The challenge of non-covalent interactions: Theory meets experiment for reconciling accuracy and interpretation

The challenge of non-covalent interactions: theory meets experiment for reconciling accuracy and interpretation

Cristina Puzzarini[1], **Lorenzo Spada**[1,2], **Silvia Alessandrini**[1,2], **Vincenzo Barone**[2]

[1] Dipartimento di Chimica "Giacomo Ciamician", Via F. Selmi 2, I-40126 Bologna, Italy

[2] Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

E-mail: cristina.puzzarini@unibo.it

November 2019

Abstract. In the past decade, many gas-phase spectroscopic investigations have focused on the understanding of the nature of weak interactions in model systems. Despite the fact that non-covalent interactions play a key role in several biological and technological processes, their characterization and interpretation are still far from being satisfactory. In this connection, integrated experimental and computational investigations can play an invaluable role. Indeed, a number of different issues relevant to unravelling the properties of bulk or solvated systems can be addressed from experimental investigations on molecular complexes. Focusing on the interaction of biological model systems with solvent molecules (e.g., water), since the hydration of the biomolecules controls their structure and mechanism of action, the study of the molecular properties of hydrated systems containing a limited number of water molecules (microsolvation) is the basis for understanding the solvation process and how structure and reactivity vary from gas phase to solution. Although hydrogen bonding is probably the most widespread interaction in nature, other emerging classes, such as halogen, chalcogen and pnicogen interactions, have attracted much attention because of the role they play in different fields. Their understanding requires, first of all, the characterization of the directionality, strength, and nature of such interactions as well as a comprehensive analysis of their competition with other non-covalent bonds.

In this review, it is shown how state-of-the-art quantum-chemical computations combined with rotational spectroscopy allow for fully characterizing intermolecular interactions taking place in molecular complexes from both structural and energetic points of view. The transition from bi-molecular complex to microsolvation and then to condensed phase is shortly addressed.

Keywords: Molecular complexes; intermolecular interactions; microsolvation; quantum chemistry; rotational spectroscopy.

Submitted to: *J. Phys.: Condens. Matter*

Contents

1 Introduction

In the past decade, many gas-phase spectroscopic investigations have focused on the understanding of the nature of weak interactions in model systems, and a great effort has been devoted –for example– to studies dealing with complexes formed by small biomolecules with either water or other solvents (see, e.g., refs.[1, 2, 3, 4]). Although non-covalent interactions play a key role in several biological and technological processes (see, e.g.,[5, 6, 7, 8, 9, 10, 11]), their characterization and interpretation are still far from being satisfactory, especially when non-hydrogen bond interactions are considered. In this connection, integrated experimental and computational investigations can play an invaluable role provided that the accuracy of the results is accompanied by their rigorous and understandable interpretation (see, e.g., [12, 13, 14, 15, 16]). Along with the well-established hydrogen bonds, emerging classes of noncovalent interactions, such as those involving a pnicogen or chalcogen atom, are attracting an increasing attention [12, 13, 17, 18, 19, 20].

Concerning biological model systems in the gas phase, the focus is on interactions between the molecules themselves, such as the case of DNA bases pairing (e.g., ring stacking) [21, 22, 23], or on complexation solvent molecules (e.g., hydrogen bond) [4, 24, 25]. With respect to the latter case, hydrogen bond cooperativity has been identified as one of the most important factors that contribute to the stabilization

of hydrated biomolecules [26, 27, 28]. Because the hydration of the nucleic acids controls their structure and mechanism of action, the study of the interactions of water with the individual nuclear bases represents the first step toward the understanding of the effects of hydration on DNA. The study of the molecular properties of hydrated systems containing a limited number of water molecules (microsolvation) is the basis for unveiling the details of the solvation process in terms of the structure and reactivity modifications accompanying the passage from isolated molecules to solution [3, 4, 29, 30]. In parallel, non-covalent pnicogen-bond and chalcogen-bond interactions have attracted much attention because of the fundamental role they play in different fields such as catalysis, drug design, self-assembly processes, and crystal packing [10, 11, 31, 32, 33, 34]. Understanding the mechanisms at the basis of these technological processes requires the characterization of the directionality, strength, and nature of such interactions as well as a comprehensive analysis of their competition with other non-covalent bonds, also taking into account the tuning of these properties by different environments [12, 13, 25].

Among the various spectroscopic techniques, rotational spectroscopy is one of the powerful tools to unveil the structure of molecules and/or clusters because the leading terms for rationalizing rotational spectra, i.e. the rotational constants, are inversely proportional to the corresponding moments of inertia [35], which in turn only depend on the molecular structure and on the isotopic composition. As a consequence, different mass distributions of a given molecule or molecular complex, namely different isomers have different rotational constants and, thus, different rotational spectra. Furthermore, the dependence on the isotopic composition allows the unbiased discrimination among different isotopic species (isotopologues).

The last decade has witnessed an increasing interaction between experiment and theory in the field of molecular spectroscopy (see, e.g., [5, 36, 37, 38, 39, 40, 41]). The key factor to accomplish a complete spectroscopic and structural characterization of the system under consideration is indeed to rely on an integrated experiment-theory approach. In particular, by combining the capability of rotational spectroscopy with high-level quantum-chemical calculations, it is possible to unveil the nature of intra-/intermolecular non-covalent interactions. On the one hand, the analysis of experiments is seldom straightforward because of the subtle interplay of several different effects, which are not easy to evaluate and isolate, and/or the complexity of the system itself. On the other hand, recent advances have enabled computational spectroscopy to provide results that are more and more often comparable to those delivered by accurate experimental techniques. For this reason, computational methodologies are increasingly employed to guide, support and complement experimental studies.

In this review, after a brief introduction of rotational spectroscopy and of the computational methodology, selected results are presented and discussed. We start by presenting a benchmark study of our computational methodology for the accurate determination of intermolecular interaction energies. Then, we move to the structural determination by exploiting the interplay of experiment and theory. After the discussion of emerging classes of non-covalent interactions, namely the pnicogen and chalcogen

bonds, we characterize the transition from hydrogen bond to microsolvation, to conclude with the last step toward the condensed phase.

2 Spectroscopic characterization: Interplay of experiment and theory

The combination of rotational spectroscopy with supersonic expansion paves the way to the investigation of molecular and cluster isomers in an environment free from the solvent, matrix or crystal packing effects encountered in solid- and liquid-phase studies, making it possible the complete characterization of the formation of stable weakly interacting complexes by means of a joint experimental-computational approach.

Before addressing specific examples of the interplay of experiment and theory sketched above, a brief introduction of rotational spectroscopy and computational methodology is needed.

2.1 Rotational spectroscopy

The most important terms of the effective rotational Hamiltonian, within the semi-rigid rotor approximation, can be summarized as follows [42]:

$$H_{rot} = H_R + H_{qcd} + H_{scd} + ... \quad , \tag{1}$$

where:

(i) H_R is the rigid-rotor Hamiltonian, which contains the equilibrium rotational constants B_τ^{eq} (τ denoting a generic inertial axis):

$$B_\tau^{eq} = \frac{\hbar^2}{2hcI_{\tau\tau}^{eq}} \quad , \tag{2}$$

where $I_{\tau\tau}^{eq}$ is the τth diagonal element of the equilibrium inertia tensor, $\mathbf{I}^{eq}$;

(ii) H_{qcd} and H_{scd} are the quartic and sextic centrifugal terms, respectively [35];

(iii) the dots refer to the possibility of including higher-order centrifugal terms or different contributions.

The Hamiltonian of Eq. (1) does not account for the effect of molecular vibrations, which –however– cannot be neglected. The dependence of the rotational constants on the vibrational motion can be conveniently described by means of second-order vibrational perturbation theory (VPT2) [42, 43]. Focusing on the vibrational ground state, which is –in the majority of the cases– that of interest, the rotational constant is given by the equilibrium contribution augmented by a corrective term:

$$B_\tau^0 = B_\tau^{eq} + \Delta B_\tau^0 = B_\tau^{eq} - \frac{1}{2}\sum_{i=1}^{N} \alpha_{i,\tau} \quad , \tag{3}$$

where B_τ^0 denotes a generic rotational constant of the vibrational ground state and $\alpha_{i,\tau}$'s are the so-called vibration-rotation interaction constants (the sum running over the vibrational normal modes).

B_τ^{eq} provides the most important contribution to the corresponding rotational constant, its value indeed accounting for about 90% to 97% of the entire term. From a computational point of view, equilibrium rotational constants are straightforwardly obtained from geometry optimizations, $\mathbf{I}^{eq}$ only depending, as already mentioned in the Introduction, on the molecular structure and isotopic masses. Anharmonic force field calculations are instead required to obtain the vibrational corrections to equilibrium rotational constants. Moving to the centrifugal-distortion Hamiltonian, the quartic terms only depend on the harmonic part of the potential, while the computation of the sextic centrifugal-distortion constants involves harmonic, anharmonic, and Coriolis perturbation terms, thus requiring anharmonic force field computations.

An important issue to be addressed in rotational spectroscopy is the effect due to the presence of a quadrupolar nucleus in the system under consideration, i.e. an atom with nuclear spin $I \geq 1$. This leads to an interaction (defined as hyperfine) that, in turn, is responsible for characteristic features in the rotational spectrum, the so-called hyperfine structure, which arises from the splitting of the rotational energy levels, and consequently, of the rotational transitions. The interaction mentioned above involves the nuclear quadrupole moment and the electric-field gradient at the corresponding nucleus. In such a case, a specific Hamiltonian term should be added to Eq. (1) [35], whose leading terms are the nuclear quadrupole coupling constants $\chi_{\tau\eta}$. Their importance lies on the fact that, while rotational constants provide information on the mass distribution, quadrupole coupling constants yield information on the electronic environment of the quadrupolar nuclei and can be decisive to identify conformers/isomers with similar mass distributions (i.e., similar rotational constants) but different intramolecular interactions.

2.1.1 Experimental techniques. Spectroscopic investigations of systems showing intermolecular bonds can be performed in the solid, liquid or gas phase. However, high-resolution rotational spectroscopy, which operates in the gas phase, allows for gaining insights into the structure without any solvent and/or environmental effects. Furthermore, the structural changes occurring upon complex formation with, e.g., one or more water molecules in the gas phase permit to bridge the gap between the results of gas-phase spectroscopy and the studies of liquid samples. Since its dawn, rotational spectroscopy has allowed for accurately determining molecular structures and internal dynamics. Its combination with supersonic jet expansion has opened up the possibility of forming weakly bound clusters, thus paving the way to the accurate investigation of the genuine nature of the non-covalent interactions governing molecular complexes [25, 30, 44]. Furthermore, rotational spectroscopy in supersonic expansion can even exploit the specificity of non-covalent interactions between two different chiral compounds to form diastereomers, thus making it possible to determine the absolute configuration of an enantiomer by means of chiral tagging [45].

In a typical experiment, a sample/inert gas mixture (usually with less than 1% of sample) at thermal equilibrium (room temperature) and a stagnation pressure on the order of a few bars is adiabatically expanded in a high vacuum (about $10^{-7}/10^{-9}$

bars) chamber through the nozzle (0.5-2 mm range circular orifice) of a solenoid valve to form a pulsed supersonic jet [46]. The collisional energy transfer taking place in the initial stages of the jet-formation results in very low molecular rotational and vibrational temperatures, so that weakly bound clusters can be formed and isolated in the collision-free expansion. The structures of these *in situ* formed clusters can therefore be characterized by collecting their microwave spectra, the commonly used frequency range being 2-26 GHz [46]. The experimental sequence consists of the gas-injection, microwave excitation, and detection of the molecular response. Once the gas sample is introduced as a pulsed jet (typically ≤ 0.5 ms) by means of a solenoid valve, a microwave excitation pulse (typically ≤ 2.0 μs) causes an oscillating macroscopic dipole moment, i.e. polarization of the molecular ensemble in a time shorter than the relaxation time to rotational decoherence. Once the microwave radiation is removed, the resulting transient molecular spontaneous emission is collected in time domain (up to several hundreds of μs) and Fourier transformed into the frequency domain to get the microwave spectrum [46].

The capability to phase-invariantly repeat and to accumulate the molecular coherence signals (with a stability that even allows for collecting millions of individual molecular excitation pulse responses collected over hours or days) makes it possible to experimentally observe very weak transitions, e.g. those belonging to large molecular complexes. Specific conformer, isomer, or isotopologue identification, in contrast to many other spectrometric techniques, can be performed by means of the assignment of tenth, hundreds or even thousands of discrete rotational transitions, thus leading to unrivaled specificity and accuracy.

To observe and record the rotational spectra of molecular complexes, two principal spectrometer designs are employed [44, 46, 47, 48, 49, 50], which allow for broad- or narrow-band spectral acquisition of a single molecular pulse event. In both cases, the observation of the molecular coherence at high resolution and high sensitivity makes it possible to spectroscopically characterize not only large molecules but also transient complexes and, in particular, even large non-covalently bonded molecular clusters. The two designs are based on different microwave excitation signals: a single-tone impulse and a chirp-pulse characterize the fixed frequency excitation Fourier Transform Microwave spectrometer (impulse-FTMW) and the swept frequency fast-passage Fourier Transform Microwave spectrometer (chirp-FTMW), respectively. These two techniques are complementary. Indeed, the chirp-FTMW spectrometer allows for covering a broadband spectral region (up to 10 GHz) for sufficiently polar species in a single experimental event, while the impulse-FTMW spectrometer covers only narrowband section (less than 1 MHz). However, the latter is able to achieve very high sensitivity and is also suitable for less polar compounds when a coaxially oriented beam-resonator arrangement (COBRA) is exploited, since bandwidth-limiting high-Q Fabry-Perot type resonators can be employed.

