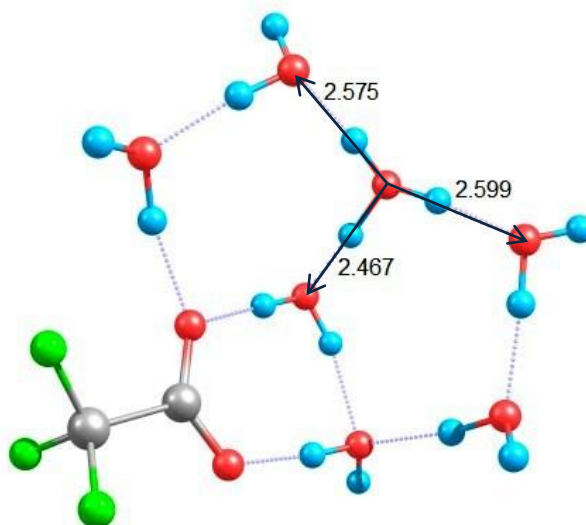


Fig. 11 Structure of $\text{TFA}\cdot(\text{H}_3\text{O}^+)\cdot(\text{H}_2\text{O})_6$. The values shown correspond to the O...O distances of the Eigen complex

There is no evidence for Zundel-type structures in minimum energy structures of TFA hydrates. However, metastable conformers corresponding to transition states (TS) display distorted Zundel complexes. The latter are characterized by O...O distances very similar to those found in undistorted Zundel ions. However, the shared proton is not equidistant to the solvation water molecules but rather prefers to sit $\sim 0,1$ Ang closer to the water molecule that becomes an hydrogen bond acceptor in the corresponding minimum. The sequence undissociated TFA : TS : dissociated TFA provides a visual representation of the Grothuss mechanism, as shown in Fig. 12.

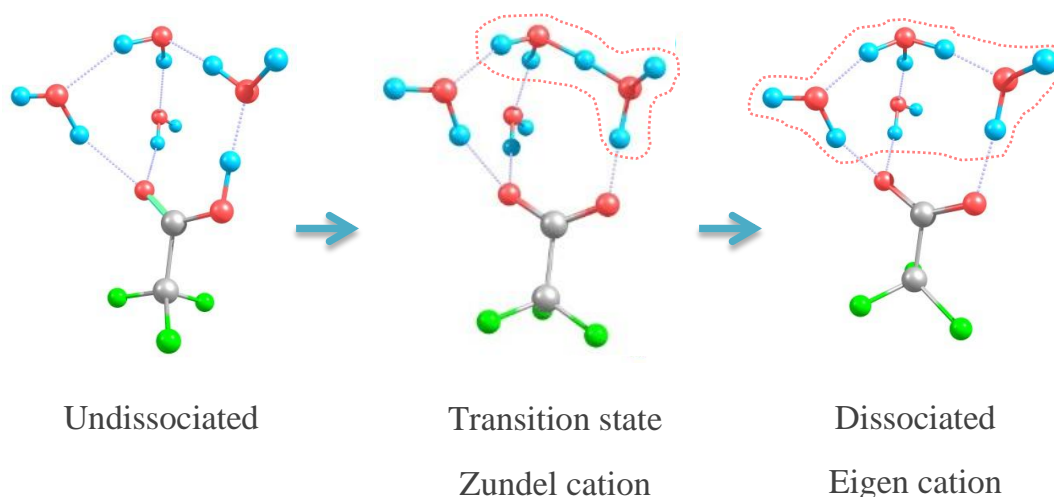


Fig. 12 Proton transfer pathway for the $n=4$ TFA hydrate, from the undissociated energy minimum, up to the transition state which forms a distorted Zundel complex and downhill again unto the dissociated hydrate in which an Eigen complex is found

4.3.7 Bridging the gap between micro- and bulk solvation

Do gas-phase hydrates really provide clues for understanding bulk dissociation or are they an entirely separate world?

Although the study of cluster systems is often pointed out as a useful probe for understanding the bulk state, one may argue that those are two entirely different worlds. The geometry of the TFA solvation shell in gas-phase clusters, for example, is likely to be quite different from its equivalent in the bulk since the chemical environment is completely changed. However, some fundamental features do remain constant. In the specific case of TFA hydration, for example, a few links connecting the nano- and macro- dissociation phenomena can be found. As pointed out in the introduction, TFA dissociation involves the formation of a contact ion-pair intermediate. The hydrates with $n=5$ and $n=6$ presented herein may thus represent the gas-phase counterparts of the contact ion-pair intermediates found in bulk, while the larger hydrates, in which hydronium occupies the second solvation shell, illustrate a possible solvent separated structure. Furthermore, according to a large angle x-ray scattering (LAXS) study by Takamuku et. Al ^[79], the first solvation shell of trifluoroacetate in bulk water is estimated to contain 4 water molecules, on average. Such

is the condition of the dissociated hydrates with $n=8$ and $n=9$. This information confirms our conclusion that this size range is the minimum required to fairly represent the bulk picture. In a nutshell, gas-phase hydrates can never be regarded as small scale replicas of the bulk, nevertheless they do provide insightful clues as to the nature of the latter.

4.4 Conclusion

The most important findings of this study are, in a nutshell

- A minimum number of 4 water molecules is necessary to abstract a proton from TFA although at this cluster size dissociation is energetically unfavorable;
- 5 water molecules are needed to render the dissociated TFA hydrate more stable than its undissociated counterpart;
- The hydronium ion, in all dissociated TFA hydrates, is found in the Eigen complex form;
- In TFA hydrates with 5 to 9 water molecules, the predominant geometric motifs are fused pentagonal rings;
- In large dissociated TFA hydrates ($n=8,9$), the solvation shell exhibits characteristics which coincide with those found for aqueous dissociation, namely, the hydronium cation is located in the second solvation shell and the carboxylate moiety is tetracoordinated with 4 H_2O molecules. Smaller hydrates ($n=5,6$) form contact ion-pairs, reported to be intermediate structures in TFA dissociation^[67].

The major weakness of this study lies in the absence of a thorough PES search. Due to limited time and computational resources, a narrow conformational search was performed. Thus, there is no way to ascertain if the local minima herein found are anywhere close to the global minima in the PES.

Additionally, the perspective of TFA dissociation presented is a static one. It compares two possible configurations, one in the dissociated and another in the dissociated

state. A dynamic simulation would be necessary to determine the dissociation time scale and more effectively identify metastable states involved in the process.

Thus, my recommendations for those interested in further studying TFA hydrates is to perform *ab-initio* MD simulations, which simultaneously allow a more thorough conformational search and probing the time scale of proton transfer events.

Chapter 5

Calculation of the pKa of organic acids using a hybrid explicit-implicit scheme

5. Calculation of the pKa of organic acids using a hybrid explicit-implicit scheme

5.1 Introduction

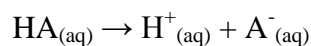
Acid-base phenomena are ubiquitous in both natural and industrial processes. The rate determining step in many chemical reactions involves proton transfer, hence their efficiency is correlated with ease of protonation. The pharmacological properties of bioactive compounds depend largely on whether the former are in their neutral or protonated/deprotonated states. Hence, when screening drug candidates, the ease of proton dissociation is usually the first parameter to account for^[80].

Although there are several experimental techniques for measuring acidity, such as spectroscopy, potentiometry, conductimetry and titrimetry there is an urgent need for reliable computational methods for acidity prediction. Theoretical prediction is especially important in drug design, which demands a fast screening of drug candidates, in the assessment of specific acidic sites in large biomolecules, such as enzymes, and for those species for whom experimental pKa measurement is not feasible such as metastable intermediates or very strong and very weak acids^[59].

This section briefly reviews current theoretical methods for acidity prediction, starting with the basic definitions and equations they employ.

5.1.1 Equations and definitions

The aqueous acidity constant, K_a , is defined as the equilibrium constant for the dissociation of an acid species HA into its conjugate base A^- and a proton H^+ in water. For the sake of illustrating the concept, the proton is usually represented as H^+ although in reality it is transferred to a nearby water molecule.



K_a is usually expressed in the logarithmic scale as pK_a .

$$pK_a = -\log K_a$$

pK_a may be predicted using *ab-initio* methods either directly, using thermodynamic cycles to calculate the Gibbs free energy of dissociation in aqueous solution (ΔG^*_{aq}) or indirectly, using parameterized models relating the pK_a of a given class of compounds with molecular descriptors such as bond lengths and energies. The former strategy is, in our perspective, more interesting since pK_a is truly calculated *ab-initio*, not requiring a parametrized model built with experimental data.

We do acknowledge that, at this time, correlation models offer a better performance at a typically lower computational cost than *ab-initio* methods. The latter are interesting to us not due to greater performance, since the contrary is the case, but rather due to the core idea behind them: if *ab-initio* methods are capable of accurately predicting ΔG^*_{aq} then the calculated pK_a must converge with the experimental one. A perfect and universal convergence between theory and experiment, preferably at moderate computational expense, would be the ultimate triumph of computational chemistry. The current scenario is a far bleaker one, however the shortcomings of *ab-initio* methods in pK_a prediction are valuable symptoms of the underlying sources of error, be they intrinsic calculation errors or unrealistic dissociation frameworks. Thus, only *ab-initio* methods of pK_a prediction shall be further discussed, as correlation models fall out of the scope of the present work.

5.1.2 Thermodynamic cycles or how to obtain $\Delta G^*_{(aq)}$

The acidity constant K_a may be calculated from the standard state free energy change of the dissociation reaction according to

$$\Delta G^*_{aq} = -RT \ln K_a$$

which may be rearranged to directly yield pK_a as

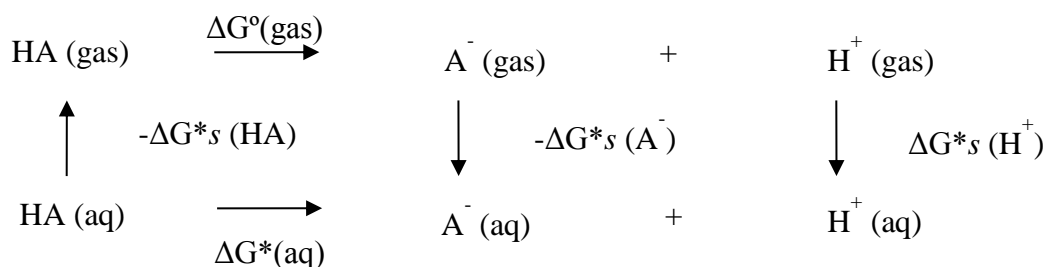
$$pK_a = \frac{\Delta G^*_{aq}}{RT \ln 10}$$

The quantity ΔG^*_{aq} is the free energy difference between products and reactants in aqueous solution at 298K.