An essential requirement for the experimental setup is the gas-phase sample generation to form the molecular pulse. Therefore, depending on the vapor pressure

of the sample in the conditions of the experiment, different sources can be employed to handle compounds in different state of matter and thus to get the aforementioned less than 1% ratio in the sample/carrier gas mixture. In this respect, it is useful to mention that, in addition to the classical heating and cooling systems, according to the stability/reactivity of the samples or to the particular species to be formed in the gas-phase, different sources are available, thus opening the way to the characterization of a great variety of non-covalent bound clusters. Several molecular complexes have been unveiled by means of the fast-mixing nozzles [51, 52], in which a pair of concentric tube keeps separated the two samples until the expansion, thus avoiding them from reacting, in favor of the weakly bound complex formation. On the other hand, the employment of laser ablation techniques to vaporize, for example, biologically relevant molecules such as aminoacids, allows for overcoming decomposition issues occurring with sample heating, and has been successfully applied to reveal the hydrogen bonding features of, e.g., the glycine-water [53] and alanine-water [54] clusters. In detail, for the investigation of the alanine-water and glycine-water complexes mentioned above, a cylindrical rod (obtained from the pulverised solid by appropriate pressing) is translated and rotated by a motor support, and then it is vaporized by a Nd:YAG laser pulse (355 nm and 532 nm wavelengths, respectively) and then supersonically expanded together with water vapor coming from a water reservoir place before the nozzle. Finally, electrical discharge sources can also be used to generate *in situ* clusters in which a monomer is a radical species, as in the case of the hydrogen bonded HO_2-H_2O [55] and H_2O-HO [56] molecular complexes.

2.2 The computational approach

To exploit an integrated experimental computational strategy, the first step is the calculation of the spectroscopic parameters involved in the prediction and analysis of rotational spectra [37, 39, 41, 57, 58]. Concerning rotational constants, according to Eq. (3), two contributions are required. As already mentioned, the equilibrium rotational constants only depend on the equilibrium structure, thus requiring –from a computational point of view– geometry optimizations. Since B_τ^{eq} provides by far the largest contribution to B_τ^0, high accuracy is required for its determination, with this being achieved by means of the composite schemes introduced later in the manuscript. Vibrational corrections account only for a few percents and their evaluation involves an anharmonic force field [37, 39, 41, 57], which can be effectively computed at a less refined level of theory. As mentioned above, harmonic and anharmonic force field calculations are needed for obtaining quartic and sextic centrifugal-distortion constants, respectively. Analogously to rotational constants, the nuclear quadrupole coupling constants at the equilibrium can be accurately evaluated by means of composite schemes, with vibrational corrections possibly determined at a less refined level of theory (see, e.g., refs. [39, 59]).

The ultimate model for quantum-chemical (QC) computations is the full configuration interaction (FCI) method [60] in conjunction with extrapolation to the

complete basis set (CBS) limit and the proper account of relativistic and non-adiabatic effects. However, except for the case of few-electron complexes, this level of theory is computationally prohibitive, hence more approximate yet accurate methods have been developed. For those systems that do not show strong static correlation effects, the most successful strategy –also for non-covalent molecular complexes– is based on the coupled cluster (CC) theory [61]. In particular, the CC model that accounts for the full treatment of single and double excitations and the perturbative inclusion of triple excitations, CCSD(T) [62], has become the so-called "gold standard" of the contemporary computational chemistry. In fact, due to a fortuitous but systematic error compensation, CCSD(T) performs actually even better than the model including the full treatment of triple excitations (CCSDT) [63, 64]; therefore, in order to obtain an improvement over CCSD(T), it is required to move to the CCSDTQ [65] method or to the CCSDT(Q) model [66, 67], the latter including a perturbative treatment of quadruples. CCSDT(Q) can be considered the "gold standard" for open-shell systems, especially when involving non-negligible delocalization effects, in view of the worse convergence behavior of the perturbative expansion in terms of excitation classes with respect to closed-shell species. Another attractive feature of the CC approach is that the extent of multireference character can be inferred by inspecting the cluster amplitudes.

In the present contribution, we will focus our interest on closed-shell single-reference molecular complexes; therefore, CCSD(T) will represent our reference for accurate calculations. However, high-accuracy results can be obtained only if CCSD(T) is coupled with the extrapolation to the CBS limit and the inclusion of core-valence (CV) correlation effects, thus leading to the model denoted as CCSD(T)/CBS+CV [68]. The same considerations apply to the computation of energy derivatives, which are required to determine equilibrium geometries (gradients), non-potential energy contributions to thermodynamic functions (at least second derivatives), and vibrational averaging effects (at least semi-diagonal third derivatives).

The second central idea behind the success of accurate QC computations is the additivity approximation, which opens the way to composite schemes (see, e.g., refs [24, 59, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77]). Within these approaches, hierarchic basis sets, whose accuracy increases consistently with the cardinal number, allow the use of robust extrapolation procedures to reach the CBS limit. In general terms, using a triple-zeta basis set (e.g. cc-pVTZ [78] or ANO1 [79, 80] or pC1 [81]) and its quadruple-zeta analog reliable extrapolations to the CBS limit can be performed using the different extrapolative formula available in the literature (see, e.g., refs. [82, 83, 84]). Some of them rely on the separate extrapolation of the Hartree-Fock (HF) and post-HF contributions, also requiring the employment of a third and thus larger basis set (e.g. quintuple-zeta). To improve the description of second-row atoms, tight polarization functions can also be added, thus leading to the cc-pV(n+d)Z basis sets [85]. However, these conventional basis sets are not adequate for the description of non-covalent interactions, which are sensitive to the tails of the wavefunctions of the partners and thus require the use of diffuse functions (denoted as augmented (aug)

basis sets [86, 87, 88]). In the past few years, a number of systematic investigations has shown that CCSD(T) computations with augmented triple and quadruple-zeta basis sets followed by extrapolation to the CBS limit provide sufficiently accurate results in most cases. The effects of CV correlation need also to be added by means of additive approaches based on purposely tailored basis sets (e.g. cc-pCVnZ [89] or cc-pwCVnZ [90]), which contain tight functions for describing the region intermediate between core and valence electrons. The same basis sets are usually sufficient to take into account also the core-core contributions, which are –however– generally less important. The situation is analogous for energy derivatives (this is not always the case for electric or magnetic properties, but we will not consider these aspects in the present contribution), but the required computer resources increase by roughly one order of magnitude for each degree of derivation.

Composite schemes are largely employed to obtain accurate molecular structures, spectroscopic parameters and energetics. Focusing on the understanding of intermolecular interactions, the interest is mainly on the structural and energetic characterization. As far as equilibrium structure determinations are concerned, the so-called "gradient" scheme [68, 69, 91] is based on the minimization of an energy gradient that is set up according the target accuracy and the dimension of the system under consideration. The scheme including up to the quadruple excitations, shortly denoted as CCSD(T)/CBS+CV+fT+fQ, is able to provide geometriacal parameters with an accuracy better than 0.001 Å for bond lengths and 0.05 degrees for angles (see, for example, refs. [39, 68, 69, 92, 93, 94, 95] and references therein). However, inclusion of fT and fQ corrections (i.e., corrections due to the full treatment of triple and full quadruple excitations, respectively) makes this approach extremely expensive from a computational point of view. On the contrary, the CCSD(T)/CBS+CV scheme, which is characterized by a very limited loss of accuracy (e.g. refs. [37, 68, 95, 96]), is affordable for small molecular adducts such as the formamide dimer [97]. If it is assumed that the molecular properties and, in particular, the structural parameters show the same behavior as the energy, then the additivity approximation can be directly applied to the properties themselves (see, e.g., refs. [59, 98, 99, 100]), thus defining the so-denoted "geometry" composite schemes. Among them, of interest in this context is the so-called "cheap geometry" scheme, which –starting from CCSD(T) calculations in conjunction with a triple-zeta quality basis set and within the frozen-core approximation (fc)– includes additional contributions such as the extrapolation to the CBS limit using Møller-Plesset theory to second order (MP2) [101]. This scheme, whose denomination is due to the strong reduction of the computational cost with respect to an approach entirely based on CCSD(T) computations, is well tested for semi-rigid and rather flexible systems [59, 74, 96, 102] and has been recently applied to molecular complexes [13, 103]. When the dimension of the system under consideration is such that a composite scheme is computationally too much expensive, the double-hybrid B2PLYP functional [104] in conjunction with a partially or fully augmented triple-zeta offers a good alternative provided that, especially when dealing with molecular

complexes, dispersion corrections are taken in proper account (e.g. by resorting to Grimme's D3 formulation [105, 106, 107]). While the accuracy obtainable by B2PLYP-D3BJ calculations (with BJ standing for Becke-Johnson damping function [107]) for intermolecular parameters is currently under investigation, for intramolecular bond distances the discrepancies from very high-level determinations range from 0.001 Å to 0.003 Å (see e.g. refs. [94, 95, 108, 109]). In addition, for molecular complexes, to the so-called basis set superposition error (BSSE), which plagues QC calculations whenever basis sets of limited size are employed, has to be recovered by including the so-called counterpoise (CP) correction [110].

Moving to energetics, the HEAT protocol [72, 73, 111], besides being entirely independent of experimental data and not containing any empirical parameter, is able to reach the so-called sub-kJ accuracy. This scheme is analogous to the CCSD(T)/CBS+CV+fT+fQ approach mentioned above, but it also accounts for first-order spin-orbit coupling, diagonal Born-Oppenheimer correction, and scalar relativistic effects. Other high-precision protocols have been introduced such as the Weizmann-3/4 (W3, W4)[76, 77], focal-point analysis (FPA)[71, 112], and Feller-Dixon-Peterson (FDP)[113] approaches. However, all these model chemistries are computationally expensive. To reduce the computational cost without loosing accuracy, one can resort to the CCSD(T)/CBS+CV scheme mentioned above, which is rather well tested in the literature (see, e.g., refs. [93, 114, 115, 116]) and provides results with an accuracy well within 1 kJ mol^{-1}. This approach is anyway still too much computationally demanding for addressing molecular complexes formed by medium-sized molecules.

To extend the applicability of the composite schemes presented above to larger systems, an effective solution is to apply the "cheap" composite approach (denoted as ChS in the following) to the evaluation of electronic energies:

$$E_{cheap} = E(CC/VTZ) + \Delta E(CBS) + \Delta E(CV) + \Delta E(aug) \ , \qquad (4)$$

where, except for the first term on the right-hand side (evaluated at the fc-CCSD(T) level in conjunction with the cc-pVTZ basis set), all contributions are computed using the MP2 method. The second and the third term have already been mentioned, while the last contribution accounts for the effect of diffuse functions in the basis set, which is particularly important in the accurate description of molecular complexes. The extrapolation to the CBS limit is performed using the cc-pVTZ and cc-pVQZ basis sets by resorting to the n^{-3} formula [83]. Triple-zeta quality sets are used for the CV and 'aug' contributions: cc-pCVTZ and aug-cc-pVTZ, respectively. However, a benchmark study in progress in our laboratories has pointed out that the inclusion of the effect of diffuse functions in the basis set by means of a separate contribution leads to a worsening of the accuracy with respect to the scheme that it does not account for it at all. A relevant improvement is obtained by using, for both CCSD(T) calculations and the extrapolation to the CBS limit, diffuse-augmented basis sets. To avoid any huge increase of the computational cost, the "seasonal" jun-cc-pVnZ sets [88] can be

employed, thus leading to the scheme denoted as jun-ChS:

$$E_{cheap} = E(CC/jun-VTZ) + \Delta E(jun-CBS) + \Delta E(CV) \ . \qquad (5)$$

The performance of different variants of the ChS will be discussed later in the text.

The rate determining step of the ChS model chemistries is the CCSD(T) energy evaluation in conjunction with a triple-zeta quality basis set, which can become unaffordable for large systems. On the other hand, the use of a double-zeta basis set (such as aug-cc-pVDZ, jul-cc-pVDZ or jun-cc-pVDZ) reduces too much the accuracy for conventional approaches. Although the extrapolation to the CBS limit can be avoided when using explicitly correlated (F12) coupled-cluster ansatzes [83], reduction of the basis set below the aug-cc-pVTZ quality still reduces the accuracy of the model chemistry down to the so-called "silver standard" level [117]. Since the MP2 extrapolation to the CBS limit is quite effective (and can be further fastened by resolution of identity and analogous techniques [118]), the bottleneck remains the CCSD(T) calculation, whose (T) part does not benefit from current F12 implementations. Local correlation models can become the methods of choice for large systems [119], but the robustness of their implementation and their general applicability for non-covalent complexes have not yet been fully tested.

As mentioned above for structural determinations, another important aspect when addressing molecular complexes is the so-called basis set superposition error (BSSE). To recover it, we resort to the full inclusion of the counterpoise (CP) correction. By definition, the CP corrected interaction energy, E_{CP}^{int}, is computed as follows:

$$E_{CP}^{int} = E_{AB} - (E_A^{AB} + E_B^{AB}) \ , \qquad (6)$$

where E_{AB} is the molecular complex energy, while E_A^{AB} and E_B^{AB} are the energies of the monomers, A and B, calculated using the basis set of the molecular complex (AB) at the geometry they assume in the latter. The corresponding non-CP-corrected (NCP) energy value is obtained by considering the two fragments in their basis set and at the corresponding geometry in the molecular adduct (E_A^A and E_B^B).

2.2.1 Analysis of the interaction energy: NBO and SAPT. While accurate interaction energies can be evaluated by means of the state-of-the-art methodologies described above, important –at least semiquantitative– information about the type and number of non-covalent interactions occurring as well as on various contributions to such interactions can be obtained by the natural bond order (NBO) [120] and symmetry-adapted perturbation theory (SAPT) [121, 122] analyses.

Energy decomposition analyses (EDAs) allow the partitioning of the overall interaction energy into more intuitive chemical contributions, namely electrostatic, polarization, charge-transfer, exchange and dispersion. The EDA schemes can be classified in variational models, in which the interaction energy is decomposed by use of intermediate wavefunctions, and perturbative models, in which the interaction between the fragments is seen as a perturbation to the non-interacting description and the interaction is built in terms of corrections resulting from different physical-chemical

effects. Both types of EDAs employed in our analyses (NBO second-order perturbation and SAPT), which are sketched below, belong to the family of perturbative approaches.

The natural bond orbital (NBO) approach describes a molecular system in terms of a wavefunction that follows the classical Lewis interpretation of bonds and lone-pairs [123]. A first outcome of this model are atomic charges, which are much more transferable and less prone to basis set effects than the conventional Mulliken charges. Furthermore, a second-order perturbative approximation provides an estimate of the charge-transfer contribution between NBO pairs (from a bonding to an antibonding NBO) to the total interaction energy [124]. This energy is expressed by the following equation:

$$\Delta E_{ij}^{(2)} = \frac{-w_i |F_{ij}|^2}{\varepsilon_j^{NL} - \varepsilon_j^{L}} \,, \tag{7}$$

where $-w_i$ is the donor orbital occupancy (approximately 2), F_{ij} is the Fock-matrix element for the donor-acceptor orbital interaction, and ε_j^{NL} and ε_j^{L} are the energies of the donor and acceptor orbital, respectively. In this way, useful insight can be gained into the non-Lewis interaction of an atom within a molecule with the neighboring functional groups, thus allowing the study of particular fragments of chemical interest. When L and NL orbitals belong to different molecules, it is possible to analyze inter-molecular interactions in terms of specific functional groups. Further insights about different contributions can be obtained in this framework by the natural (N) EDA approach [125], which belongs to the family of variational EDAs, but it is fully defined only at the Hartree-Fock or DFT level. As a consequence, the NBO approach provides at most qualitative information and should be replaced by more advanced models when quantitative aspects are of interest.

This task can be effectively accomplished by the symmetry-adapted perturbation theory (SAPT) [121]. Within this model, the electrostatic contribution includes Coulombic multipole-multipole-type interactions as well as the interpenetration of charge clouds. Exchange-repulsion is a repulsive force which arises from the monomer wavefunction overlap and the fermionic anti-symmetry requirements of the dimer wavefunction. Induction includes both polarization from each monomer's response to the other's electric field as well as charge transfer, although these two contributions are non-separable in SAPT. Dispersion is an attractive force resulting from the dynamical correlation between electrons on one monomer with those on another.

Thanks to recent advances in wave-function based SAPT, it has become possible to apply this model to large molecular systems as well as to more complex clusters, such as those mimicking surfaces [126, 127, 128]. In the present connection (i.e. accurate treatment of medium-sized systems), the major advance is the use of natural orbital (NO) truncations to reduce the size of the virtual space in high-order SAPT computations (denoted SAPT2+, SAPT2+(3), SAPT2+3) [129].

The simplest SAPT method, SAPT0, treats the monomers at the Hartree-Fock level and adds explicit dispersion terms issuing from second-order perturbation theory to the electrostatics, exchange, and induction terms obtained from a HF dimer treatment [130].

A slightly modified variant including the scaling of some terms has also been introduced (referred to as sSAPT0), which is claimed to lead to much improved results without any significant increase of computational time [122]. While employing the SAPT0 treatment for dispersion, SAPT2 adds terms for electrostatics, exchange, and induction up to the second order with respect to the intra-monomer electron correlation. This is one of the most commonly applied levels and it performs similarly to MP2, hence it is suitable only for electrostatic-, non-dispersion-, dominated systems [121].

In SAPT2+ and beyond, intra-monomer corrections to dispersion are summed through second-order at a level similar to MP4 [101] and thus it requires a steep O(N^7) scaling. As far as SAPT2+ and SAPT2+(3) are concerned, they only differ for two O(N^6)-scaling terms. SAPT2+(3) is a more complete method and it would seem to be preferred. The heavy cost of the perturbative triples term may be substantially reduced by expressing it in terms of MP2 natural orbitals then truncating the virtual orbital space by half or more with a minimal loss of accuracy. SAPT2+3 extends SAPT2+(3) with third-order induction terms as well as the coupling between induction and dispersion, making it more computationally expensive by a small fraction. The third-order induction terms in SAPT2+3 are included in an updated, third-order approach.

These high-order SAPT methods are found to reliably describe all but the most difficult dispersion-bound systems, electrostatically dominated complexes, whose description is often improved by the addition of a MP2 correction (δE_{MP2}) evaluated as the difference of the interaction energies computed at the MP2 (CP corrected) and SAPT2 levels. For SAPT2+, SAPT2+(3), and SAPT2+3, the analogous MP2-corrected methods are generated by adding δE_{MP2} to each method, thus leading to SAPT2+δMP2, SAPT2+(3)δMP2, and SAPT2+3δMP2, respectively. For the three high-order wavefunction-based SAPT methods, one has a choice to use MP4-level dispersion terms or to compute dispersion using the t_2 amplitudes issuing from CCD calculations, thus giving rise to six SAPT(CCD) variants (the three methods described above with and without δMP2). This choice affects only the dispersion term, thus leaving electrostatics, exchange, and induction unaltered. The CCD dispersion should lead –in principle– to a more robust treatment.

A detailed analysis of the accuracy of different SAPT models has been performed based on four databases considering different binding motifs, with examples ranging from strong hydrogen bonds, to mixed interactions, to dispersion dominated bonding (S22, NBC10, HBC6, and HSG) [121, 120]. This study ended up with the definition of the following standards in terms of the overall mean average error (MAE) with respect to CCSD(T)/CBS reference values and of the relative timings evaluated for the adenine-thymine complex (RT-AT):

(i) Bronze: sSAPT0/jun-cc-pVDZ (MAE = 2 kJ mol^{-1}; RT-AT = 1)

(ii) Silver: SAPT2+3/aug-cc-pVDZ (MAE = 1.2 kJ mol^{-1}; RT-AT = 170)

(iii) Gold: SAPT2+(3)δMP2/aug-cc-pVTZ (MAE = 0.6 kJ mol^{-1}; RT-AT = 2000)

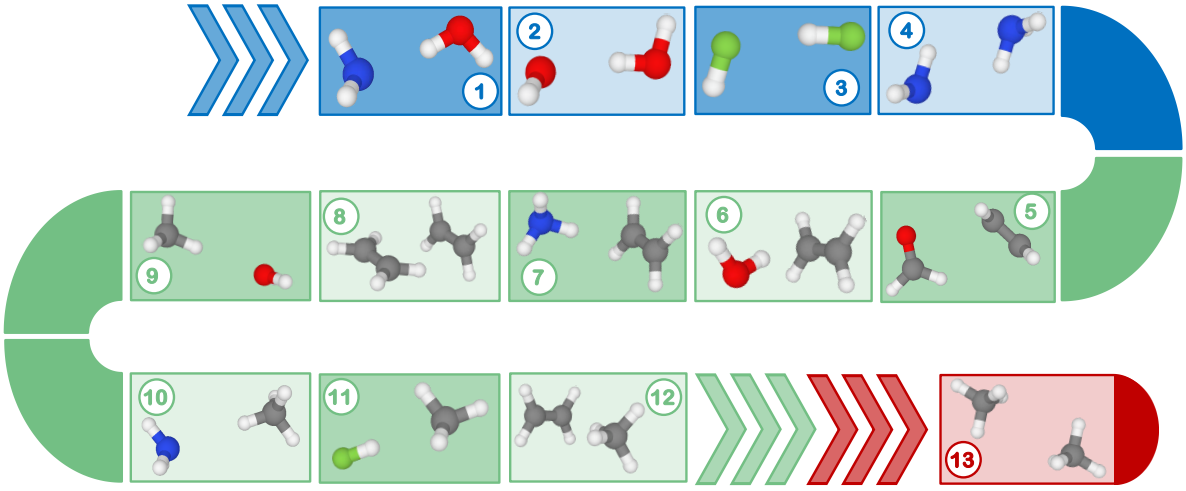


Figure 1. Molecular complexes selected for the benchmark analysis. Hydrogen bonds in blue, mixed electrostatic/dispersion interactions in green, and dispersion interactions in red.

2.3 Energetic characterization: small molecular complexes as test cases

To accurately determine interaction energies of molecular complexes, it is necessary to set up a reliable and accurate model chemistry protocol at a reduced computational cost. As mentioned above, this is envisaged in the ChS approach. To test its limitations and performance, we have carried out a benchmark study based on highly reliable reference values for the interaction energy of a set of 24 small-sized non-covalent molecular complexes containing up to 13 first-row atoms (the A24 set), reported in ref. [131]. The small size of these dimers allows for pushing theory toward its limit and thus for performing a particularly accurate and reliable benchmark. Since our target is to investigate molecular complexes at their equilibrium structure containing elements belonging to the first two-rows of the periodic table and essential for life (i.e. H, C, N, O, F, P, S, Cl), we have selected 13 of the above mentioned systems (for which reliable post-CCSD(T) contributions are also available), which are displayed in Figure 1. Furthermore, four additional systems containing second-row atoms have been considered.

Since non-covalent bondings span a huge range of interactions, from the relatively strong electrostatic, hydrogen bonds (10-30 kJ mol^{-1}) to the much weaker dispersion interactions (1-10 kJ mol^{-1}), discussions based on absolute errors might be meaningless. Indeed, while a target error of 0.2 kJ mol^{-1} is fully satisfactory in the first case, this might lead to an excessive uncertainty in the second one. As a consequence, we prefer to resort to percentage errors with a target of $\sim$1%, without any outlier above 3%. In any case, absolute deviations above 0.2 kJ mol^{-1} are considered as outliers. While the ChS composite scheme [59, 74, 102], which has been described above, is proved to deliver accurate results for the structures and energies of a large panel of semi-rigid [59, 102]

Table 1. Performance of different ChS composite schemes: interaction energies in kJ mol^{-1}.

Complex	"best-ref"[a]	ChS[b]	ChS-$\Delta\alpha$[b]	aug-ChS[b]	jul-ChS[b]	jun-ChS[b]	jun-ChS[c]
1. NH_3-H_2O	-27.39	-26.61	-28.77	-27.51	-27.43	-27.32	-27.57
2. H_2O-H_2O	-21.07	-20.48	-21.50	-21.25	-21.05	-20.98	-21.11
3. HF-HF	-19.18	-18.63	-18.70	-19.45	-19.23	-19.23	-19.47
4. NH_3-NH_3	-13.26	-13.03	-13.95	-13.31	-13.22	-13.27	-13.31
5. C_2H_4-H_2CO	-6.84	-6.72	-8.33	-6.92	-6.86	-6.96	-6.85
6. C_2H_4-H_2O	-10.82	-10.59	-12.44	-10.91	-10.85	-10.87	-10.81
7. C_2H_4-NH_3	-5.84	-5.67	-6.85	-5.89	-5.85	-5.89	-5.86
8. C_2H_4-C_2H_4	-4.64	-4.58	-5.76	-4.73	-4.68	-4.75	-4.69
9. H_2O-CH_4	-2.85	-2.57	-2.85	-2.83	-2.79	-2.81	-2.80
10. NH_3-CH_4	-3.26	-2.94	-3.03	-3.29	-3.24	-3.23	-3.20
11. HF-CH_4	-6.95	-6.88	-8.38	-7.08	-7.01	-7.03	-7.08
12. C_2H_4-CH_4	-2.17	-2.06	-2.51	-2.18	-2.16	-2.17	–[d]
13. CH_4-CH_4	-2.27	-2.11	-2.78	-2.26	-2.19	-2.26	–[d]
MAE		0.28	0.83	0.08	0.03	0.05	0.08
Max Abs. Error		0.78	1.62	0.27	0.08	0.12	0.29
%MAE		3.93	12.21	0.94	0.77	0.80	0.90
Max Rel. Error		9.82	24.14	1.94	3.52	2.37	1.87

[a] Ref. [131].

[b] Computed using the CCSD(T)/CBS reference geometries from ref [133].

[c] Computed using the B2PLYP-D3BJ/may-cc-pVTZ reference geometries.

[d] Excluded from the analysis because of too large differences between CCSD(T)/CBS and B2PLYP-D3BJ/may-cc-pVTZ reference geometries.

and flexible [96, 74] molecules, its extension to molecular complexes [12, 13, 24, 103, 132] cannot be given for granted, and indeed the additive inclusion of the effect of diffuse functions has casted some doubts.

The results for different ChS schemes are summarized in Table 1, where the "best-ref" interaction energies are taken from [131] and they account for the extrapolation to the CBS limit of CCSD(T) energies within the frozen-core approximation, which are augmented for (*i*) core-valence and core-core contributions extrapolated to the CBS limit as well, (*ii*) the contribution due to quadruple excitations from all electron-CCSDT(Q) calculations in conjunction of the aug-cc-pVDZ basis set, and (*iii*) the inclusion of the scalar relativistic effects, calculated using CCSD(T) in conjunction with the aug-cc-pCVQZ-DK basis set [134].

The first approach considered is the ChS scheme obtained from Eq. (4) without inclusion of the last term, which is denoted as the $\Delta\alpha$ contribution. The resulting relative mean absolute error (%MAE) is 3.93%. Not only the %MAE is well outside our target accuracy, but also the maximum relative error ($\sim$10%) is too large. In Table 1, in addition to relative errors, the absolute errors are also given. The latter are rather small, ranging –in unsigned terms– from 0.04 kJ mol^{-1} to 0.78 kJ mol^{-1}. However, these values are not satisfactory because more than half of the molecular complexes

considered are characterized by small interaction energies and, therefore, a discussion in terms of relative errors is more sound. The next step is the inclusion of the effect of diffuse functions by means of a separate contribution, thus assuming that the diffuse functions contribution follows the additivity approximation. For this purpose, the $\Delta\alpha$ term has been computed using different families of basis sets including diffuse functions, the results collected in Table 1 being those obtained with the aug-cc-pVTZ set. In all cases, the relative errors are considerably higher than those without the $\Delta\alpha$ term, confirming that diffuse functions cannot be treated by additive schemes. In particular, when the aug-cc-pVTZ basis set is used, %MAE is as large as 12.21%, with a max relative error of 24.14%.