Unfortunately, the calculation of free energies in solution is still difficult. Typically, continuum solvation models such as the Polarizable Continuum Model (PCM) or the recently proposed Solvation Model D (SMD)^[81], are used to implicitly model the solvent, since the explicit alternative would be computationally inefficient.

Continuum solvation models^[82], also named implicit solvation models, provide a way to efficiently describe how a given solute responds to a solvent by substituting the latter with a dielectric continuum whose dielectric constant is that of the solvent of interest. Thus, the solute is treated with atomic detail while confined in a molecule-sized cavity which is surrounded by the dielectric medium. Both PCM and SMD belong to a popular class of implicit solvation models in which the dielectric medium responds to the charge distribution of the solute by undergoing polarization.

Although time-efficient and practical, continuum solvation models are not able to predict free energies in solution with enough accuracy. This is problematic, since an error of merely 5.7 kJ/mol in the estimation of the free energy of dissociation gives an error of 1 pKa unit. In order to promote accurate results, ΔG^*_{aq} may be calculated through a thermodynamic cycle. There are many possible thermodynamic cycles, the simplest being the “direct method”, displayed in Scheme 2.



Scheme 2 Direct thermodynamic cycle

Firstly, the geometries of reactants and products are optimized in the gas phase and, subsequently, their free energies in solution are computed. For best results, a re-

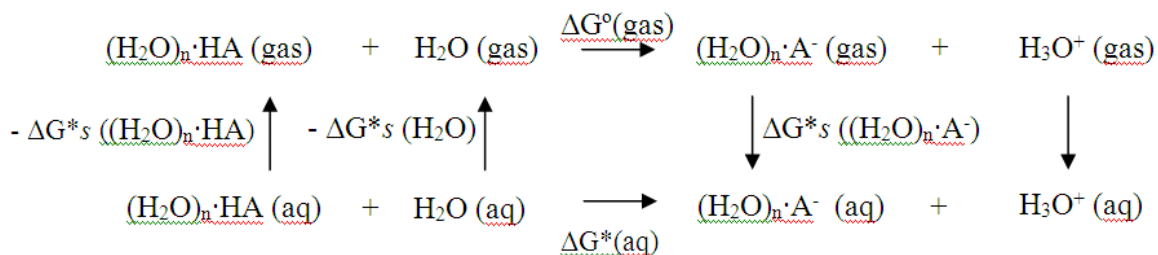
optimization in the solvent model should be performed. The free energy differences between solutes in the gas and solution phase provides the solvation free energy ΔG^*_s . The advantage of using a thermodynamic cycle instead of simply calculating the free energy difference in solution is that the gas-phase calculation, being less time consuming, is performed at a higher level of theory than the calculations in a dielectric continuum. It is expected that the high accuracy of the ΔG° value somehow compensates for the inherent error of the solvation free energies.

At this point, $\Delta G^*(aq)$ would be easily calculated by following the thermodynamic cycle. But how to account for $\Delta G^*_s(H^+)$? A proton has no electrons, thus cannot be submitted to calculation. Instead, experimental values have been used. There is a large discrepancy between the several reports of experimentally determined $\Delta G^*_s(H^+)$ [83-84] and that is an additional source of error when this basic thermodynamic cycle is used. The problem is avoided by adding one water molecule to the reactants and substituting H^+ for H_3O^+ in the products.

The basic thermodynamic cycle shown in Scheme 2, the direct model, is the most commonly encountered in the literature. It is fairly accurate at predicting *relative* pKa values, however fails at estimating absolute values. This indicates that the dissociation free energy calculated using this scheme is strongly correlated to pKa, however persistent systematic errors lead the calculated absolute values to deviate from the experimental ones.

Many alternative thermodynamic cycles have been proposed as an attempt to surpass the aforementioned limitations [59, 80, 85-86]. Out of the proposed models, the cluster-continuum hybrid approach [59, 87-88] has caught our attention. It includes explicit solvation water molecules in order to better account for the interactions between solute and solvent, which are poorly estimated by continuum solvation models. The latter is especially true for ionic species, since their unbalanced charge distribution evokes a more “dramatic” electrostatic response from the bulk – for example, charge transfer phenomena – than neutral species. Aqueous solvation free energies calculated using continuum models have errors of up to 4 kJ/mol for typical neutral solutes and ~17 kJ/mol for ionic solutes [59]. Consequently, ionic species frequently require a different parametrization of the continuum model than the one used for neutral species. This is quite bothersome and the explicit representation of solvent molecules is a more straightforward way to improve the

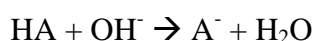
representation of ionic solutes. In the latter case, the solute encased in the continuum solvent cavity is no longer the bare acid/base but rather its microsolvated version. The cluster-continuum thermodynamic cycle is presented in Scheme 3.



Scheme 3 Cluster-continuum thermodynamic cycle

Many variations of this cycle have been used. While some authors explicitly solvate both the acid and the conjugate base, others decide to do so only for ionic solutes. The explicit solvation of the hydronium ion to form, when in aqueous media, the Eigen complex, has also been tested.

The cluster-continuum model may be applied in the direct way, as depicted in Scheme 3 or using other dissociation schemes, such as the proton-exchange model, in which proton transfer takes place between the acid of interest and the hydroxide ion, as follows

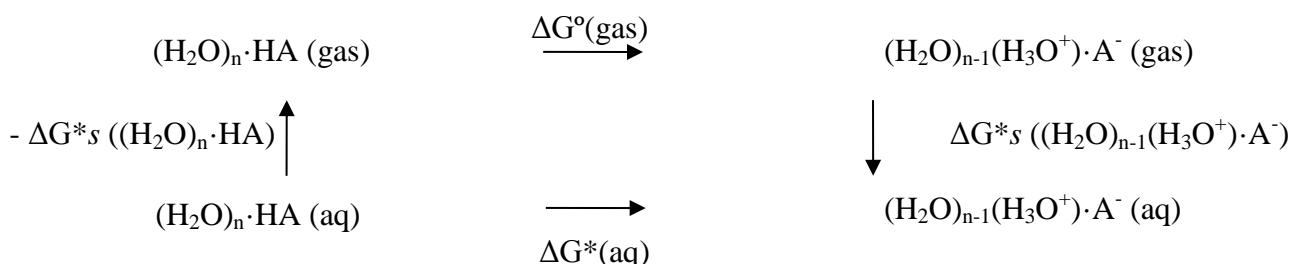


The advantage of the proton-exchange model is that ionic species are present on both sides of the equation promoting the cancellation of errors.

The cluster-continuum method has been applied to carboxylic acids ^[89] yielding better results than those usually obtained through the direct method. A major drawback of the cluster-continuum method is that the calculated pKa depends on the number and location of water molecules of solvation. The number of water molecules that yields a better pKa estimate varies with each acidic compound. Hence it is difficult to determine *a priori* which number represents bulk solvation ideally. However, Cramer, Kelly and

Truhlar^[90] have shown that adding a single water molecule significantly improves the pKa prediction for compounds which interact strongly with the solute. If the solute-solvent interaction is weak, the cluster-continuum method does not perform better than the direct method. Since TFA strongly interacts with water it is reasonable to include explicit solvent molecules in the pKa calculation.

Despite the elegance of the method discussed above, one detail has spiked our curiosity: why are the anionic and cationic species resulting from dissociation always treated separately? The reason for this cognitive dissonance stems from our studies with TFA hydrates, in which aggregates of neutral TFA and water are compared with their analogs in the dissociated state. The free energy difference between undissociated and dissociated clusters of the same size is undoubtedly linked with pKa. With this notion in mind, we have built yet another thermodynamic cycle in which dissociation occurs within the cluster, as shown in Scheme 4.



Scheme 4 Our proposed thermodynamic cycle

Here, ΔG° is free energy difference between the undissociated and dissociated clusters in the gas phase,

$$\Delta G_{gas}^0 = G_{gas}^0((\text{H}_2\text{O})_{n-1}(\text{H}_3\text{O}^+)(\text{A}^-)) - G_{gas}^0((\text{H}_2\text{O})_n(\text{HA}))$$

And ΔG^*_{aq} is computed by adding or subtracting the solvation free energies ΔG_s^* .

$$\Delta G_{(aq)}^* = \Delta G_{gas}^0 + \Delta G_s^*((\text{H}_2\text{O})_{n-1}(\text{H}_3\text{O}^+)(\text{A}^-)) - \Delta G_s^*((\text{H}_2\text{O})_n(\text{HA}))$$

The use of this thermodynamic cycle only makes sense if one computes the gas-phase dissociation free energy at a higher level of theory than the one used for the

calculation of solution-phase energies. We decided to skip this step altogether and simply computed the solution-phase energies of both dissociated and undissociated clusters, so that $\Delta G^*(aq)$ results directly from

$$\Delta G_{(aq)}^* = G_{(aq)}^*((H_2O)_{n-1}(H_3O^+)(A^-)) - G_{(aq)}^*((H_2O)_n(HA))$$

5.1.3 Comment on the nature of pKa

As the results will demonstrate, our proposed thermodynamic cycle provides a very satisfying degree of accuracy. But is it more accurate because it is more realistic or because it promotes a better cancellation of errors? Is it a simple artifact? Only by applying the proposed method to a wide variety of acid compounds with known pKa may those questions be safely answered. It is however amusing to ponder on the nature of pKa at a fundamental level. How is dissociation defined? Is an acid species fully dissociated only if the resulting anion and cation are infinitely apart? What if they form a contact ion pair, does that count as dissociation? What if there is only one solvent shell separating them?