Therefore, it seems necessary to introduce diffuse functions in both CCSD(T) calculations and in the MP2 extrapolation term. Resorting to the Dunning and coworkers hierarchical basis sets, namely the aug-cc-pVnZ family [86], leads to a relevant improvement, the results being collected in the fourth column (aug-ChS) of Table 1. Indeed, the %MAE, being 0.94%, is about one fourth of that of the starting approach, i.e. the ChS without addition of the effect of diffuse functions, with a maximum relative error of 1.94%. Impressive is also the absolute MAE, this being 0.08 kJ mol^{-1}. However, while the composite scheme based on the aug-cc-pVnZ family of basis sets can represent an accurate approach to describe any type of non-covalent interactions, the full inclusion of diffuse functions strongly affects its applicability to medium- and large-sized systems. Furthermore, it has to be noted that this scheme leads to energies that are overestimated for almost all complexes, thus suggesting that full inclusion of diffuse functions is not the best choice for an effective extrapolation. This overestimation is probably due to the well-known overshooting of intermolecular correlation by the MP2 model [135, 136, 137]. Therefore, the systematic removal of some sub-shells of diffuse functions present in the aug-cc-pVnZ basis sets might provide improved results at a reduced computational cost. For this reason, we have thus investigated systematically the "seasonal" set of bases [88]. In particular, starting from the aug-cc-pVnZ, the removal of the diffuse functions on hydrogen and helium leads to the definition of the jul-cc-pVnZ basis sets, which are commonly used for the description of weakly-bound compounds (see, e.g., ref. [138]), and further removal of the highest-angular momentum diffuse functions for non-hydrogen atoms leads to the jun-cc-pVnZ basis sets. Several tests have shown that a further reduction of the number of diffuse functions leads to unsatisfactory descriptions, whereas the results of jul- and jun- families of basis sets are comparable. Indeed, the jul-ChS (corresponding to the scheme of Eq. (5), where jul-cc-pVnZ basis sets are used in the determination of the two contributions of the right-hand term) and jun-ChS (Eq. (5)) approaches further improve the mean deviations, %MAE being 0.77% and 0.80%, respectively. However, the maximum relative errors worsen with respect to aug-ChS, these being 3.52% and 2.37%, respectively. In Figure 2, the interaction energies obtained with the different types of ChS protocols are graphically represented for 5 selected molecular complexes in terms of the relative differences with respect to the corresponding "best-ref" reference values.

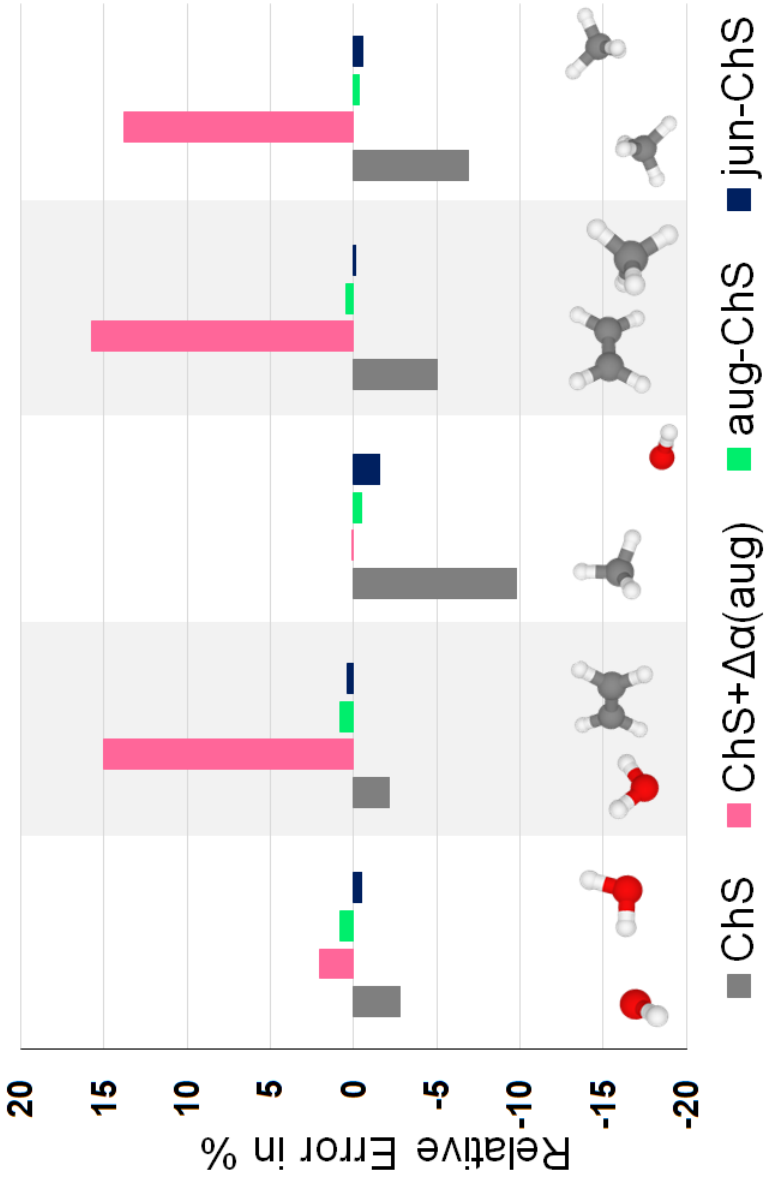


Figure 2. Interaction energy differences (%) with respect to "best-ref" values for 5 selected complexes (n. 2, 6, 9, 12, and 13; see Figure 1). Comparison of different ChS approaches (for their definition, see text): grey bars refer to ChS, pink bars to ChS+$\Delta\alpha$(aug), green bars to aug-ChS, and blue bars to jun-ChS.

Table 2. Second-row atom-containing complexes: jun-ChS energies (kJ mol^{-1}) [a].

Complex	fc-CCSD(T)/ junTZ	ΔE^{∞}_{MP2}/ jun(T,Q)Z	ΔE^{CV}_{MP2}	E^{int}_{CP} This work	E^{int}_{CP} Lite.[b]	Δ^{def}	D_e This work	D_e Lite.
FH_2P-NH_3	-24.8311	-4.4465	-0.2669	-29.5445		-2.123	-27.42	-28.49[c]; -25.52[d]
FH_2P-SH_2	-10.7386	-3.9130	-0.2836	-14.9352		-0.512	-14.42	-15.40[c]; -12.89[d]
FHS-OFH	-11.3921	-2.6460	-0.0930	-14.1311		-0.366	-13.36	-15.15[c]; -12.55[d] -13.56[e]
CH_3OH-HCl	-23.7406	-3.1925	-0.3500	-27.2831	-26.6061	-1.324	-25.96	

[a] CP-B2PLYP-D3BJ/jun-cc-pVTZ references geometries are used. The tight *d*-functions-augmented jun-cc-p-V(n+d)Z sets[85] have been employed for second-row elements [b] MP2-F12/cc-pV(T,Q)Z-F12 half-CP + [CCSD(T)-MP2]/jul-cc-pV(T,Q)Z half-CP results from ref. [141]. [c] NCP results from ref. [140]. [d] CP results from ref. [140]. [e] Value at the CP-DLPNO-CCSD(T)/aug-cc-pV5Z level from ref. [140].

All the above computations were performed by using accurate reference geometries evaluated at the CCSD(T)/CBS level [133], whose cost can become prohibitive already for medium-sized systems. On the grounds of previous experience on semi-rigid molecules, one can resort to reference structures obtained at the B2PLYP-D3BJ/may-cc-pVTZ level. The results for the jun-ChS approach applied on top the latter reference geometries are collected in the last column of Table 1. They are indeed quite satisfactory, since the %MAE only slightly worsen, i.e. from 0.80% to 0.90%, with respect to the case in which ChS is used in combination with CCSD(T)/CBS reference structures. For what concerns the absolute MAE, the worsening is very small as well, i.e. from 0.05 kJ mol^{-1} to 0.08 kJ mol^{-1}. In conclusion, even by employing B2PLYP-D3BJ/may-cc-pVTZ reference geometries, the jun-ChS protocol fulfils well the target accuracy, with a maximum absolute error as small as 0.29 kJ mol^{-1}, 1.87% in relative terms.

The A24 set does not consider any molecular complexes containing second-row elements and, more generally, accurate interaction energies for this class of systems are very limited in the literature [139, 140, 141]. For these reasons, we have decided to investigate four small non-covalent complexes bearing at least one second-row atom for which reasonably accurate computations are available, namely $FH_2P\cdots NH_3$, $FH_2P\cdots H_2S$, FHS$\cdots$OFH and $CH_3OH\cdots$HCl. To compare our results with those reported in the literature, the interaction jun-ChS energies need to be corrected for the deformation contribution, Δ^{def}, in order to obtain the equilibrium dissociation energies (D_e), the Δ^{def} term being defined as:

$$\Delta^{def} = E_A^{A,A} + E_B^{B,B} - E_A^{A,AB} - E_B^{B,AB} \ , \tag{8}$$

where the superscript indicates the basis set and the geometry used, while the subscript refers to the monomer considered. Therefore, to give an example, $E_B^{B,AB}$ is the energy of the monomer B computed using its basis set (B) at the geometry the monomer assumes in the complex (AB). The level of theory here considered is the jun-ChS approach and B2PLYP-D3BJ/jun-cc-pVTZ is used for geometry optimizations. Once the deformation energy is computed, the equilibrium dissociation energy is straightforwardly derived as $D_e = E^{int}_{CP} - \Delta^{def}$, the results being summarized in Table 2.

The first three complexes ($FH_2P\cdots NH_3$, $FH_2P\cdots SH_2$, and FHS$\cdots$OFH) are

representative of pure or mixed pnicogen/chalcogen bonds, whose interaction is ruled more by electrostatic rather than dispersive forces. Reference values for geometries and interaction energies have been obtained at the CCSD(T)/aug-cc-pVTZ level, and reported in ref. [140]. It is noted that our results, collected in Table 2, lie in between the CP and NCP binding energies reported in ref. [140], with deviations from the denoted half-CP values, i.e. the arithmetic mean of CP and NCP results, smaller than 0.5 kJ mol^{-1}, thus supporting the idea that half-CP results provide the most reliable estimates when QC calculations do not account for the convergence to the CBS limit. For FHS$\cdots$OFH, a more accurate value, obtained at the CP-DLPNO-CCSD(T)/aug-cc-pV5Z level, is also given in Table 2 (see supplementary material of ref. [140]), and is in good agreement with our jun-ChS result, the difference being only 0.2 kJ mol^{-1} and thus well within the error bar of about 0.4 kJ mol^{-1} introduced by the DLPNO approximation.

The reference interaction energy for the $CH_3OH\cdots HCl$ adduct (see Table 2) was obtained by means of the composite approach defined as "Gold2", which assumes as best result the half-CP corrected MP2-F12/cc-pV(T,Q)Z-F12 combined with a high-order term derived from the difference between conventional, half-CP corrected CCSD(T) and MP2 results extrapolated to the CBS limit using the jul-cc-pVTZ and jul-cc-pVQZ basis sets [141]. The corresponding result, however, does not employ diffuse-function augmented basis sets in the MP2-F12 extrapolation to the CBS limit and does not include the core-valence correlation term and tight *d*-functions on the chlorine atom. Since the latter two contributions amount to about -0.3 kJ mol^{-1}, the remaining difference of $\sim$0.4 kJ$\cdot$mol^{-1} should be ascribed to the intrinsic limits of our scheme, to the fact that we consider CP-corrected energies (and not half-CP), and to possible differences in the reference geometries. Furthermore, the missing effect of diffuse functions in the MP2-F12 CBS extrapolation can also play a role, as demonstrated in ref [142].

2.4 Structural characterization: semi-experimental intermolecular parameters

Focusing on structural determinations, two different joint theory-spectroscopy strategies can be envisaged, which are referred to as "top-down" and "bottom-up". The first approach, denoted as the *semi-experimental approach* [57, 143, 144], relies on extracting from experimental outcomes the equilibrium structure details by using QC computations for providing the missing information. However, when molecular complexes are considered, it is very difficult to obtain a set of experimental data able to lead to a complete structural determination. In such cases, the range of application of the "top-down" approach can be further extended by employing either the *method of predicate observations* [145] or the *template approach* [57], in which QC computations are also used to complete the set of experimental data. Whenever the lack of experimental information is too extended, one can resort to the "bottom-up" approach, which consists in verifying the computed equilibrium geometry by means of a comparison between calculated and experimental spectroscopic parameters that probe structural characteristics.

As already mentioned, the spectroscopic information on geometries are mainly contained in rotational constants, but also nuclear quadrupole-coupling constants are extremely informative in this respect. Whenever a sufficient number of isotopologues is available, semi-experimental (SE) equilibrium structures can be obtained by a non-linear least-square fit of the SE equilibrium rotational constants, which are obtained from the experimental ground-state rotational constants by correcting them for computed vibrational effects:

$$B^e_{SE} = B^0_{exp} - \Delta B^{vib}_{comp} \ . \tag{9}$$

Several studies have shown that vibrational corrections computed using the hybrid B3LYP or, even better, double-hybrid B2PLYP functionals have the required accuracy to obtain reliable results [57, 145, 146, 147]. The application of such an approach has led to the development of a large database of semi-experimental equilibrium structures containing information for more than 60 species [57, 146, 148], which include the most relevant building blocks of prebiotic systems.

For large systems and/or non-covalent molecular complexes, a full set of deuterated species is often not experimentally available; therefore, semi-experimental X–H bond lengths cannot be derived. To overcome this issue, the database mentioned above can be used to exploit a linear regression approach (LRA), which allows for correcting DFT (B3LYP or B2PLYP) X–H bond lengths. Within this approach, a given computed parameter (i.e. the X–H bond distance) is augmented by the Δr contribution derived from the linear regression of the fit of semi-experimental structural parameters versus the corresponding theoretical ones. For CH, NH, and OH bonds, for which a statistically robust sample is available, the LRA mean average errors are well below 0.001 Å. A similar procedure can be applied to other common bond lengths, such as C–O and C–C.

While LRA can be applied to derive corrective terms for the principal classes of bond distances and angles, the situation can be more involved for non-covalent molecular complexes, for which very often only a limited number (or even just one) of isotopic data is experimentally available. In such circumstances, intramolecular geometrical bond lengths and angles can be obtained by the so-called template approach (TMA), which consists in correcting the structural parameters obtained at the DFT level (usually B2PLYP(-D3BJ) in conjunction with a triple-zeta quality basis set) by using either semi-experimental or high-level calculated values for a smaller system, referred to as templating model (TM):

$$r_e = r_e^{B2} + \Delta TM \ , \tag{10}$$

where r_e^{B2} is the structural parameter at the B2PLYP(-D3BJ) level for the system under investigation, and ΔTM is the difference between the B2PLYP(-D3BJ) equilibrium parameter and the corresponding reference value:

$$\Delta TM = r_e^{Ref}(TM) - r_e^{B2}(TM) \ . \tag{11}$$

In the case of non-covalent molecular adducts, the TMs are the monomers, for which SE structures are likely available in the above-mentioned database. If this is not the case,

CCSD(T)/CBS+CV or ChS geometry optimizations can provide the required reference parameters.

The intramolecular parameters being fixed at the corresponding LRA, TM, CCSD(T)/CBS+CV or ChS values, the number of experimental data is then sufficient to optimize the intermolecular parameters. In the few cases for which this is not the case, partial geometry optimizations of a few intermolecular parameters can be carried out at the ChS level.

In the following, the effectiveness of this general approach is illustrated by means of two test cases. The first system investigated in this respect is the planar formamide-water complex (see Figure 3 for structure and atom labeling), which is representative of a medium-strength hydrogen bond between the simplest model for the amide moiety present in peptides and proteins with a water molecule. The rotational constants of five different isotopologues have been reported in the literature, namely the parent species together with the ^{13}C, ^{15}N, D$_{cis}$(N-formamide) and D$_c$(O-water) singly-substituted isotopologues [149]. Although this set of isotopic species could allow a robust fit of several geometrical parameters, we have selected only three intermolecular parameters, namely the (N)H$_{cis}\cdots$O distance together with the N–H$_{cis}\cdots$O and H$_{cis}\cdots$O–H$_c$ angles. Both SE and ChS structures are available for both fragments, so that we can compare the accuracy of the TMA protocol proposed above using different TM models. The essential results are collected in Table 3. As a first remark, we note that B2PLYP-D3BJ and ChS structures of the fragments show a similar accuracy when compared to their SE counterparts; therefore, the bare B2PLYP-D3BJ or ChS TM models should be nearly equivalent in the analysis of the intermolecular parameters. Moving to the molecular complex, the situation remains essentially unvaried for intramolecular parameters, whereas the CP-B2PLYP-D3BJ/jun-cc-pVTZ level underestimated significantly the H$_{cis}\cdots$O intermolecular distance. In this connection, the results of Table 3 show that the ChS and SE C=O bond lengths are quite different; thus, this should be either fitted in the determination of the SE structure of the molecular complex or fixed within the TM model at the corresponding SE value of the formamide fragment. As far as intermolecular parameters are concerned, their values differ when ChS or SE intramolecular parameters are employed in the TMA procedure, with the differences being slightly above the estimated uncertainties. This points out that the error bar of the ChS geometries can lead to inaccuracies within about 0.01 Å and 1 degree for intermolecular distances and angles, respectively. At the same time, CP-B2PLYP-D3BJ/jun-cc-pVTZ structures coupled to the TMA procedure employing the SE structures of the fragments allows for obtaining accurate intermolecular parameters without any highly expensive QC (e.g. CCSD(T)) computations and already when only two isotopic substitutions are available from the experimental investigation. The vibrational corrections to rotational constants required for the determination of the SE parameters have been computed at the CP-B3LYP-D3BJ/SNSD level.

To further test the accuracy of the proposed procedure, we have investigated the complex formed by dimethylsulfide (DMS) and SO_2 (see Figure 4 for structure and atom

Table 3. Structural parameters of the planar formamide-water complex, formamide, and water.[a]

Parameter		SE structure		
Formamide-H_2O	CP-B2PLYP-D3BJ/ jun-cc-pVTZ	CP-B2PLYP-D3BJ/[b] jun-cc-pVTZ	ChS TM[c]	SE TM[d]
r(C=O)	1.2246	1.2246	1.2198	1.2225
r(C-N)	1.3462	1.3462	1.3428	1.3432
r(C-H)	1.0979	1.0979	1.0966	1.0943
r(N-H_{cis})	1.0114	1.0114	1.0090	1.0141
r(N-H_{trans})	1.0024	1.0024	1.0000	1.0168
∠(OCN)	125.08	125.08	124.88	124.51
∠(HCN)	113.45	113.45	113.65	113.32
∠(CNH_{cis})	119.63	119.63	119.42	120.14
∠(CNH_{trans})	120.27	120.27	120.20	119.61
r(O-H_c)	0.9745	0.9745	0.9703	0.9728
r(O-H_b)	0.9586	0.9586	0.9543	0.9569
∠(HOH)	107.02	107.02	106.80	105.77
r($H_{cis}\cdots$O)*	2.0573	2.197(2)	2.145(2)	2.150(1)
∠(N-$H_{cis}\cdots$O)*	136.13	131.16(8)	133.50(8)	132.55(5)
∠($H_{cis}\cdots$OHc)*	82.49	69.6(5)	69.7(4)	70.1(2)
σ^2		1.77×10^{-2}	4.53×10^{-3}	1.96×10^{-3}
Formamide	B2PLYP/ jun-cc-pVTZ	ChS	SE structure[e]	
r(C=O)	1.2141	1.2093	1.212	
r(C-N)	1.3571	1.3536	1.354	
r(C-H)	1.1006	1.0993	1.097	
r(N-H_{cis})	1.0053	1.0028	1.008	
r(N-H_{trans})	1.0027	1.0002	1.017	
∠(OCN)	124.77	124.57	124.2	
∠(HCN)	112.53	112.73	112.4	
∠(CNH_{cis})	119.39	119.19	119.9	
∠(CNH_{trans})	121.16	121.09	120.5	
H_2O	B2PLYP/ jun-cc-pVTZ	ChS	SE structure[f]	
r(O-H)	0.9605	0.9563	0.9573	
∠(HOH)	104.78	104.56	104.53	

[a] Stars denote the fitted intermolecular parameters of the formamide-water complex.

[b] Non-fitted parameters fixed at the CP-B2PLYP-D3BJ/jun-cc-pVTZ level.

[c] Non-fitted parameters fixed at the scaled values using the ChS structure in the TMA procedure.

[d] Non-fitted parameters fixed at the scaled values using the SE structure in the TMA procedure.

[e] Taken from ref. [97].

[f] Taken from ref. [57].

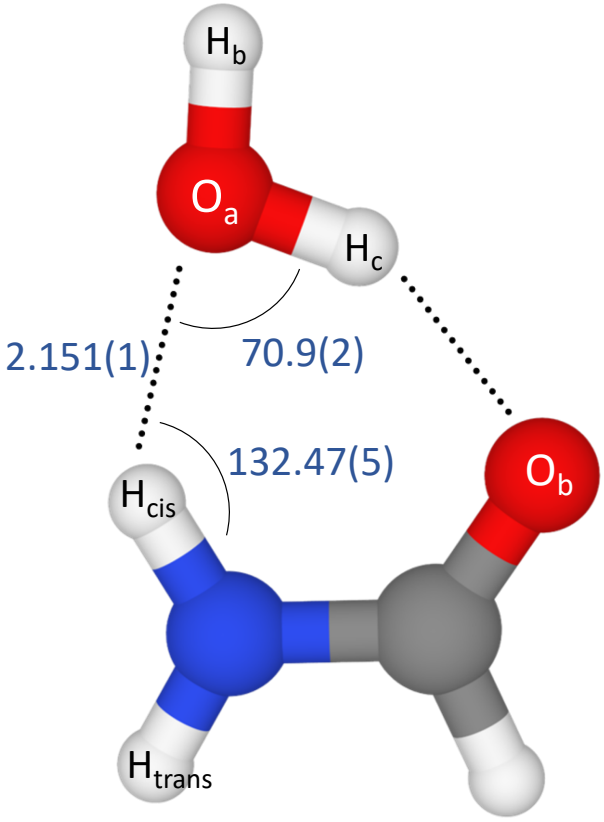


Figure 3. The formamide-water complex: structure and atom labeling. The three SE intermolecular parameters determined using the SE TM model are reported (distance in Å, angles in degrees).

labeling), which involves a different kind of intermolecular interaction (chalcogen bond) and heavier atoms. The rotational constants for seven different isotopic species have been reported in the literature [13], namely the singly substituted ^{34}S (on both S), ^{33}S (on both S), ^{13}C, ^{18}O isotopologues in addition to the main species. To exploit the SE approach, the vibrational corrections to rotational constants have been computed at the B2PLYP-D3BJ/maug-cc-pVTZ-*d*H level. Although this set of data allows for determining also intramolecular structural parameters (see ref [13]), for the present fit, we have selected four intermolecular parameters, namely the S···S distance, the C–S···S and O–S···S valence angles, and the C–S···S–X dihedral angle, with X defining the bisector of the ∠OSO angle. The results, collected in Table 4, confirm the general trends discussed above; however, in this case, ChS and SE TM models lead to very similar outcomes because the ChS structures of both monomers are remarkably accurate. As mentioned above, in ref [13], two intramolecular parameters (the C–S and S–O bond lengths in DMS and SO_2, respectively) were also determined. It is noteworthy that for both distances the TMA estimates are well within the final error bars, whereas this is not the case for the CP-B2PLYP-D3BJ/jun-cc-pVTZ values.

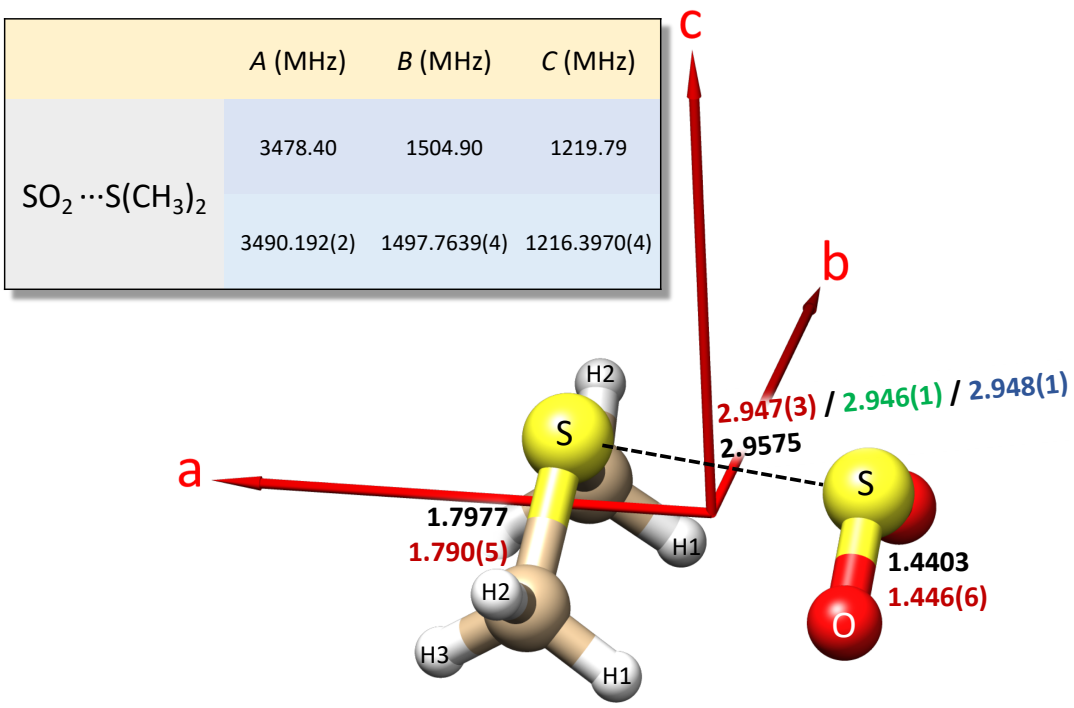


Figure 4. The DMS-SO_2 complex: selected structural parameters (bond distances in Å). Computed (ChS level) and SE values in black and in dark red, respectively, from ref [13]. The SE S···S intermolecular distance from the TMA procedure is also reported: in blue and green the results from the ChS TM and SE TM models, respectively. The principal inertia axes are also shown. In the inset, the comparison between computed (ChS equilibrium values augmented by vibrational corrections at the B2PLYP-D3BJ/maug-cc-pVTZ-*d*H level) and experimental rotational constants are given.

In conclusion, the availability of a large database of SE geometries (for potential monomers) and the effectiveness of the ChS composite method for missing structural parameters paves the way to the determination of accurate structures for non-covalent molecular complexes employing a reduced number of isotopic species, while exploiting the effective B2PLYP-D3BJ computational model for geometry optimizations and anharmonic force field calculations (for vibrational corrections to rotational constants). For the latter, low-cost B3LYP-D3BJ calculations carried out in conjunction with a double-zeta quality basis set are already show the required accuracy.