Smiechowski *et al* ^[56] performed an analysis of carboxylic acid hydrates similar to our own and calculated the equilibrium constant for the process $(H_2O)_nHA \rightarrow (H_2O)_{n-1}(H_3O^+)(A^-)$ in aqueous solution but did not use it to calculate the pKa of the respective acids nor did they ever refer such equilibrium constant to be the acid dissociation constant. The latter is always treated in pKa calculation studies as the equilibrium constant for the process where the conjugate base and hydronium ion become fully separate and are thus computed separately.

Obviously, in the bulk solution, the proton rapidly hops across the hydrogen bond network after dissociation and in a short period of time distances itself from the anion but the question is, in average, how distant?

A good starting clue can be found in the recent work of Gu, Frigato, Straatsma and Helms ^[62] who performed an MD simulation of acetic acid dissociation in water. An acetic acid molecule was placed in a simulation box with explicit water molecules and the

trajectory of the acetic acid proton was then followed for 50 ns. During this time-window, several proton transfer (PT) events took place.

In more than 90% of the PT events, the proton hopped to the closest water molecule but quickly returned to the acetate ion, protonating it again. In the remaining cases the proton managed to escape and travel through the hydrogen bond network but eventually hopped back to the acetate ion. As the radial distribution function of this simulation evidences, the preferred position of the proton is around 2 Å from the carboxylate oxygen, corresponding to the case of fast proton swapping with the closest water molecule, and it goes no further away than 12 Å upon escaping the pull of the acetate ion.

Bearing in mind that many of the biologically interesting acids being studied today are as strong as, and in many cases weaker than, acetic acid, it does not seem at all reasonable to assume that upon dissociation the resulting hydronium and conjugated base are infinitely separated. Thus we believe that even though the proposed scheme may not capture the nature of pKa as it has been traditionally defined, it might actually provide a more realistic description of the chemical environment of conjugated bases after dissociation takes place.

5.2 Computational methods

The geometries of all studied hydrates and their dissociated counterparts were optimized at the SMD/B3LYP/6-311++G** level with default parameters using the Gaussian09^[47] electronic structure calculation package. SMD is a universal solvation model – here the term “universal” denotes its applicability to all solvents – recently proposed by Marenich, Cramer and Truhlar^[81] and recommended by the Gaussian09 manual for the computation of free energies of solvation.

Vibrational analysis confirmed, by the absence of imaginary frequencies, that all the considered structures correspond to local minima on the PES.

The Gibbs free energy of all hydrates was calculated at 298K and 1 atm. For the calculation of the Gibbs free energy Gaussian09 assumes the system behaves as an ideal gas and that its first and higher excited states are inaccessible. The software uses standard statistical thermodynamic formulae in order to obtain the electronic, translational, rotational and vibrational contributions to entropy and enthalpy and then retrieves the Gibbs free energy through the equation $G=H-TS$ (for detailed information please consult ref ^[68]).

The greatest source of error in the calculation of thermodynamic quantities regards the vibrational contribution, which is the largest. Those “vibrations” whose frequency is lower than 625cm^{-1} may not be true vibrations but internal rotations and their treatment as vibrations induces error. When such low frequencies are present, as is the case of all clusters in our data set, it is recommended that one treats them separately. After doing so, we found that the newly estimated pKa values were less accurate than those obtained including the low frequencies. For this reason, the pKa values here reported were calculated using the Gibb’s free energies directly provided in the calculation output.

5.3 Results and discussion

5.3.1 Global performance assessment

Table 5 List of calculated and experimental pKa values, along with individual errors (ΔpK_a), mean unsigned error (MUE) and mean signed error (MSE)

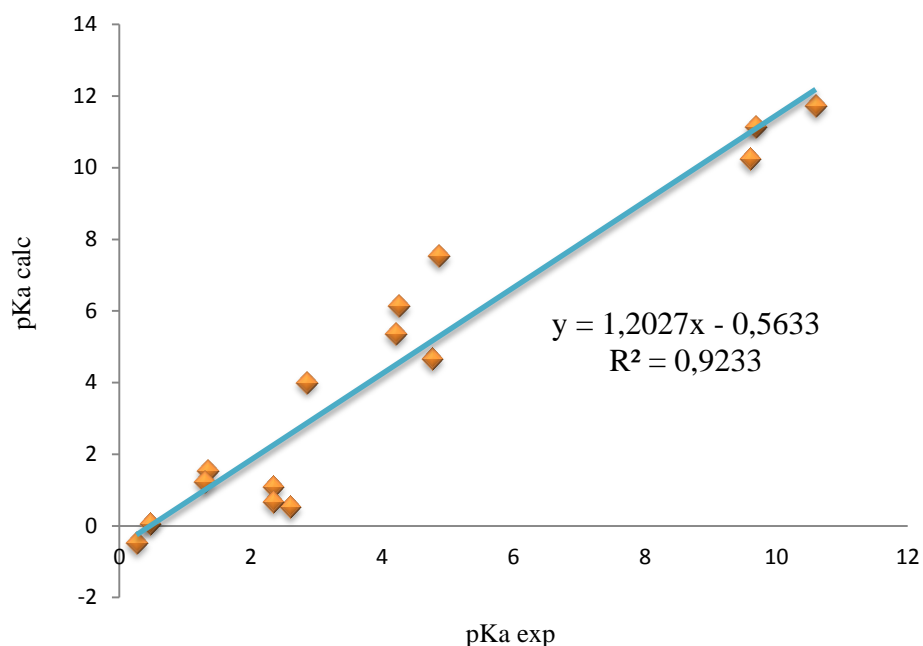
Acid	pKa (calc)	pKa (exp)	ΔpK_a
Acetic acid	4,67	4,76	-0,09
Propanoic acid	7,53	4,86	2,67
Trifluoroacetic acid	0,05	0,47	-0,42
Difluoroacetic acid	1,55	1,34	0,21
Fluoroacetic acid	0,52	2,60	-2,08
Trichloroacetic acid	-0,47	0,26	-0,73
Dichloroacetic acid	1,24	1,30	-0,06
Chloroacetic acid	4,01	2,85	1,16
Acrylic acid	6,13	4,25	1,88
Benzoic acid	5,36	4,20	1,16
Isopropylammonium	11,73	10,60	1,13
Glycine pKa1	1,10	2,34	-1,24
Glycine pKa2	10,24	9,60	0,64
Alanine pKa1	0,67	2,34	-1,67
Alanine pKa2	11,15	9,69	1,46
		MUE	1,11
		MSE	0,27

The pKa values of the tested set of 10 carboxylic acids, 1 amine and 2 aminoacids, estimated according to the proposed scheme are displayed in Table 5 along with the individual errors, mean unsigned error (MUE) and mean signed error (MSE). The greatest deviation for this group of carboxylic acids is 2.67 pKa units (propanoic acid) and the lowest merely 0.06 (dichloroacetic acid) while the MUE is 1.11, indicating reasonable accuracy. The last statement may sound quite extraordinary, as in practical terms, this error is one hundred times greater than that offered by any pH meter found in the laboratory. However, the art of estimating pKa values *ab-initio*, contrary to potentiometry, is still in its infancy, and the average MUE of the calculation schemes suggested in the literature is ~2 pKa units.

Although the MUE is a useful, quick measure of accuracy, some other indicators are important. It is customary to plot the calculated pKa's against their experimental values and perform a linear regression analysis. The slope, y intercept and correlation of the regression line provide valuable information. The ideal slope is 1, however it is common to find slopes of 0.5 especially among direct continuum methods. The y intercept, when large, indicates the presence of systematic errors in the calculation method. Finally, the correlation factor measures the consistency of each method's pKa predictions. If only systematic errors were present, the correlation would be close to 1 while the MUE could very well be large.

The slope of the regression line, shown in Plot 10, for our proposed method is 1.2 which is quite reasonable and the correlation 0.92, indicating there is significant but not strong correlation between the calculated and experimental results.

The suggested calculation scheme is reasonably accurate



Plot 10 Plot of the estimated pKa values against the experimental ones

When one proposes yet another route for first-principles pKa estimation, it is common practice to focus most of the discussion around the accuracy issue and how it compares with the accuracy of other schemes that have been used to test the same family of acids. More often than not, the discussion aims at promoting the proposed method to the detriment of the previously published ones instead of honestly attempting to discern which characteristics promote success in a calculation scheme. As a result, all sort of “smoke and mirrors” tricks are found in the literature that twist and cut calculated results, forcing them to adjust to experimental ones. Examples include using parametrization functions, “adjusting” the calculated solvated free energies by a fixed amount, using different experimental values for $G^\circ(\text{H}^+)$ and selecting the one that yields the best pKa estimates or, quite naively, declaring that the error is lower than 1 pKa unit without specifying that they’re referring to the Mean Signed Error, which is usually lower than the Mean Unsigned Error since the errors with opposite signs cancel each other out.

It is thus quite frustrating to conduct a serious performance comparison, one of the reasons why the following one is kept short.

Pliego and Riveros ^[87] used the cluster-continuum approach in conjunction with proton-exchange method to estimate the pKa's of 17 species. Explicit water molecules were added on both the hydroxide anion (reagents) and the conjugate base anion (products) although the number of added water molecules varied for each solute, having been selected to maximize the stability of said molecule in solution. With a MUE of **1.77** and a slope of **1.07**, Pliego's method performance is as good as ours. The author also estimated the pKa of the whole data set using pure continuum methods, which performed fairly worse than the cluster-continuum model.

Klamt and Eckert ^[91] calculated the pKa of 94 acids using the COSMO solvation model. They made 3 distinct predictions for each acid: without explicit water molecules, adding one or two water molecules. Surprisingly, the MUE and correlation of explicitly solvated systems was worse than those in the purely continuum method. The mono- and disolvated systems lead to correlation coefficients of **0.94** and **0.89** and MUEs around **2** pKa units while the purely continuum method yielded a MUE of **1.4** and correlation coefficient of **0.97**. Conversely, the slope was improved by the addition of water molecules, lowering from **1.5** in the pure continuum case to close to **1** in the explicitly solvated case. The authors state that the cluster-continuum method is a promising way of solving the slope problem, however new problems are found. For example, the loss of overall correlation with each added water molecule.

Sutton, Franks and da Silva ^[92], in a freshly published paper, present an extensive study on the performance of cluster-continuum schemes using the SMD solvation model. The testing set consists of 8 carboxylic acids whose solvation free energies are computed at the DFT level using different functional/basis sets combinations. 3 thermodynamic cycles are tested and 4 explicit hydration scenarios (including no explicit solvation, one water molecule on all species, one water molecule on all aqueous species and one water molecule solely on anions). All possible combinations were tested so that a total of 804 individual pKa values were computed. The MUEs varied from **0.6** to **7.08**.

The combination yielding the most accurate results, with MUEs below 1 pKa unit, is the computation of solvation energies at the M05-2X/cc-pVTZ level (M05-2X is a density functional known for its improved treatment of medium-range correlation) using explicit water molecules on all solutes. The absence of explicit solvation consistently leads to worse estimates but its application solely on the anions and not on the neutral species results in ever worse performance than when the pure continuum model is used.

A rare case of high accuracy must be here noted, even though it does not apply the cluster-continuum model. Liptak and Shields^[93] used the direct method to estimate the pKas of 6 carboxylic acids. These were optimized in the gas-phase at the HF level of theory and their single-point energy calculated using complete basis set methods, while using the conductor-like polarizable continuum model (CPCM) to calculate the solvation free energies. The best results are obtained when using the CBS-QB3 energies and stand less than ½ a pKa unit from the experimental values.

The estimation of the acidity constants of aminoacids has been receiving increasing attention in the past few years. These systems present a heavier burden due to their greater size and conformational freedom, thus requiring longer optimization steps.

Kiani and his team^[94] estimated the pKas of 3 aminoacids and 2 peptides using 0 to 4 explicit water molecules. For each dissociation constant, the ensemble resulting in the best estimate was selected. The resulting MUE is 0.3, a quite outstanding feat, although it was accomplished at a high computational expense due to the high number of solvation ensembles tested.

Vyas and Ojha^[95] estimated the pKa of 4 aminoacids using the direct thermodynamic cycle, without adding explicit water molecules. The aminoacid structures were optimized in the gas-phase at the B3LYP and HF levels with the 6-311++G(d) basis set and then solvated in the PCM framework. The resulting MUE was 1.3 for the pKa values calculated at the B3LYP level and 2.3 for the HF values. It must however be noted that several sets of pKa estimates were calculated using values of $\Delta G_{\text{sol}}(\text{H}^+)$ between -262 and -266 kcal/mol in order to choose the one which yielded the best pKa estimates for all aminoacids.

Mangold and his team ^[96] used density functional based molecular dynamics to estimate the pKa of 7 aminoacids. The novelty of this work is that it explicitly represents the bulk solvent (or at least a fraction of it) by simulating the aminoacid dissociation inside simulation cells with ~50 water molecules. Unfortunately, the uncertainty of the trajectories' finite lengths, of ± 1.1 pKa units, weighs heavily upon the overall MUE, which reaches 2.1 pKa units.

In this study only 2 aminoacids were studied, and the 4 resulting pKas were estimated with a MUE of 1.25 which is fairly reasonable given the current scenario.

Overall, the scheme herein proposed does not deliver an outstanding performance but a quite reasonable and balanced one. Another important point is that our scheme dispenses the gas-phase calculation step. This step is usually performed using high-level methods and large basis sets.

Our method bypasses the gas-phase calculation and simply takes the free energy difference of reactants and products in solution, calculated at a moderate level of theory. This "shortcut" is likely to be challenged, since the thermodynamic cycle has been adopted as the norm for the calculation of pKas and its indispensability is never questioned. Sutton, Franks and da Silva ^[92] argue *"As the solvent models have been parametrized on Gibbs energies of solvation, not solution phase reaction energies, more accurate results can be obtained when they are used to calculate the energy of solvation, in association with a gas basicity calculation performed at a higher level of theory"* This statement reflects the whole "Weltanschauung" of the research community interested in pKa determination. It seems, apparently, that our work contradicts this assumption since even without using a thermodynamic cycle the herein proposed scheme performs with accuracy comparable to that of other methods, which are without exception based on thermodynamic cycles.

The most important contribution of our method, however, is not an improvement in accuracy but a widening of perspectives, an opening of new pathways for pKa calculation and an additional exploration of the advantages and pitfalls of using explicit solvent water molecules. The latter issue is addressed in the following discussions, in which different solvation scenarios are compared and their effect on accuracy discussed.

5.3.2 On the optimum number of explicit water molecules

Hydrogen bonds play a pivotal role in acid dissociation events taking place in aqueous media. Hence, a good description of the solvation shell of the solutes improves the accuracy of the pKa estimate. The polarizable continuum model cannot efficiently account for hydrogen bonds between solvent and solute. In order to compensate for this deficiency one may add a few explicit water molecules in the solvation shell of the acidic moiety responsible for proton transfer (in carboxylic acids, the carboxyl group). Although this method has been oftentimes successful it is not yet clear how many explicit water molecules, and in which configuration, yield the best description for each acid species.

Kelly, Cramer and Truhlar ^[90] provide a short but illustrative example of this conundrum. They have calculated the pKa of the HCO_3^- anion while adding 0 to 3 water molecules to the CO_3^{2-} anion. Three continuum solvation methods have been used: solvation model 6 (SM6), solvation model 5.43R (SM5.43R) and the dielectric polarizable continuum model 98 (DPCM/98). For SM6, the results follow the intuitive notion that adding more explicit water molecules improves accuracy, as the latter increases in a stepwise fashion with each added solvent molecule. The other solvation models, however, behaved differently. For SM5.43R the best pKa estimate is given by the mono-solvated scheme while the best DPCM/98 result corresponds to the purely continuum scenario, and each explicit water molecule added progressively worsens the pKa estimate. Therefore, the optimum number of explicit water molecules for each particular case depends not only on the species but also on the continuum solvation model used and even, perhaps, on the level of theory.

This unpredictable behavior has also been reported by Ho and Coote ^[59] who used the cluster-continuum method, via proton exchange, to estimate the pKas of 32 acids using various continuum solvent models and 0 to 3 explicit water molecules. Their results are summarized in Fig. 13, where the Mean Absolute Deviation (MAD, the same as MUE) for each solvation model is plotted against the number of water molecules. Once again, when using the SM6 model each additional explicit water molecule lowers the MUE, as also

happen to a much lesser degree for IPCM. For the other models, however, the optimum number of water molecules is two and the addition of a third one renders the pKa estimates as bad as when no explicit water molecules are used. This divergent behavior of continuum solvation models undoubtedly arises from their considerably different parametrization.

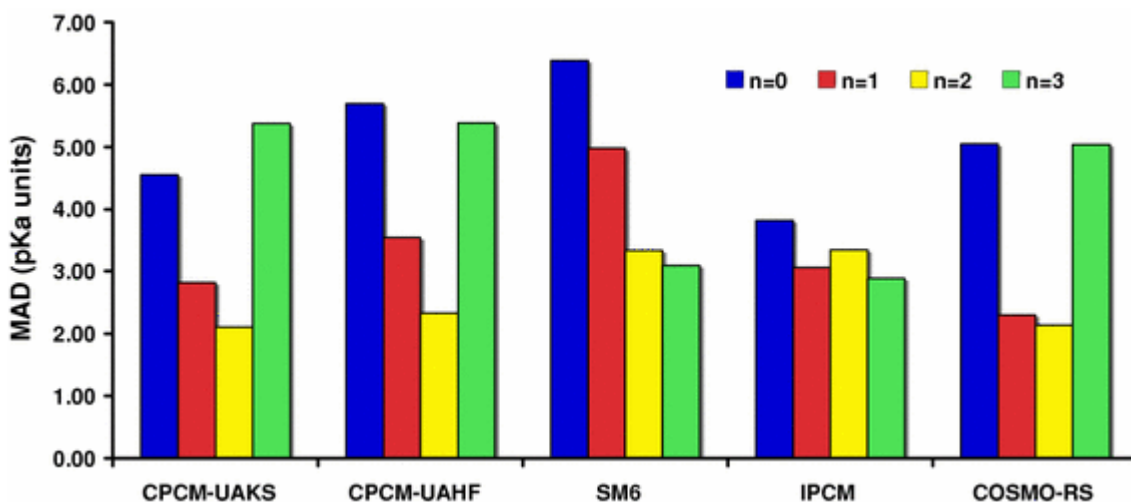


Fig. 13 Mean Absolute Deviation (MAD, same as MUE) for pKa values calculated within different solvation models and using 0 to 3 explicit solvation water molecules. From ^[59]

Zhang ^[88] estimated the pKa of a large number of species from different families, including carboxylic acids, alcohols and amines, using 0, 1 or 2 explicit solvation water molecules on ionic or both neutral and ionic species. His results suggest that, alongside with all the other variables the cluster-continuum method is subject to, the chemical class of the species in study also affects the effectiveness of the scheme. Adding explicit water molecules on both species improves the theoretical slope while not affecting the MUE and slightly decreasing R^2 for carboxylic acids, improves the theoretical slope while also increasing the MUE and decreasing R^2 in the group of alcohols and phenols and has no effect in the predicted pKa's of amines and anilines.

The following example from our own work illustrates a case in which the addition of a large number of explicit water molecules of solvation has a detrimental effect upon the pKa estimate. In the framework of our model, the explicit water molecules have the “task” not only of improving the description of the solvent but also of inducing proton transfer from the solute to the surrounding water molecule network.

As evidenced in the previous chapter, a minimum number of water molecules is needed to stabilize the hydronium ion in acid hydrates. As a general trend, the weaker the acid is, the higher the number of water molecules required to drive its dissociation.

Isopropylammonium is a very weak acid, with a pKa of 10.6, which lead us to assume it would require a large number of water molecules in order to dissociate.

A cluster with 8 water molecules, isopropylamine and an hydronium ion in the second solvation shell was then built and optimized in SMD. It proved to be a minimum and the pKa calculation was thus carried out according to our proposed method, yielding a pKa estimate which is 3,3 pKa units higher than the real value.