3 Spectroscopic characterization: Results for some case studies

The qualitative and quantitative description of non-covalent interactions is still a great challenge for all areas of physical chemistry. If, on the one side, the investigation of structural and spectroscopic properties of intermolecular systems can shed light on the characterization of molecular complexes, this is non-trivial and a combined rotational spectroscopy - quantum chemistry approach has been suggested as the best strategy to be pursued. From the results discussed in the previuous sections, it is apparent that

Table 4. Structural parameters of the DMS-SO_2 complex, DMS, and SO_2.[a]

Parameter			SE structure				
DMS-SO_2	CP-B2PLYP-D3BJ/ jun-cc-pVTZ	ChS	CP-B2PLYP-D3BJ/ [b] jun-cc-pVTZ	ChS TM[c]	CCSD(T)/ CBS+CV TM[d]	SE TM[e]	Ref. [13]
r(S$\cdots$S)*	2.9344	2.9575	2.932(4)	2.946(1)	2.953(2)	2.948(1)	2.947(3)
r(S-O)	1.4535	1.4403	1.4535	1.4423	1.4337	1.4401	1.446(6)
∠(OSS)*	95.21	94.32	95.46(2)	95.11(7)	95.0(1)	95.07(7)	95.0(2)
∠(O-S-S-X)	-121.10	-121.16	-121.10	-121.18	-120.73	-121.08	
r(C-S)	1.8040	1.7977	1.8040	1.7963	1.7935	1.7951	1.790(5)
∠(CSS)*	91.33	91.13	91.4(2)	91.58(7)	91.6(1)	91.59(7)	91.7(2)
∠(C-S-S-X)*	129.93	130.37	132.5(8)	130.7(3)	129.9(4)	130.5(3)	
r(H_1-C)	1.0885	1.0885	1.0885	1.088	1.0887	1.0899	
∠(H_1CS)	110.68	110.55	110.68	110.48	110.55	110.44	
∠(H_1CSC)	-65.53	-64.17	-65.53	-65.41	-65.43	-65.32	
r(H_2-C)	1.0865	1.0863	1.0865	1.086	1.0867	1.0880	
∠(H_2CS)	107.14	107.33	107.14	106.94	107.30	107.28	
∠(H_2CSH_1)	-118.65	-118.77	-118.65	-118.77	-118.75	-118.86	
r(H_3C)	1.0871	1.0870	1.0871	1.0868	1.0874	1.0886	
∠(H_3CS)	109.93	109.82	109.93	109.73	109.79	109.68	
∠(H_3CSC)	57.03	58.16	57.03	56.92	56.93	56.83	
σ^2			1.52×10^{-2}	1.47×10^{-3}	3.41×10^{-3}	1.66×10^{-3}	
SO_2	B2PLYP-D3BJ/ jun-cc-pVTZ	ChS	CCSD(T)/[e] CBS+CV	SE structure [f]			
r(O-S)	1.4443	1.4331	1.4245	1.4308			
∠(OSO)	119.28	119.11	120.01	119.31			
∠(OSXX)	-120.34	-120.43	-119.98	-120.33			
DMS	B2PLYP-D3BJ/ jun-cc-pVTZ	ChS	CCSD(T)/[g] CBS+CV	SE structure [h]			
r(C-S)	1.8075	1.7998	1.7970	1.7986			
∠(CSC)	99.03	98.41	98.55	98.58			
r(H_2-C)	1.0870	1.0865	1.0873	1.0886			
∠(H_2CS)	107.28	107.44	107.44	107.42			
r(H_1-C)=R(H_3-C)	1.0882	1.0879	1.0885	1.0897			
∠(H_1CS)=∠(H_3CS)	110.93	110.73	110.80	110.69			
∠(H_1CSH_2)=∠(H_3CSH_2)	-118.85	-118.96	-118.95	-119.05			
∠(CSXX)	130.47	130.77	130.80	129.15			
∠(H_1CSC)=∠(H_3CSC)	-61.15	-61.04	-61.05	-60.95			

[a] Stars denote the fitted intermolecular parameters of the DMS-SO_2 complex. X defines the bisector of the ∠OSO angle.

[b] Non-fitted parameters fixed at the CP-B2PLYP-D3BJ/jun-cc-pVTZ level.

[c] Non-fitted parameters fixed at the scaled values using the ChS structure in the TMA procedure.

[d] Non-fitted parameters fixed at the scaled values using the CCSD(T)/CBS+CV structure in the TMA procedure.

[e] Non-fitted parameters fixed at the scaled values using the SE structure in the TMA procedure.

[e] Taken from ref. [150].

[f] M. Mendolicchio, private communication.

[g] Taken from ref. [151].

[h] Taken from ref. [152].

rotational spectroscopy is able to provide accurate information on structural parameters [3, 29, 30, 153, 154], but it requires to be strongly complemented by theory. For a quantitative analysis of interaction energies, one has instead to mostly rely on theory.

3.1 Non-covalent pnicogen-bond and chalcogen-bond interactions

In the last years, the inventory of donor-acceptor moieties involved in non-covalent interactions has been considerably expanded from the traditional hydrogen bond

involving electronegative donors (NH, OH) and acceptors (N, O) to low-electronegative donors (CH, PH, SH) and low-electronegative or π-acceptors, but also to the halogen bond and to attractive directional interactions between atoms of groups 5 and 6. The group-name of the electrophilic atoms is used to distinguish these last types of interaction from the cases where they participate as nucleophilic electron donors. Pnicogen and chalcogen bonds are established by interactions of σ- and π-hole regions (or, in a NBO picture, by the charge transfer from lone-pairs of the donor to an antibonding orbital of the acceptor) and, similarly to hydrogen bonds, they include variable amounts of attractive electrostatic, polarization, charge-transfer and dispersion interactions, which more than counterbalance the exchange repulsion. However, contrary to other cases, a comprehensive analysis of the structural and thermochemical characteristics of pnicogen and chalcogen bonds is still lacking and can strongly benefit from integrated spectroscopic and quantum-chemical studies. In this framework, we have recently investigated two prototypical systems involving either pnicogen or chalcogen bond, namely the trimethylamine-nitroethane (TMA-NE) complex [12] and the DMS-SO_2 adduct [13], respectively.

As mentioned above, along with the well-established hydrogen and halogen bonds, emerging classes of non-covalent interactions have attracted an increasing attention. Among them, the R–X$\cdots$Y interactions [17], with R–X (with X=N, P, As, Sb or Bi) and Y being the pnicogen bond donor and acceptor, have been recently characterized for the cases of heavy pnicogen$\cdots\pi$ interactions [155] and of the P$\cdots$P [156], P$\cdots$N [18, 157], and P$\cdots$O non-covalent bonds [18]. However, much less information is available for nitrogen acting as pnicogen bond donor [18, 158]. The dual nature of nitrogen as pnicogen bond donor and, more widely, as bond acceptor (such as in hydrogen and halogen bonds) has raised a number of questions concerning (*i*) the possibility of experimentally observing a complex showing the nitrogen-nitrogen non-covalent bond in the gas phase, (*ii*) the competition of the homo pnicogen non-covalent bond with other non-covalent interactions, (*iii*) the magnitude of the binding energy with respect to that typically observed for hydrogen bonds, and –finally– (*iv*) the which contributions (dispersions, electrostatics, or polarization) stabilize the molecular adduct. To answer all these questions the TMA-NE complex has been investigated by means of a combined experimental-computational approach by exploiting, on both sides, state-of-the-art methodologies.

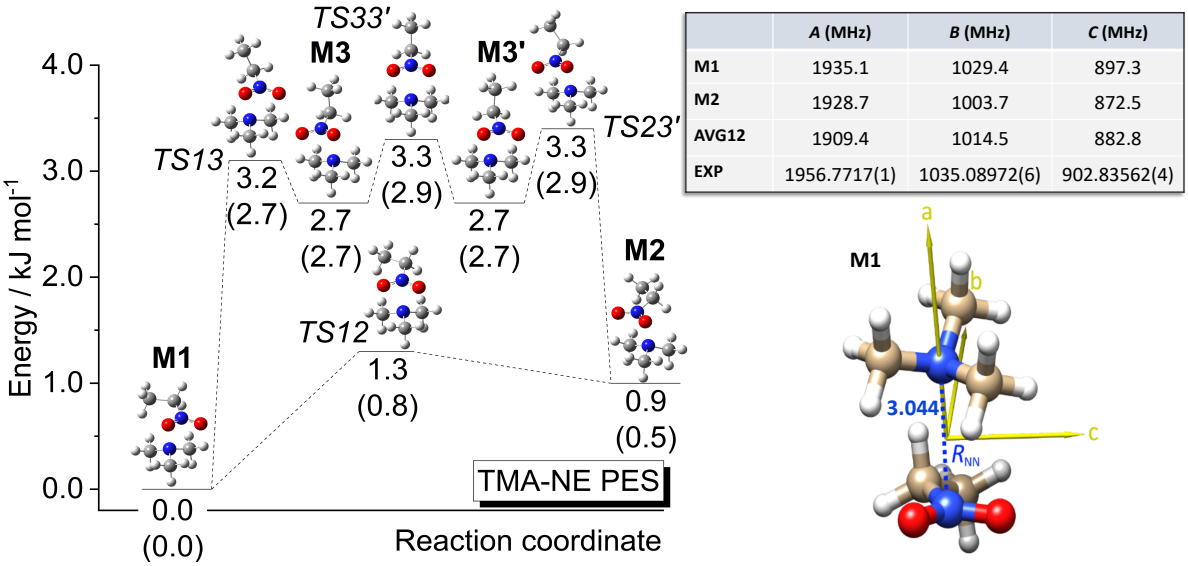
The simplest tertiary amine (TMA) is an effective partner in pnicogen bonds because the lack of NH bonds avoids any competition with strong hydrogen bonds. At the same time, the negatively charged nitrogen atom of TMA should lead to a stabilizing electrostatic interaction with the positively charged nitrogen of nitroethane (NE), whose low-energy N–O antibonding orbitals should be also effective acceptors for the lone-pair electrons of the TMA nitrogen. This picture was further validated by a NBO analysis, which not only confirmed the orbitals involved in the non-covalent interaction, but also showed a charge flux and an electrostatic interaction. Indeed, it was found that the NBO charges on the two interacting nitrogens vary upon complexation:

the charge on the nitrogen atom belonging to TMA becomes more negative, while that on the NE nitrogen becomes more positive. The NBO analysis furthermore pointed out that the nitrogen-nitrogen non-covalent interaction largely overcomes the weaker C–H···O and C–H···N hydrogen bonds, with the overall binding energy –computed at the ChS level (also incorporating zero-point vibrational energy, ZPE, and CP correction)– being ∼24 kJ mol^{-1}, which is of the same order of magnitude as those of relatively strong neutral hydrogen bonds. The value is in line with the less accurate one obtained from the SAPT analysis (at the SAPT2+3/aug-cc-pVDZ-RI level), thus confirming the reliability of this perturbative approach. By inspecting the single contributions, it is noted that the dominant role is played by the repulsive exchange term, which is nearly compensated by the attractive induction and dispersion components, the electrostatic term thus ruling the overall interaction energy.

To guide the experimental investigation, the potential energy surface of the TMA-NE complex was explored, with four different isomers being characterized (see Figure 5). However, two of them (which are equivalent and separated by a low energy barrier) resulted to be quite unstable because of the very small barrier ruling their relaxation to the two more stable structures. The latter two isomers were found nearly isoenergetic and separated by a quite small energy barrier. As a consequence, the large amplitude motion governing their interconversion should be properly sampled when determining the effective experimental structure. However, the overall mass distribution and the key structural parameters of the two stable isomers are actually very similar. In turn, this leads to very similar rotational constants, so that the averaged values are also close to those of each one of them (see the inset of Figure 5). The computed spectroscopic parameters were used to guide the investigation of the rotational spectrum of the TMA-NE adduct, which pointed out the presence of only one isomer. Even if QC computations were not able to clarify whether one of the two stable isomers or, more likely, an averaged structure was observed, the very good agreement between calculated rotational and nuclear quadrupole-coupling constants demonstrates the accuracy and reliability of the computed structures, thus allowing us to point out that the intermolecular N···N distance lies in the 3.044-3.051 Å interval.

As previously mentioned, non-covalent chalcogen-bond interactions play a role in different fields [31, 32, 33, 34, 159]. For this reason, an increasing effort is being devoted to a better characterization of the directionality, strength, and nature of such interactions [160]. Among chalcogens, sulfur can act either as a bond donor (due to a σ- or π-hole on sulfur) or as acceptor (because of a lone-pair on sulfur, as in thioethers). In this respect, DMS is a good candidate as acceptor because the lack of SH bonds avoids any competition between chalcogen and hydrogen bonds . On the other hand, the strongly positive sulfur atom of SO_2 should lead to significant electrostatic interaction with the negatively charged sulfur in DMS. Therefore, the DMS-SO_2 complex has been studied by combining rotational spectroscopy in supersonic expansion and state-of-art QC calculations.

QC calculations, carried out at the ChS level, pointed out that the equilibrium



	A (MHz)	B (MHz)	C (MHz)
M1	1935.1	1029.4	897.3
M2	1928.7	1003.7	872.5
AVG12	1909.4	1014.5	882.8
EXP	1956.7717(1)	1035.08972(6)	902.83562(4)

Figure 5. Left panel: potential energy surface of the TMA-NE complex showing the structures of the four minima (M1, M2, M3, M3') and the transition states ruling their interconversion; the ChS relative electronic energies, also corrected by harmonic ZPE (within parentheses, at the B2PLYP-D3BJ/maug-cc-pVTZ-*d*H level), are reported. In the inset (right, top panel), the computed (B2PLYP-D3BJ/maug-cc-pVTZ-*d*H equilibrium values augmented by vibrational corrections at the B3LYP-D3BJ/SNSD level) rotational constants for M1, M2 and their averaged structure (Avg12) are compared to experimental values. Right panel (bottom): the structure of the M1 isomer together with the computed N···N intermolecular distance (in Å) is shown.

structure of the DMS-SO_2 complex is characterized by a stacked geometry with one of the lone pairs of the Lewis base oriented towards the σ-hole of SO_2 and each methyl groups facing one oxygen atom with one hydrogen aligned to maximize a secondary CH···O interaction (see Figure 4). The corresponding rotational constants, together with other computed spectroscopic parameters, guided the investigation of the rotational spectrum, whose analysis confirmed the presence of only one isomer as well as the accuracy and reliability of the computed structure (see the inset of figure 4). As mentioned above (and already discussed in section 2.4), the characterization of the rotational spectra for seven isotopologues allowed the determination of a partial semi-experimental structure. The semi-experimental S···S distance obtained (2.947(3) Å; see Figure 4) is, interestingly, 0.65 Å shorter than the sum of the van der Waals radii, thus confirming the presence of a S···S interaction. Moreover, this difference is significantly larger than that observed for the $(CH_3)_2O\cdots SO_2$ complex, for which the retrieved S···O distance of 2.884(2) Å is 0.44 Å shorter than the corresponding sum of the van der Waals radii.