Could this large error be caused by excessive noise introduced in the calculation by the large number of explicit water molecules? Remember that we are attempting to accurately measure very small energetic differences between the dissociated and undissociated state. The higher is the dimension of the system the harder it gets to accurately measure such small differences.

Thus we started to exclude, one by one, the explicit water molecules that were not directly hydrogen bonded to either the hydronium ion or the amine/ammonium moiety in isopropylamine/ammonium. The result is displayed in Table 6 and Plot 11 which clearly show how the pKa estimate improves as the number of explicit water molecules decreases from 9 to 6.

Table 6 pKa values of isopropylammonium calculated using microsolvated systems with 6 to 9 explicit water molecules

Acid	pKa (calc)	pKa (exp)	Δ pKa
Isopropylammonium n=9	13,91	10,60	3,31
Isopropylammonium n=8	12,34		1,74
Isopropylammonium n=7	11,79		1,19
Isopropylammonium n=6	11,73		1,13

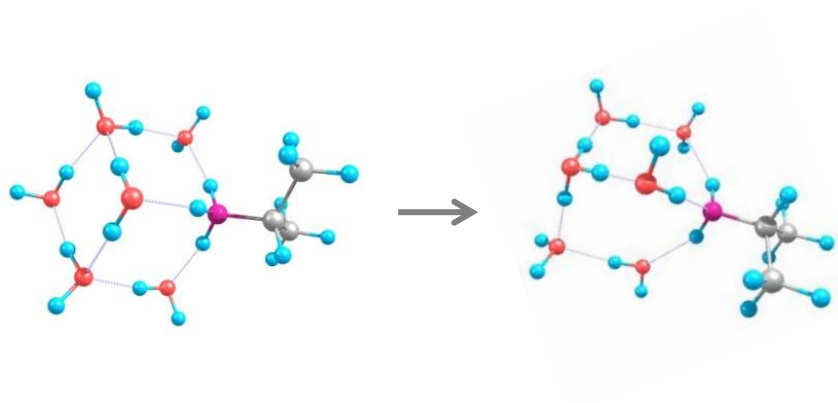
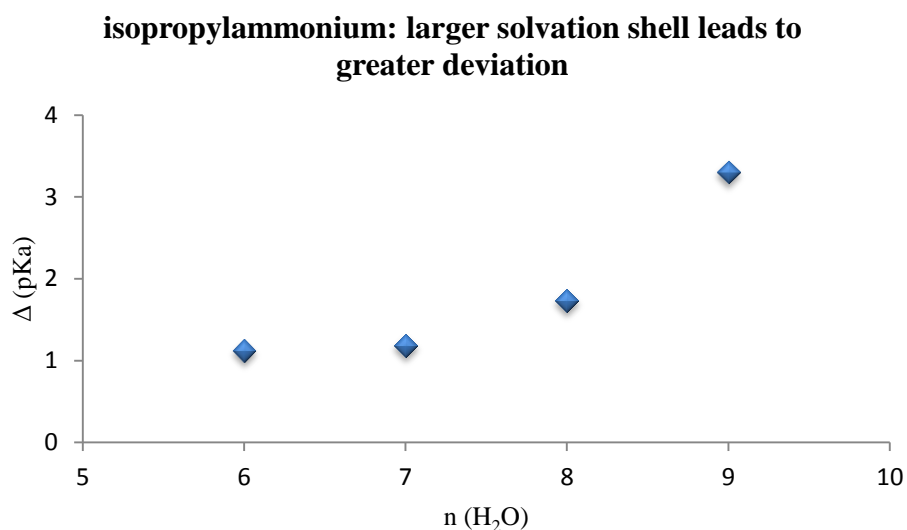


Fig. 14 Pair of undissociated and dissociated isopropylammonium hydrates with 6 water molecules, the ensemble that yields the best pKa estimate



Plot 11 Variation of ΔpK_a with the number of explicitly added water molecules for isopropylammonium

The balance is indeed delicate: an insufficient number of explicit water molecules does not accurately represent the solvent or allow proton transfer while a large number of water molecules increases the computational cost and introduces unnecessary noise into the calculation.

5.3.3 Conformation affects the pKa estimate: the case of acetic acid

Table 7 Estimated pKa values of acetic acid for two solvation shells with the same size but different conformation

Acid	pKa (calc)	pKa (exp)	Δ pKa
Acetic acid Struct. A	2,72	4,76	-2,04
Acetic acid Struct. B	4,67	4,76	-0,09

Geometry matters: given the same solvation shell size, some geometries yield a better pKa estimate than others

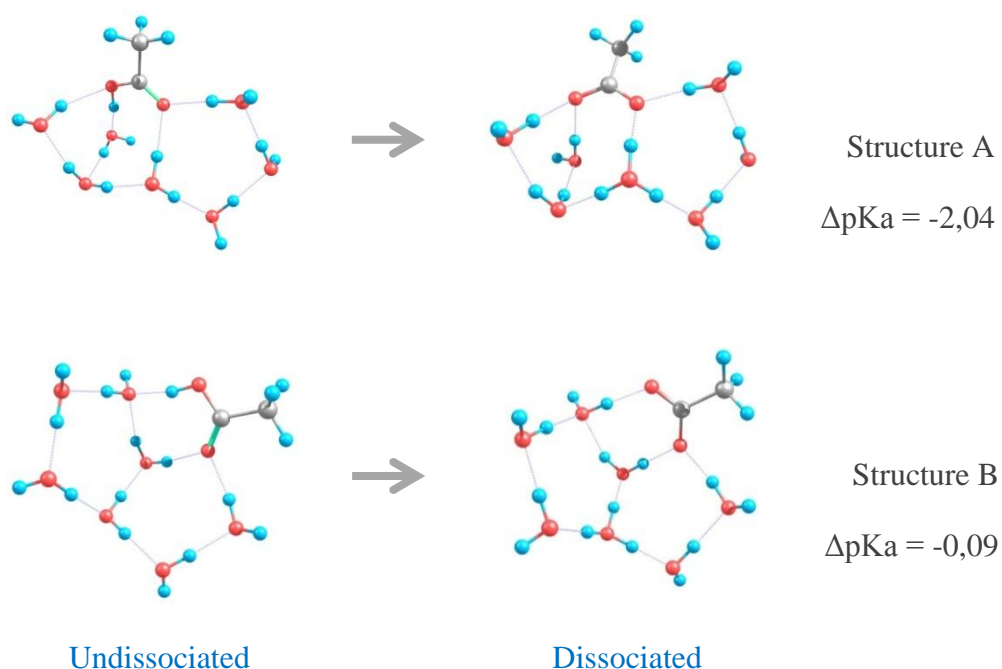


Fig. 15 Acetic acid hydrates with solvation shell structures A and B

The determination of the optimum number of explicit water molecules is not the only puzzle one must solve when applying the proposed method. The conformation water molecules adopt in the solvation shell also affects the accuracy of the dissociation constant calculation. Consider the following illustrative case: acetic acid hydrates with 7 water molecules where built with 2 different conformations, as shown in Fig. 15. One of these, labeled as “Structure A”, closely mimics one of the conformations proposed by Smiechowski and his team ^[56]. The authors analyzed the FTIR spectra of formic, acetic and propanoic acid in aqueous media containing isotopically diluted HDO. Through empirical correlations between stretching frequencies and hydrogen bond length, the oxygen radial distribution function for each system was determined. The vibrational analysis indicated the formation of a contact ion-pair between the hydronium cation and the carboxylate moiety and also that the latter forms a total of 4 hydrogen bonds with the solvent molecules. Several hydrates of the three carboxylic acids were then built and optimized at the B3LYP/6-311++G(d,p) level and the dissociated structure mimicked in the present work was chosen due to its close agreement with the aforementioned data. Recall the work of Gu and his team ^[62] mentioned in this chapter introduction which confirms that during proton hopping events in acetic acid dissociation, the hydronium cation is preferably located in the first solvation shell, forming a contact ion-pair with carboxylate moiety, and only occasionally hopping further to the second solvation shell.

The solvation shell in structure B is the same as the one in the TFA hydrates with $n=7$ proposed in the previous chapter. This structure represents the scenario where the hydronium ion resides in the second solvation shell. It is seemingly a poor representative of the most likely conformation of acetic acid in solution since it does not form a contact ion pair and the carboxylate moiety only establishes 3 hydrogen bonds with the surrounding water molecules.

One would easily assume, given all these considerations, that the pair of clusters with structure A, being supposedly a better representation of the first solvation shell of acetic acid in bulk media, would yield a better pKa estimate. Such is not the case, as this ensemble leads to a ΔpK_a of -2,04 pKa units while the dissociation scheme B yields a very accurate estimate, with an error of 0,09 pKa units.

5.3.4 A lesson from glycine: the importance of explicitly solvating all hydrogen bonding sites

When deciding which sites of a solute molecule should be explicitly solvated, one may be tempted to contemplate only the site that is directly involved in proton transfer events, in order to keep the number of explicit water molecules to a minimum and thus avoid large calculations. Although for simple molecules this is a fairly reasonable rationale, in some cases it may prove fruitful to consider the explicit solvation of all hydrogen bonding sites. This lesson is demonstrated by the example of glycine, one of the molecules in our data set. The first dissociation constant of glycine, which involves the carboxyl group, was first estimated using the ensemble shown in Fig. 16, on the left.