The binding energy of DMS-SO_2, computed at the ChS level (also including CP and ZPE corrections), amounts to 23.5 kJ mol^{-1}, which results to be comparable to that of the TMA-NE complex and to typical hydrogen bonds. To inspect the nature of the S···S non-covalent interaction (but also to further confirm it), different analyses have been

carried out. To disclose the role played by different contributions, the SAPT analysis –at the SAPT2+(CCD)δMP2 level in conjunction with the aug-cc-pVTZ basis set – was first of all considered. This pointed out that the electrostatic interaction is by far the largest contribution, while dispersion and induction are comparable in magnitude, and altogether are larger than the exchange contribution, with the overall binding energy being in agreement with the ChS value. The NBO analysis showed that the largest contribution stabilizing the molecular complex is the interaction between the S lone-pair of DMS and the S–O antibonding orbital of SO_2. Furthermore, the NBO charges point out a large charge transfer flowing from the DMS sulfur atom, through sulfur in SO_2, to its oxygen atoms, which therefore increase their negative charges. This outcome was also confirmed by the analysis of the changes occurring in the ^{33}S nuclear quadrupole coupling constants and in the potential barriers for the rotation of the methyl groups as a consequence of the DMS complexation. In both cases, the observed changes were found to be good diagnostic of a partial charge transfer mentioned above.

In summary, the two examples analyzed above were able to demonstrate the ability of a gas-phase combined rotational spectroscopy - quantum chemistry investigation in unveiling the nature of non-covalent intermolecular interactions from both structural and energetic points of view.

3.2 From hydrogen bonding to microsolvation

Non-covalent interactions as those involved in solvation play a decisive role in nature. For example, in biomolecules in aqueous environment, the surrounding molecules strongly affect the molecular structure and biological functions. Other examples concern water mediation in molecular recognition and protein folding (see, e.g., refs [161, 162, 163]) as well as confomational and tautomeric equilibria (see, e.g., refs [164, 165]).

To bridge the gap between gas-phase and condensed-phase behavior, the best approach relies on the investigation of microsolvated clusters, which represent the initial step of the hydration process and allow for probing water-solution interactions without additional environmental effects. By definition, microsolvation is the situation in which an isolated molecule is surrounded by a reduced number of water molecules and which can be experimentally created in a supersonic jet [166]. To understand the solvation forces at the molecular level, a step by step approach can be devised, which requires the study of the molecule of interest in an isolated environment, i.e. in the gas phase, as well as by varying the degree of hydration. Going more into details, first, the molecular species is characterized in the gas phase in order to derive its intrinsic properties. Then, the molecular complexes formed with an increasing number of water molecules are investigated.

In this section, some case studies of increasing complexity will be described: from the molecular complexes formed by one molecule that can be seen as a biological model system and one solvent molecule to molecular complexes containing several water units.

3.2.1 Hydrogen bonds. Before addressing the issue of microsolvation, the characterization of molecular complexes formed with one single water molecule is deserved. An important feature of the hydrogen bond is its cooperativity, i.e., the fact that the local hydrogen bond strength is influenced by the neighboring water molecules [167]. Therefore, hydrogen bond cooperativity affects, for example, the vibrational spectrum of the O–H stretches [168], properties of low-temperature liquid phases [169], and site-specific reaction rates for chemical processes like acid dissociation in water clusters [170]. By definition, the "net cooperativity" that affects a local hydrogen bond is given by the sum of the number of occupied hydrogen bond positions in the donor and acceptor water that lead to cooperative strengthening of the bond, to which the number of the so-defined "anticooperative" occupied sites that weaken the hydrogen bond should be subtracted. As a consequence, hydrogen bond cooperativity produces distinctive distortions in the oxygen atom framework geometries, which in turn lead to variations in the hydrogen bond strength by a factor of two or more. As a result, the hydrogen bond network geometry can influence reaction rates of molecules "solvated" in a cluster geometry [170].

Hydrogen bond cooperativity has also been identified as one of the most important factors that contribute to the stabilization of hydrated biomolecules. Since the hydration of the nucleic acids controls their structure and mechanism of action, the study of the interaction of water with the individual nuclear bases represents the first step toward the understanding of the effects of hydration on DNA. As an example, we consider the computational investigation of the complex formed by 2-thiouracil (2TU), the most stable isomer of thiouracil, and one molecule of water [24]. The interest on thiouracil and its derivatives is due to the fact that they are minor components in natural t-RNAs and represent an example of biosystem including the thiocarbonyl (–C(=S)–) bond pattern.

To accurately determine the structural and spectroscopic properties of the 2TU-water complex, the ChS approach was exploited, thus allowing for pointing out the geometrical modification occurring upon complexation. In Figure 6, some selected structural parameters for both the adduct and the isolated molecule are shown. It is noted that complexation with water determines an evident shrinking of the C=S and N–H distances by more than 0.01 Å, which is –in relative terms– about 1%. On the contrary, the parameters not involved in the formation of intermolecular bonds are marginally affected, with variations on the order of 0.002-0.003 Å, at most. In the inset, the four lowest-energy complexes are depicted. Analogously to uracil [171], the conformer denoted as W1 was found to be the most stable with a population of about 90%. The energies reported have been obtained at the B2PLYP/aug-cc-pVTZ level on top of B3LYP/SNSD geometries, and their accuracy is expected to be –in relative terms– within 1 kJ mol^{-1} of the "cheap" counterpart, once the CP correction is incorporated.

As well-known, the structures of proteins and other biological molecules are determined by a delicate balance between several molecular interactions [172, 173], which also include weak O–H$\cdots\pi$ contributions. While strong hydrogen bonds are well characterized, the latter are less explored. To fill this gap, the O–H$\cdots\pi$ interaction

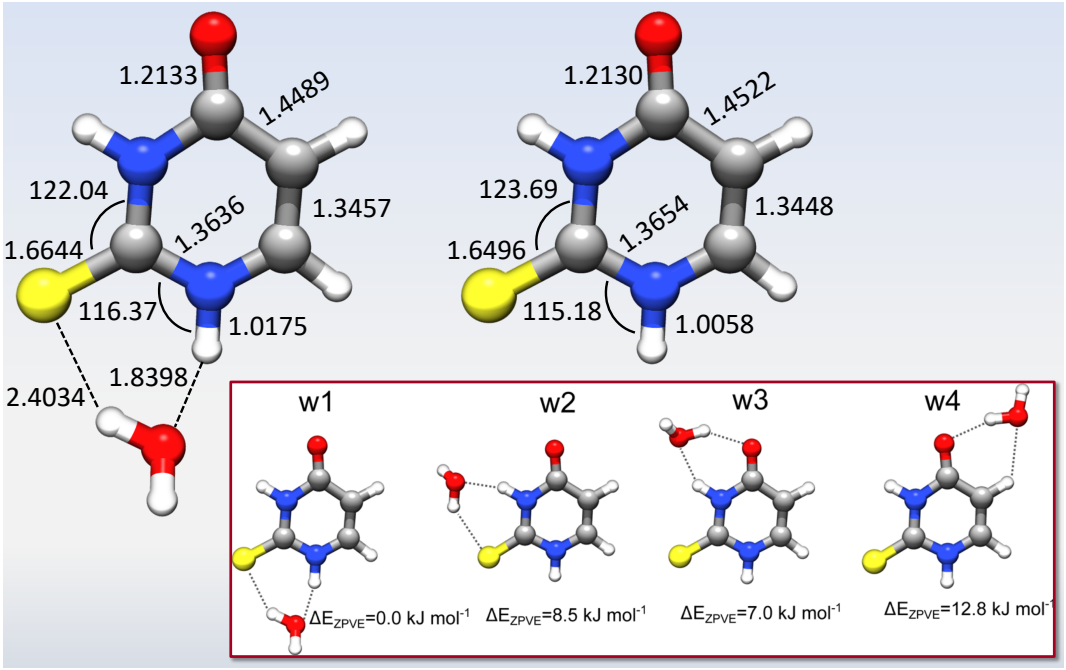


Figure 6. Comparison of selected structural parameters for the 2TU-water complex and the 2TU isolated molecule. In the inset, the structures and relative energies (at the B2PLYP/aug-cc-pVTZ level) of the four most stable 2TU-water complexes are shown.

occurring in cycloalkene-water, C_nH_{2n-2}-H_2O (with n = 3-7), adducts has been recently investigated [14], where the choice of cycloalkanes as model systems also allow for finding a relation between the size of the carbon ring and the binding energy of the water complexes.

From the inspection of Figure 7 (left panel), which summarizes the results of SAPT2+3/aug-cc-pVDZ-RI calculations, the role played by the dimension of the cycloalkene ring in determining the intermolecular interaction energy and a quantitative understanding of the chemical nature of the non-covalent interaction can be pointed out. This Figure shows that the largest term stabilizing the cluster is the electrostatic contribution, whose absolute value monotonically increases along the series. The same behavior is presented by the other two stabilizing terms, induction and dispersion, the latter being only a few kJ mol^{-1} smaller, in absolute terms, than the electrostatic contribution. The repulsive exchange term also shows an increasing trend by enlarging the size of the cycloalkene monomer. Overall, the total intermolecular interaction energy, but also the dissociation energy (at the B2PLYP-D3BJ/maug-cc-pVTZ-dH level, also accounting for CP and ZPE corrections) that follows a similar trend, increases with the dimension of the cycloalkene ring and then reaches a nearly constant value for the series with $n \geq$5. This suggests that –when $n \geq$5– the steric strain in the cycloalkene family no longer affects the O–H$\cdots\pi$ interaction. For this reason, in ref. [14], the cyclopentene-water adduct, which is therefore the smallest member of the cycloalkene-water complex series showing a converged behavior with respect to interaction and dissociation energies,

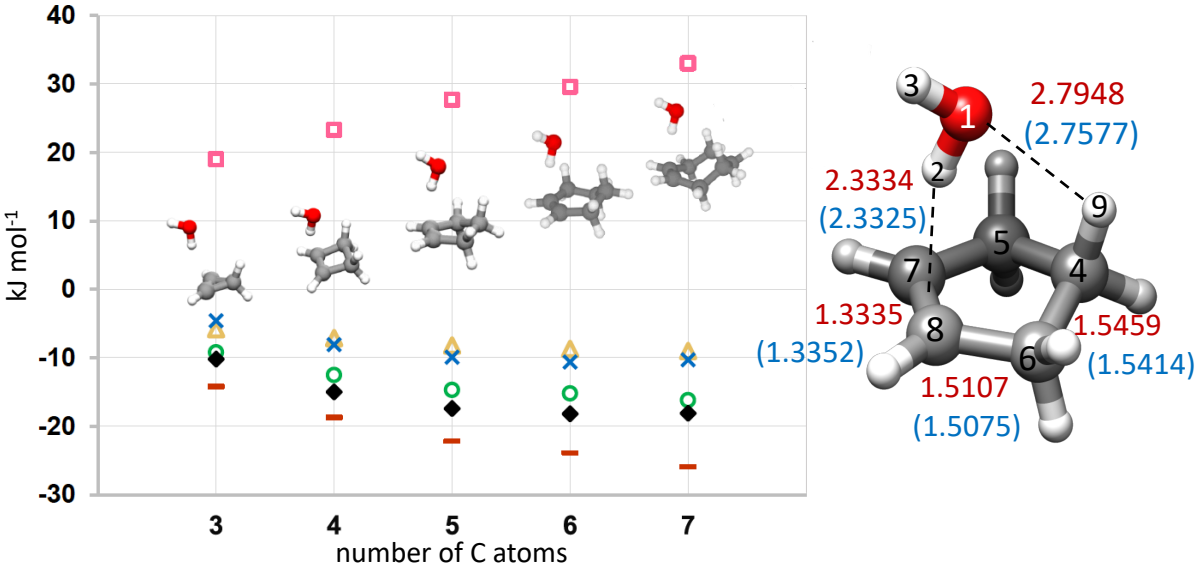


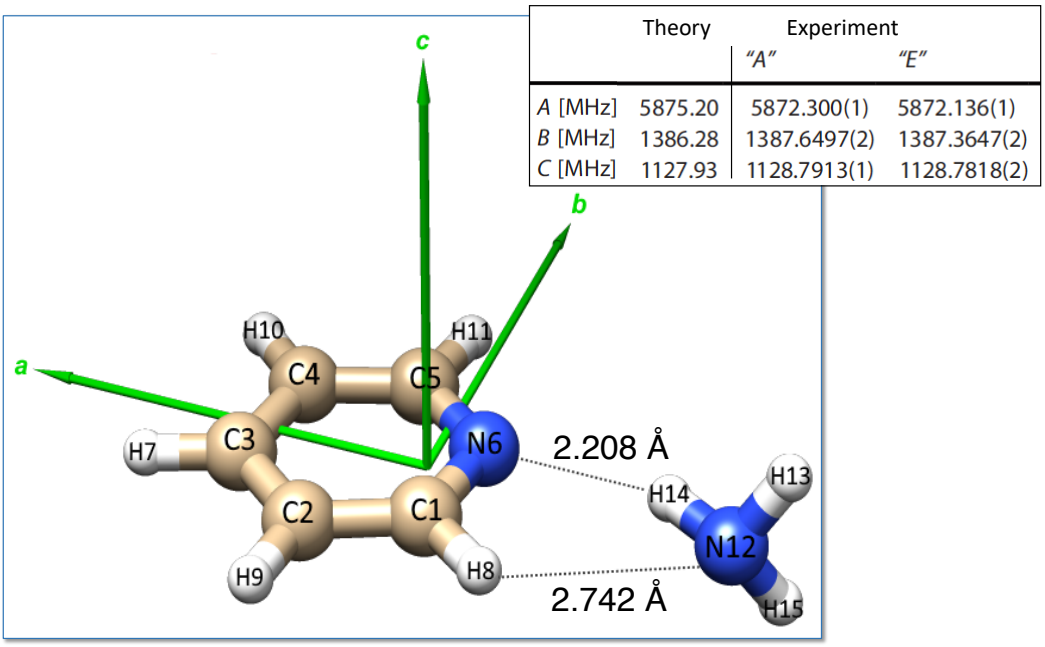
Figure 7. Left panel: Exchange-repulsion (pink squares), induction (orange triangles), dispersion (green circles), and electrostatic (red dashes) contributions, and total intermolecular interaction energies (black diamonds) together with dissociation energies at the CP- and ZPE-corrected B2PLYP-D3BJ/maug-cc-pVTZ-*d* level of theory (blue crosses) as a function of the number of carbon atoms in cycloalkene-water complexes. Right panel: Selected semi-experimental equilibrium structural parameters in dark red compared to the computed counterparts at the ChS level.

was selected for a combined rotational spectroscopy - quantum chemistry investigation.