All hydrogen bonding sites should be explicitly solvated

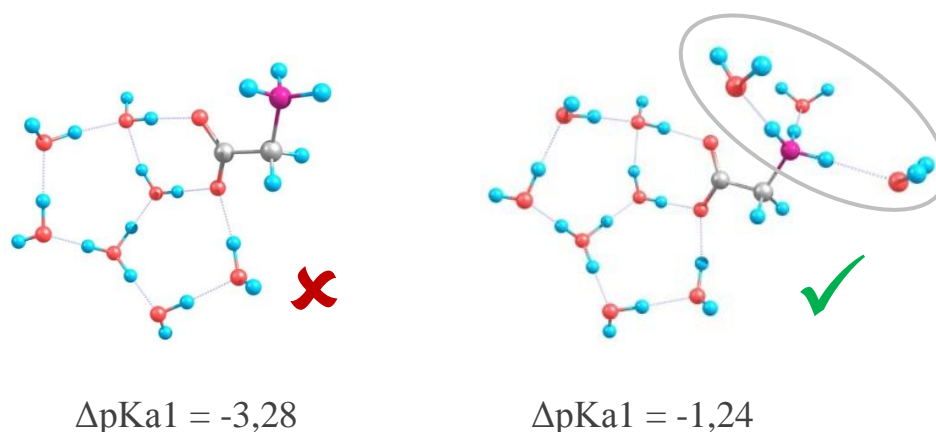


Fig. 16 Complete and partial microsolvation schemes for glycine and the consequent error in the pKa estimation

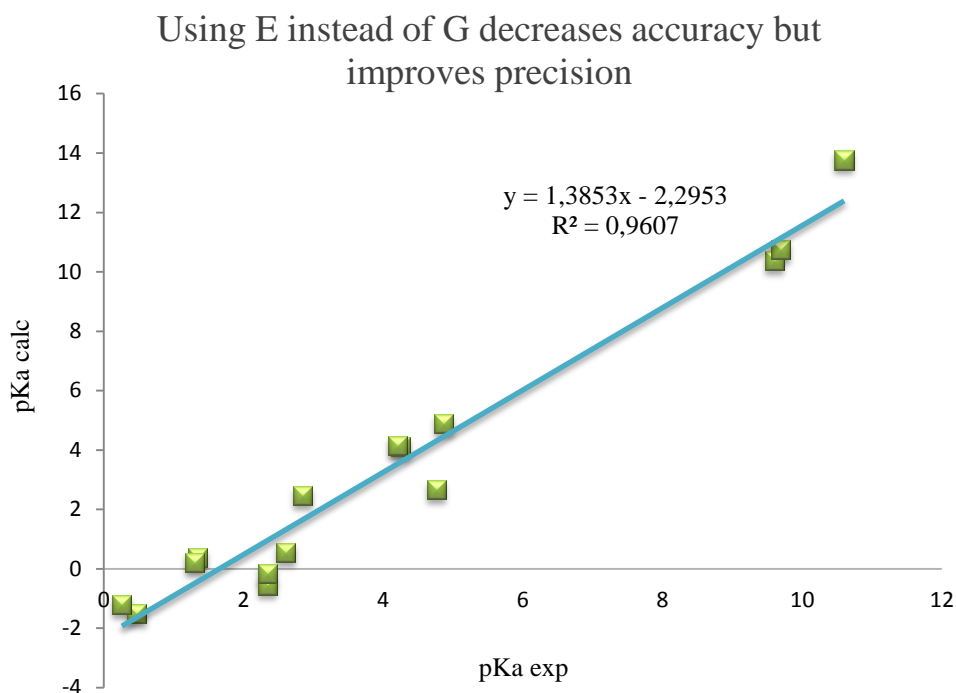
Since the NH_3^+ group is not involved in proton transfer at this stage no explicit solvent molecules were added to it. The resulting error was quite large, which prompted us to test whether the explicit solvation of the amino terminal would improve the pKa estimate. This assumption proved to be correct and the calculated pKa1 error was reduced by 2 pKa units. Hence, whenever possible, all hydrogen bonding sites within a given solute

should be explicitly solvated, as this will greatly improve the accuracy of the free energy calculation and, consequently, of the final pKa estimate.

5.3.5 Calculating “pKa” using the electronic energy of dissociation instead of the Gibbs free energy of dissociation

Table 8 List of experimental and calculated “pKa” values, estimated using the electronic dissociation energy, along with individual errors (Δ pKa), mean unsigned error (MUE) and mean signed error (MSE)

Acid	pKa (calc)	pKa (exp)	Δ pKa
Acetic acid	2,66	4,76	-2,10
Propanoic acid	4,89	4,86	0,03
Trifluoroacetic acid	-1,52	0,47	-1,99
Difluoroacetic acid	0,36	1,34	-0,98
Fluoroacetic acid	0,53	2,60	-2,07
Trichloroacetic acid	-1,21	0,26	-1,47
Dichloroacetic acid	0,19	1,30	-1,11
Chloroacetic acid	2,45	2,85	-0,40
Acrylic acid	4,10	4,25	-0,15
Benzoic acid	4,15	4,20	-0,05
Isopropylammonium	13,76	10,60	3,16
Glycine pKa1	-0,56	2,34	-2,90
Glycine pKa2	10,36	9,60	0,76
Alanine pKa1	-0,19	2,34	-2,53
Alanine pKa2	10,73	9,69	1,04
		MUE	1,38
		MSE	-0,72



Plot 12 “pKa” values calculated using the electronic dissociation energy plotted against the experimental ones

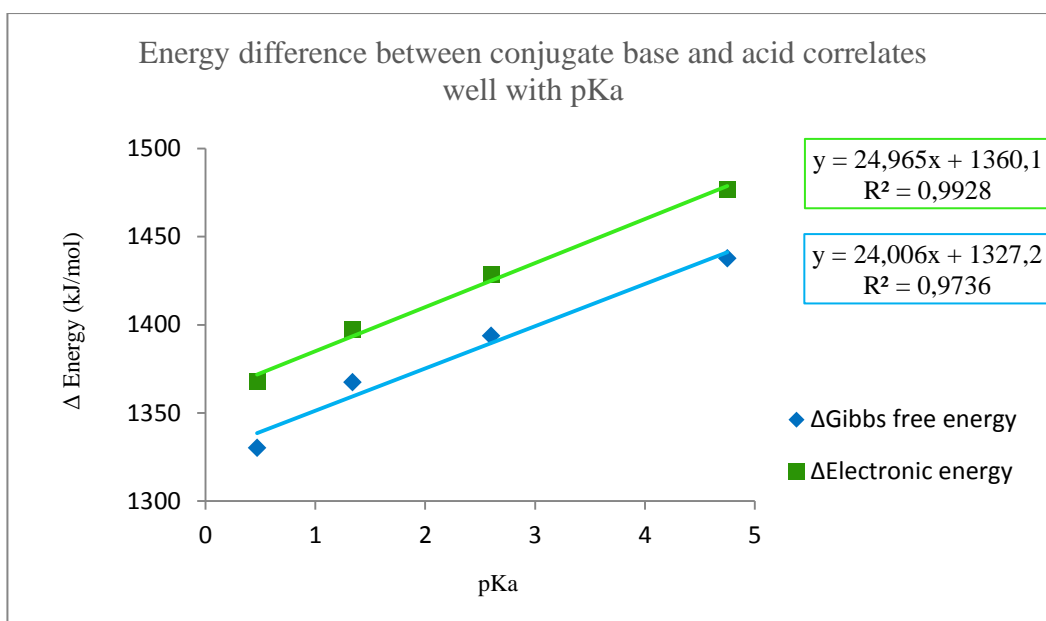
Motivated by sheer curiosity we calculated the “pKa” of the whole data set using the difference in electronic energy instead of the free energy. Of course, the calculated descriptor is not a true pKa, but we thought it would be interesting to access how well this measure correlates with the experimental pKa values.

The results are shown in Table 8 and Plot 12. Surprisingly, this route performs better at predicting relative pKa’s as evidenced by the high correlation ($R^2=0,96$) between the experimental and calculated “pKa”. However, the absolute deviation between these quantities is increased, which is to be expected since the electronic energy as opposed to the free energy was used for the “pKa” estimation.

The disturbing conclusion one may draw from this behavior is that the approximations made by Gaussian to estimate the free energy introduce a large amount of error in the calculation and this error is not systematic in nature. This situation is reminiscent of the typical scenario found for pKa estimation methods that use solely implicit solvation: the pKa ordering is fairly precise however the absolute calculated value falls short of the experimental one. Thus, the employed electronic structure calculations are

sensitive enough to qualitatively discriminate between solutes with different acidities but somehow fail to quantitatively predict the latter.

An additional clue suggests the larger error probably lies in the free energy calculation itself and not on the deficiencies of the continuum solvation models. The geometries of bare TFA, DFA, FA and AA and their respective conjugate bases have been optimized at the B3LYP/6-311++G(d,p) level of theory, in vacuum. Plot 13 displays the energy differences between the conjugate base and the acid plotted against the pKa. One set of data was calculated using the electronic energy while the other using the Gibbs free energy at 298K and 1 atm. Both correlate fairly well with pKa although using the electronic energy difference leads to a significantly better correlation than provided by using the Gibbs free energy. It is thus quite possible that the free energy, as estimated by Gaussian, is one of the weak links contributing to poor results in the *ab-initio* estimation of pKa values.



Plot 13 Gibbs free energy and electronic energy differences between the conjugate base and the acid, in vacuum, plotted against the experimental pKa of TFA, DFA, FA and AA.

5.4 Conclusion

pKa determination using *ab-initio* methods is a growing field, having still much challenges to be overcome. The cluster-continuum approach, as well as the model herein proposed, lead to a partial and debatable improvement in accuracy although entailing computationally intensive calculations and being strongly dependent on the number and topology of water molecules in the microsolvated clusters. A summary of the strengths and weaknesses of our proposed method is provided in Table 9.

The application of our method, which is moderately computationally intensive, has been limited by time and computational resources. Therefore, it was not possible to conduct a systematic study in which all acid hydrates are tested with the same of number of water molecules, in the same conformation, for many different cluster sizes and conformations. On the contrary, both cluster size and conformation were adapted to the characteristics of the solute (size, hydrogen bonding sites, ease of dissociation) and a systematic analysis is still lacking. Also, only one solvation model and level of theory was tested. Further explorations of the proposed method should seek to test these variables, as the literature suggests that the accuracy of cluster-continuum schemes is affected differently by each of them.

Further research is necessary in order to decide whether hybrid explicit+implicit models really constitute a viable avenue for development or merely stand as a quirky secondary road with no way out. Hopefully, our contribution has helped confusing matters even further and shown how tricky can be the development of a new pKa prediction method.

Table 9 Strong and weak points of our proposed scheme

Strengths	Weaknesses
No needed parametrization	Computationally demanding due to large system size
No use of experimental values	
No correction of energies	Variability of results with number of water molecules in the cluster and their conformation
No use of standard acid for relative pKa determination	
No replication of calculations (gas-phase, separately) although it might be argued that our scheme still takes longer	
No need for computationally expensive methods	
No need for thermodynamic cycle	

6. References

- [1] C. Seife, "Physics Enters the Twilight Zone", *Science* **2004**, *305*, 464-466.
- [2] R. F. W. Bader, "The Quantum Mechanical Basis of Conceptual Chemistry", *Monatshefte für Chemie / Chemical Monthly* **2005**, *136*, 819-854.
- [3] E. Schrödinger, "Quantisierung als Eigenwertproblem", *Annalen der Physik* **1926**, *384*, 361-376.
- [4] P. A. M. Dirac, "Quantum Mechanics of Many-Electron Systems", *Proceedings of the Royal Society of London. Series A* **1929**, *123*, 714-733.
- [5] D. Vasilescu, M. Adrian-Scotto, "From Democritus to Schrödinger: a reflection on quantum molecular modeling", *Structural Chemistry* **2010**, *21*, 1289-1314.
- [6] IUPAC, "Compendium of Chemical Terminology, 2nd ed. (the "Gold Book")", **1997**.
- [7] D. B. Boyd, in *Ullmann's Encyclopedia of Industrial Chemistry*, Electronic Release ed., Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, **2000**.
- [8] D. C. Young, in *Computational Drug Design: A Guide for Computational and Medicinal Chemists*, John Wiley & Sons, Inc., West Sussex, **2009**, pp. 71-74.
- [9] S. C. Smith, Q. Sun, "Molecular Modelling: All the Way from Atomistic Structure to Function in Complex Systems", *Australian Journal of Chemistry* **2010**, *63*, 343-344.
- [10] P. Tian, "Computational protein design, from single domain soluble proteins to membrane proteins", *Chemical Society Reviews* **2010**, *39*, 2071-2082.
- [11] H. M. Senn, W. Thiel, in *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations, Vol. 268* (Ed.: M. Reiher), Springer-Verlag, Heidelberg, **2007**, pp. 173-290.
- [12] C. Cramer, *Essentials of Computational Chemistry: Theories and Models*, 2nd ed., John Wiley & Sons Ltd, West Sussex, **2004**.
- [13] P. Echenique, J. L. Alonso, "A mathematical and computational review of Hartree–Fock SCF methods in quantum chemistry", *Molecular Physics* **2007**, *105*, 3057-3098.
- [14] P. Atkins, R. Friedman, *Molecular Quantum Mechanics*, 4th ed., Oxford University Press, Oxford, **2005**.

- [15] A. Hinchliffe, *Molecular Modelling for Beginners*, John Wiley & Sons, West Sussex, **2003**.
- [16] R. Grinter, *The Quantum in Chemistry: An Experimentalist's View*, John Wiley & Sons, West Sussex, **2005**.
- [17] A. Szabo, N. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 1st ed., Dover Publications, New York, **1989**.
- [18] H. J. Schneider, "Binding Mechanisms in Supramolecular Complexes", *Angew. Chem.-Int. Edit.* **2009**, *48*, 3924-3977.
- [19] K. E. Riley, P. Hobza, "Noncovalent interactions in biochemistry", *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2011**, *1*, 3-17.
- [20] J. H. Fendler, "Chemical self-assembly for electronic applications", *Chem. Mat.* **2001**, *13*, 3196-3210.
- [21] G. Chałasiński, M. M. Szczyński, "State of the Art and Challenges of the ab Initio Theory of Intermolecular Interactions", *Chemical Reviews* **2000**, *100*, 4227-4252.
- [22] K. E. Riley, M. Pitoňák, P. Jurečka, P. Hobza, "Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories", *Chemical Reviews* **2010**, *110*, 5023-5063.
- [23] S. F. Sousa, P. A. Fernandes, M. J. Ramos, "General Performance of Density Functionals", *The Journal of Physical Chemistry A* **2007**, *111*, 10439-10452.
- [24] C. D. Sherrill, T. Takatani, E. G. Hohenstein, "An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene–Methane, and Benzene–H₂S[†]", *The Journal of Physical Chemistry A* **2009**, *113*, 10146-10159.
- [25] J. H. Jensen, *Molecular Modeling Basics*, Taylor and Francis Group, Boca Raton, **2010**.
- [26] S. F. Boys, F. Bernardi, "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors", *Molecular Physics* **1970**, *19*, 553-566.
- [27] K. Morokuma, "Why do molecules interact? The origin of electron donor-acceptor complexes, hydrogen bonding and proton affinity", *Accounts of Chemical Research* **1977**, *10*, 294-300.
- [28] E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenber, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci, D. J. Nesbitt, "Definition of the hydrogen bond (IUPAC Recommendations 2011)", *Pure and Applied Chemistry* **2011**, *83*, 1637-1641.

- [29] R. Kumar, J. R. Schmidt, J. L. Skinner, "Hydrogen bonding definitions and dynamics in liquid water", *The Journal of Chemical Physics* **2007**, *126*, 204107-204112.
- [30] S. J. Grabowski, "What Is the Covalency of Hydrogen Bonding?", *Chemical Reviews* **2011**, *111*, 2597-2625.
- [31] G. Bertrand, "A reappraisal of what we have learnt during three decades of computer simulations on water", *Journal of Molecular Liquids* **2002**, *101*, 219-260.
- [32] R. Ludwig, "Water: From Clusters to the Bulk", *Angewandte Chemie International Edition* **2001**, *40*, 1808-1827.
- [33] K. Szalewicz, C. Leforestier, A. van der Avoird, "Towards the complete understanding of water by a first-principles computational approach", *Chemical Physics Letters* **2009**, *482*, 1-14.
- [34] F. N. Keutsch, R. J. Saykally, "Water clusters: Untangling the mysteries of the liquid, one molecule at a time", *Proceedings of the National Academy of Sciences* **2001**, *98*, 10533-10540.
- [35] B. J. Mhin, J. Kim, S. Lee, J. Y. Lee, K. S. Kim, "What is the global minimum energy structure of the water hexamer? The importance of nonadditive interactions", *The Journal of Chemical Physics* **1994**, *100*, 4484-4486.
- [36] S. S. Xantheas, J. T. H. Dunning, "Ab initio studies of cyclic water clusters (H₂O)_n, n=1--6. I. Optimal structures and vibrational spectra", *The Journal of Chemical Physics* **1993**, *99*, 8774-8792.
- [37] S. S. Xantheas, "Ab initio studies of cyclic water clusters (H₂O)_n, n=1--6. II. Analysis of many-body interactions", *The Journal of Chemical Physics* **1994**, *100*, 7523-7534.
- [38] B. Temelso, K. A. Archer, G. C. Shields, "Benchmark Structures and Binding Energies of Small Water Clusters with Anharmonicity Corrections", *The Journal of Physical Chemistry A* **2011**, *115*, 12034-12046.
- [39] X. Sotiris S, "Cooperativity and hydrogen bonding network in water clusters", *Chemical Physics* **2000**, *258*, 225-231.
- [40] J. C. Rasaiah, S. Garde, G. Hummer, "Water in Nonpolar Confinement: From Nanotubes to Proteins and Beyond *", *Annual Review of Physical Chemistry* **2008**, *59*, 713-740.
- [41] J. Hernández-Rojas, V. Monteseuro, J. Bretón, J. M. Gomez Llorente, "Water clusters confined in icosahedral fullerene cavities", *Chemical Physics* **2012**, *399*, 240-244.

- [42] Y. I. Neela, A. S. Mahadevi, G. N. Sastry, "Hydrogen Bonding in Water Clusters and Their Ionized Counterparts", *The Journal of Physical Chemistry B* **2010**, *114*, 17162-17171.
- [43] M. P. Hodges, D. J. Wales, "Global minima of protonated water clusters", *Chemical Physics Letters* **2000**, *324*, 279-288.
- [44] C. O. da Silva, E. C. da Silva, M. A. C. Nascimento, "Ab Initio Calculations of Absolute pKa Values in Aqueous Solution I. Carboxylic Acids", *The Journal of Physical Chemistry A* **1999**, *103*, 11194-11199.
- [45] K. R. Leopold, "Hydrated Acid Clusters", *Annual Review of Physical Chemistry* **2011**, *62*, 327-349.
- [46] L. Pu, Y. Sun, Z. Zhang, "Hydrogen Bonding in Hydrates with one Acetic Acid Molecule", *The Journal of Physical Chemistry A* **2010**, *114*, 10842-10849.
- [47] G. W. T. M. J. Frisch, 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, 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, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, **2009**.
- [48] S. S. Xantheas, J. Thom H. Dunning, "Ab initio studies of cyclic water clusters (H₂O)_n, n=1--6. I. Optimal structures and vibrational spectra", *The Journal of Chemical Physics* **1993**, *99*, 8774-8792.
- [49] S. S. Xantheas, "Ab initio studies of cyclic water clusters (H₂O)_n, n=1--6. III. Comparison of density functional with MP2 results", *The Journal of Chemical Physics* **1995**, *102*, 4505-4517.
- [50] S. S. Xantheas, J. Thom H. Dunning, "The structure of the water trimer from ab initio calculations", *The Journal of Chemical Physics* **1993**, *98*, 8037-8040.
- [51] R. Parthasarathi, M. Elango, V. Subramanian, N. Sathyamurthy, "Structure and Stability of Water Chains (H₂O)_n, n = 5–20⁺", *The Journal of Physical Chemistry A* **2009**, *113*, 3744-3749.
- [52] M. Elango, V. Subramanian, N. Sathyamurthy, "Structure and stability of spiro-cyclic water clusters", *Journal of Chemical Sciences* **2009**, *121*, 839-848.
- [53] R. Ludwig, "Water: From Clusters to the Bulk", *ChemInform* **2001**, *32*, 208-208.

- [54] P. Jena, A. W. Castleman, "Clusters: A bridge across the disciplines of physics and chemistry", *Proceedings of the National Academy of Sciences* **2006**, *103*, 10560-10569.
- [55] E. M. Cabaleiro-Lago, J. M. Hermida-Ramon, J. Rodriguez-Otero, "Computational study of the dissociation of H-X acids (X = F, Cl, Br, I) in water clusters", *The Journal of Chemical Physics* **2002**, *117*, 3160-3168.
- [56] M. Śmiechowski, E. Gojło, J. Stangret, "Hydration of Simple Carboxylic Acids from Infrared Spectra of HDO and Theoretical Calculations", *The Journal of Physical Chemistry B* **2011**, *115*, 4834-4842.
- [57] K. Chuchev, J. J. BelBruno, "Theoretical study of carboxylic acids with explicit water molecules", *Journal of Molecular Structure: THEOCHEM* **2006**, *763*, 199-204.
- [58] E. G. Goken, K. L. Joshi, M. F. Russo, A. C. T. van Duin, A. W. Castleman, "Effect of Formic Acid Addition on Water Cluster Stability and Structure", *The Journal of Physical Chemistry A* **2011**, *115*, 4657-4664.
- [59] J. Ho, M. Coote, "A universal approach for continuum solvent pKa calculations: are we there yet?", *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **2010**, *125*, 3-21.
- [60] P. R. Rablen, J. W. Lockman, W. L. Jorgensen, "Ab Initio Study of Hydrogen-Bonded Complexes of Small Organic Molecules with Water", *The Journal of Physical Chemistry A* **1998**, *102*, 3782-3797.
- [61] J. M. Park, A. Laio, M. Iannuzzi, M. Parrinello, "Dissociation Mechanism of Acetic Acid in Water", *Journal of the American Chemical Society* **2006**, *128*, 11318-11319.
- [62] W. Gu, T. Frigato, T. P. Straatsma, V. Helms, "Dynamic Protonation Equilibrium of Solvated Acetic Acid", *Angewandte Chemie International Edition* **2007**, *46*, 2939-2943.
- [63] M. V. Fedotova, S. E. Kruchinin, "Hydration of acetic acid and acetate ion in water studied by 1D-RISM theory", *Journal of Molecular Liquids* **2011**, *164*, 201-206.
- [64] V. S. Znamenskiy, M. E. Green, "Topological Changes of Hydrogen Bonding of Water with Acetic Acid: AIM and NBO Studies", *The Journal of Physical Chemistry A* **2004**, *108*, 6543-6553.
- [65] F. Ito, "Infrared matrix-isolation spectroscopy of trifluoroacetic acid hydrates", *Chemical Physics* **2011**, *382*, 52-57.
- [66] B. Ouyang, T. G. Starkey, B. J. Howard, "High-Resolution Microwave Studies of Ring-Structured Complexes between Trifluoroacetic Acid and Water", *The Journal of Physical Chemistry A* **2007**, *111*, 6165-6175.

- [67] V. Thomas, P. Maurer, R. Iftimie, "On the Formation of Proton-Shared and Contact Ion Pair Forms during the Dissociation of Moderately Strong Acids: An Ab Initio Molecular Dynamics Investigation", *The Journal of Physical Chemistry B* **2010**, *114*, 8147-8155.
- [68] J. Ochterski, "Thermochemistry in Gaussian", http://www.gaussian.com/g_whitepap/thermo.htm, last access December 2012
- [69] K. Wendler, J. Thar, S. Zahn, B. Kirchner, "Estimating the Hydrogen Bond Energy", *The Journal of Physical Chemistry A* **2010**, *114*, 9529-9536.
- [70] A. Smith, M. A. Vincent, I. H. Hillier, "Mechanism of Acid Dissociation in Water Clusters: Electronic Structure Studies of (H₂O)_nHX (n = 4, 7; X = OH, F, HS, HSO₃, OOSO₂H, OOH·SO₂)", *The Journal of Physical Chemistry A* **1999**, *103*, 1132-1139.
- [71] J.-L. Kuo, M. L. Klein, "Structure of protonated water clusters: Low-energy structures and finite temperature behavior", *The Journal of Chemical Physics* **2005**, *122*, 024516.
- [72] M. Prakash, V. Subramanian, S. R. Gadre, "Stepwise Hydration of Protonated Carbonic Acid: A Theoretical Study", *The Journal of Physical Chemistry A* **2009**, *113*, 12260-12275.
- [73] E. Gojłó, M. Śmiechowski, A. Panuszko, J. Stangret, "Hydration of Carboxylate Anions: Infrared Spectroscopy of Aqueous Solutions", *The Journal of Physical Chemistry B* **2009**, *113*, 8128-8136.
- [74] M. Vovk, M. Pavlova, V. Chizhik, A. Vorontsova, "Hydrate shell models of acetate ions according to nuclear magnetic relaxation data and quantum-chemical calculations", *Russian Journal of Physical Chemistry A, Focus on Chemistry* **2011**, *85*, 1597-1602.
- [75] IUPAC, <http://goldbook.iupac.org/MT07068.html>, last access December 2012
- [76] U. W. Schmitt, G. A. Voth, "The computer simulation of proton transport in water", *Journal of Chemical Physics* **1999**, *111*, 9361-9381.
- [77] C. Knight, G. A. Voth, "The Curious Case of the Hydrated Proton", *Accounts of Chemical Research* **2011**, *45*, 101-109.
- [78] E. S. Stoyanov, I. V. Stoyanova, C. A. Reed, "The unique nature of H⁺ in water", *Chemical Science* **2011**, *2*, 462-472.
- [79] T. Takamuku, Y. Kyoshoin, H. Noguchi, S. Kusano, T. Yamaguchi, "Liquid Structure of Acetic Acid–Water and Trifluoroacetic Acid–Water Mixtures Studied by Large-Angle X-ray Scattering and NMR", *The Journal of Physical Chemistry B* **2007**, *111*, 9270-9280.

- [80] Y. Zevatskii, D. Samoilov, "Modern methods for estimation of ionization constants of organic compounds in solution", *Russian Journal of Organic Chemistry* **2011**, *47*, 1445-1467.
- [81] A. V. Marenich, C. J. Cramer, D. G. Truhlar, "Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions", *The Journal of Physical Chemistry B* **2009**, *113*, 6378-6396.
- [82] J. Tomasi, B. Mennucci, R. Cammi, "Quantum Mechanical Continuum Solvation Models", *Chemical Reviews* **2005**, *105*, 2999-3094.
- [83] D. M. Camaioni, C. A. Schwerdtfeger, "Comment on "Accurate Experimental Values for the Free Energies of Hydration of H⁺, OH⁻, and H₃O⁺"", *The Journal of Physical Chemistry A* **2005**, *109*, 10795-10797.
- [84] M. D. Tissandier, K. A. Cowen, W. Y. Feng, E. Gundlach, M. H. Cohen, A. D. Earhart, J. V. Coe, T. R. Tuttle, "The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster-Ion Solvation Data", *The Journal of Physical Chemistry A* **1998**, *102*, 7787-7794.
- [85] K. S. Alongi, G. C. Shields, in *Annual Reports in Computational Chemistry, Vol. Volume 6* (Ed.: A. W. Ralph), Elsevier, Pittsburgh, **2010**, pp. 113-138.
- [86] J. R. Pliego Jr, "Thermodynamic cycles and the calculation of pKa", *Chemical Physics Letters* **2003**, *367*, 145-149.
- [87] J. R. Pliego, J. M. Riveros, "Theoretical Calculation of pKa Using the Cluster-Continuum Model", *The Journal of Physical Chemistry A* **2002**, *106*, 7434-7439.
- [88] S. Zhang, "A reliable and efficient first principles-based method for predicting pKa values. III. Adding explicit water molecules: Can the theoretical slope be reproduced and pKa values predicted more accurately?", *Journal of Computational Chemistry* **2012**, *33*, 517-526.
- [89] D. Du, M. Qin, Z. Zhou, A. Fu, "Density functional theoretical study of pKa of RCOOH (R=H, CH₃, and C₂H₅) using the combination of the extended clusters-continuum model", *International Journal of Quantum Chemistry* **2012**, *112*, 351-358.
- [90] C. P. Kelly, C. J. Cramer, D. G. Truhlar, "Adding Explicit Solvent Molecules to Continuum Solvent Calculations for the Calculation of Aqueous Acid Dissociation Constants", *The Journal of Physical Chemistry A* **2006**, *110*, 2493-2499.
- [91] F. Eckert, M. Diederhofen, A. Klamt, "Towards a first principles prediction of pKa: COSMO-RS and the cluster-continuum approach", *Molecular Physics* **2009**, *108*, 229-241.

- [92] C. C. R. Sutton, G. V. Franks, G. da Silva, "First Principles pKa Calculations on Carboxylic Acids Using the SMD Solvation Model: Effect of Thermodynamic Cycle, Model Chemistry, and Explicit Solvent Molecules", *The Journal of Physical Chemistry B* **2012**, *116*, 11999-12006.
- [93] M. D. Liptak, G. C. Shields, "Accurate pKa Calculations for Carboxylic Acids Using Complete Basis Set and Gaussian-n Models Combined with CPCM Continuum Solvation Methods", *Journal of the American Chemical Society* **2001**, *123*, 7314-7319.
- [94] F. Kiani, A. A. Rostami, S. Sharifi, A. Bahadori, M. J. Chaichi, "Determination of Acidic Dissociation Constants of Glycine, Valine, Phenylalanine, Glycylvaline, and Glycylphenylalanine in Water Using ab Initio Methods", *Journal of Chemical & Engineering Data* **2010**, *55*, 2732-2740.
- [95] N. Vyas, A. K. Ojha, "Calculation of dissociation constants and chemical hardness of some biologically important molecules: A theoretical study", *International Journal of Quantum Chemistry* **2011**, *111*, 3961-3970.
- [96] M. Mangold, L. Rolland, F. Costanzo, M. Sprik, M. Sulpizi, J. Blumberger, "Absolute pKa Values and Solvation Structure of Amino Acids from Density Functional Based Molecular Dynamics Simulation", *Journal of Chemical Theory and Computation* **2011**, *7*, 1951-1961.