The right panel of Figure 7 graphically depicts the structure of the most stable isomer of the cyclopentene-water complex, for which the assignment of the rotational spectra of four isotopic species (the parent isotopologue together with the adducts formed by ^{18}O-water, mono-deuterated and bi-deuterated water) allowed the determination of a partial semi-experimental equilibrium structure. Some selected parameters are reported in Figure 7 and compared with the corresponding ChS values. This figure clearly points out that the O–H$\cdots\pi_{C=C}$ interaction is assisted by a weaker hydrogen bond formed between the oxygen lone pair of the water molecule and the closest apical methylene hydrogen atom of cyclopentene. As confirmed by a NBO analysis, this secondary interaction occurs only for the n $\geq$5 elements of the series considered.

Of great interest is also the investigation of the non-covalent interactions occurring in systems formed by a prototype heterocyclic aromatic system with other solvent molecule than water. In this respect, we can mention the recent study of the pyridine-amonia complex in the gas phase [103]. In pyridine, the simultaneous presence of n and π orbitals makes the molecule particularly suitable for the formation of non-covalent linkages, that can be unravelled by an integrated quantum-chemical and rotational spectroscopy strategy.

The joint experimental-computational investigation of ref [103] incontrovertibly



	Theory	Experiment	
		"A"	"E"
A [MHz]	5875.20	5872.300(1)	5872.136(1)
B [MHz]	1386.28	1387.6497(2)	1387.3647(2)
C [MHz]	1127.93	1128.7913(1)	1128.7818(2)

Figure 8. Orientation of the pyridine-ammonia cluster with respect to the principal axes of inertia a, b, and c, also showing the N–H···N and C–H···N hydrogen bonds. In the inset, the comparison between computed (equilibrium ChS rotational constants augmented by B2PLYP-D3BJ/maug-cc-pVTZ-dH vibrational corrections) and experimental ground-state rotational constants.

pointed out that the pyridine-NH_3 complex shows a σ-type bonding structure in which ammonia interacts through two hydrogen bonds within the plane of the pyridine ring (see Figure 8), with a small degree of proton transfer from ammonia to pyridine. On the contrary, no evidence of the formation of a stable π-bound adduct was found. The reliability of the computational investigation carried out in ref [103] is confirmed by the very good agreement between computed and experimental rotational constants, as seen in the inset of Figure 8, the maximum discrepancy being well within 0.1%. The computational study was also able to explain why only one isomer was observed in the gas phase mixture. In fact, even if three stable isomers were located on the potential energy surface, the interconversion barriers connecting the two higher energy minima to the most stable species are so low that they clearly relax to the latter.

3.2.2 Microsolvation. Despite the fact that various spectroscopic techniques can be and have been employed in studying isolated microsolvated clusters (see, e.g., refs [174, 175, 176]), rotational spectroscopy is the leading technique because it can provide detailed structural information, thus allowing for gaining deeper insight on the effect of microsolvation (see, e.g., refs [4, 149, 177, 178, 3, 179, 180, 181]).

Two studies have been selected to investigate the interaction of model system with water: 2-azetidinone interacting with one and two water molecules [4] and β-propiolactone complexed with an increasing number of water molecules, from one to

five [3].

The first system has been chosen because in 2-azetidinone, the peptide functional group is involved in a four-membered ring, which therefore constrains a *cis* configuration. For this reason, it can be considered an appropriate model for investigating the interaction of the *cis* peptide group with water. The investigation of the rotational spectrum of the water complexes has been supported by QC calculations at the MP2/6-311++G(2d,p) level of theory, with all details provided in ref [4]. The *cis* configuration of the amide group in 2-azetidinone allows the interaction of the water molecule with the N–H and C=O groups simultaneously by means of two hydrogen bonds, thus forming a six membered ring. Therefore, this conformer could be considered as a model of interaction for molecules containing the amide group with a *cis* configuration.

The observation and assignment of the rotational spectra for 9 isotopic species allowed the authors to determine the effective structure (r_0) from the fit of the experimental ground-state rotational constants. The structural parameters obtained can be compared with those analogously derived for the formamide-water complex [149]. Since the strength of the hydrogen bonds can be related to their bond distances and the distortion of the hydrogen bond angles, it can be concluded that both hydrogen bonds, namely C=O$\cdots$H and N–H$\cdots$O–H, are stronger in formamide-H_2O than in 2-azetidinone-H_2O. Another important conclusion is that the dominant hydrogen bond in the complex of *cis*-2-azetidinone-H_2O is the C=O$\cdots$H–O interaction.

For the amide group, two resonant structures exist and hydrogen-bond interactions with water may increase the contribution of the ionic resonant form, the latter being known as resonance-enhanced cooperative effect or π-bond cooperative effect [182]. In agreement with this, a decrease of the C–N distance and an increase of the C=O distance by -0.013 Å and 0.011 Å, respectively, were predicted by QC computations for the *cis*-2-azetidinone-H_2O complex and by -0.019 Å and 0.016 Å, respectively, for the 2-azetidinone complex with two water molecules. For the latter adduct, three hydrogen bonds exist in the molecular complex: one linking one of the water molecules to the N–H group, another linking the second water molecule to the C=O group and a third formed between the two water molecules. Its structure is therefore denoted as distorted "chemically substituted water trimer" [183], where the non-bonded hydrogen atoms of water adopt an up-down configuration similar to that found in related complexes.

To gain information on the effect of increasing the number of water molecules, the r_0 geometries of the mono- and bi-hydrated 2-azetidinone complexes can be compared. It is noted that the addition of a second water molecule determines some interesting changes. The most important difference is the decrease in the hydrogen bond distances, with interesting insights obtained by inspecting the dissociation energy (D_e). The D_e of the *cis*-2-azetidinone-H_2O and *trans*-2-azetidinone-H_2O complexes, computed at the MP2/6-311++G(2d,p) level, are 34.9 kJ mol^{-1} and 27.1 kJ mol^{-1}, respectively. The value of D_e for the *trans*-2-azetidinone-H_2O complex can be assumed as a reasonable approximation to the energies of the isolated C=O$\cdots$H hydrogen bond. The isomer denoted as 1c for the formamide-water complex (D_e = 20.4 kJ mol^{-1}, at the same level)

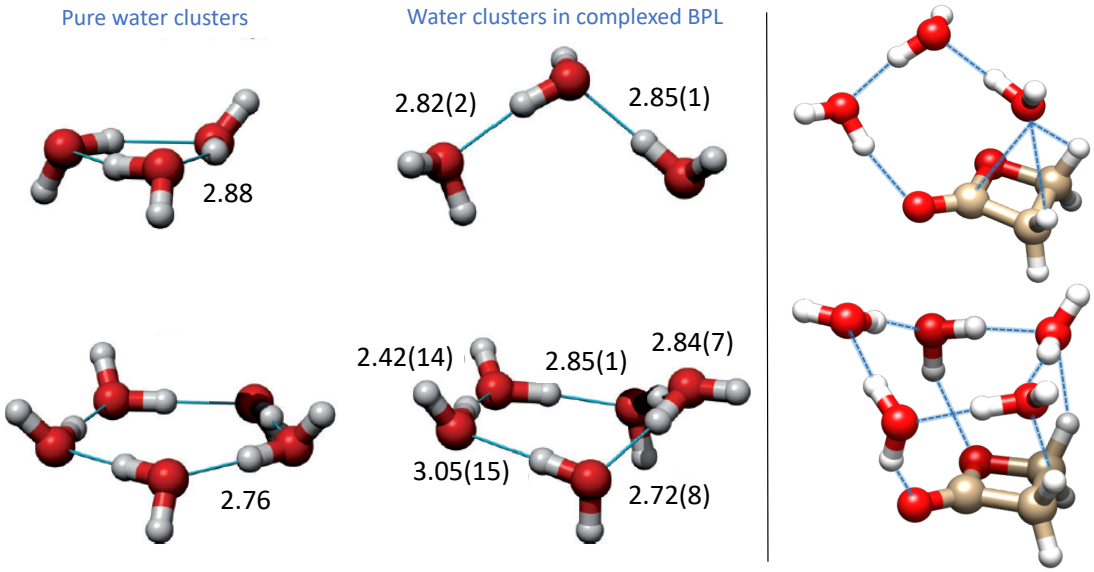


Figure 9. Left panel: Comparison of the O···O distances (in Å) between isolated pure water clusters and in BPL complexes, for $n = 3$ and 5. For the latter, the r_s structural parameters are reported, with BPL molecules being removed to aid visualization. For isolated water clusters, the experimental distances were calculated using the data and expression given in ref [184]. Right panel: visualization of the MP2/6-311++G(d,p) equilibrium structure.

can be taken as a model for the N–H···O hydrogen bond, and the dissociation energy of the water dimer calculated at the same level (19.3 kJ mol^{-1}) as the energy of the isolated O–H···O bond. The sum of these three energies, 66.8 kJ mol^{-1}, is smaller than that calculated for the 2-azetidinone-$(H_2O)_2$ complex of 76.3 kJ mol^{-1}. This increment of the energy of the cyclic sequential set of three hydrogen bonds considered with respect to the sum of the energies of three isolated hydrogen bonds should be attributed to σ-bond cooperativity.

Despite the interesting information obtained from the study presented above [4] and from similar investigations (see, e.g. refs [165, 149, 177, 178, 180, 181]) and the fact that the most favorable water binding sites can be discovered, because of the small number of water molecules involved, the molecular behavior is still far from that of diluted solutions. For this reason, we here also discuss the investigation of β-propiolactone (BPL) complexed with one to five water molecules [3], which can provide insight on the structure of an organic molecule upon the formation of its first solvation shell. The study reported in ref [3] pointed out that microsolvated BPL in a supersonic expansion allowed for disclosing the most stable water-binding sites and, more importantly, the changes caused by complexation to the local water-hydrogen-bonding network. In each complex, the water molecules form a hydrogen-bonded network which interacts at multiple sites with the BPL molecule, the most favorable BPL-water interaction being an $O_{H_2O}H\cdots O_{C=O}$ hydrogen bond.

The structures of the observed complexes reflect a trade-off between maximizing the

number of BPL-water intermolecular interactions and the minimum-energy structures of the pure water clusters. The resulting structures therefore resemble the pure water complexes, but with significant structural perturbations as a result of water-BPL hydrogen-bonding effects. For the molecular complexes observed in supersonic expansion, structural parameters of the heavy-atom framework were obtained by the measurement of the isotopically substituted ^{13}C and ^{18}O and the subsequent determination of the so-called substitution structure (r_s) [185], with the hydrogen-bond orientation derived by measuring the corresponding $(HOD)(H_2O)_{n-1}$ isotopologues.

Focusing on the larger clusters considered, complexation induces measurable structural changes in the $n = 3$, 4, and 5 complexes, n being the number of water molecules. The largest effect on the water-cluster network is observed for $n = 3$, with water molecules preferring a chain arrangement in the complex (see Figure 9). In the case of $n = 4$ system, the difference is mostly related to the orientation of the non-bonded hydrogens for two of the water molecules forming the cluster. For $n = 5$, the largest changes observed are the loss of planarity of the ring formed by water molecules and the fact that the O–H$\cdots$O distances do no longer have the same length in the BPL complex (see Figure 9). The comparison of the experimentally derived O$\cdots$O distances in pure and BPL-complexed water clusters shows that these are shorter in the BPL complexes.

3.3 From microsolvation to condensed phase

Moving from microsolvated systems to condensed phases, the significant problem of extending accurate QC computations to large systems previously addressed is worsened by the need of extensive configurational samplings, which are required to obtain a proper representation of solvent and thermal effects, and thus by the need of calculating accurate structures and properties for a huge number of snapshots [186, 187]. In fact, in most cases, a deficient sampling and/or insufficiently accurate quantum-mechanical (QM) models have a great impact on the quality of the overall results.

To treat localized phenomena, the most viable strategy is thus offered by the so-called multi-scale approaches [188, 189, 190], which focus on the target system (treated at the QM level), while the environment can be progressively blurred from an atomistic MM (MM standing for molecular mechanics) treatment up to a mean field description [191]. Therefore, the matching between the representations of different regions becomes an essential ingredient to be taken into account. For instance, when QM/MM approaches are considered for the investigation of a system (by QM) and its environment (by MM), the inaccurate description of the exchange interaction between the system and the environment by means of empirical additive terms, as well as the lack of polarization in the MM part, can cause a spurious over-polarization of the electron (QM) density, which has a major impact –for example– on solvato-chromic shift predictions [187, 192, 193]. Thus, a number of developments are still needed to extend the range of application for computational simulations.

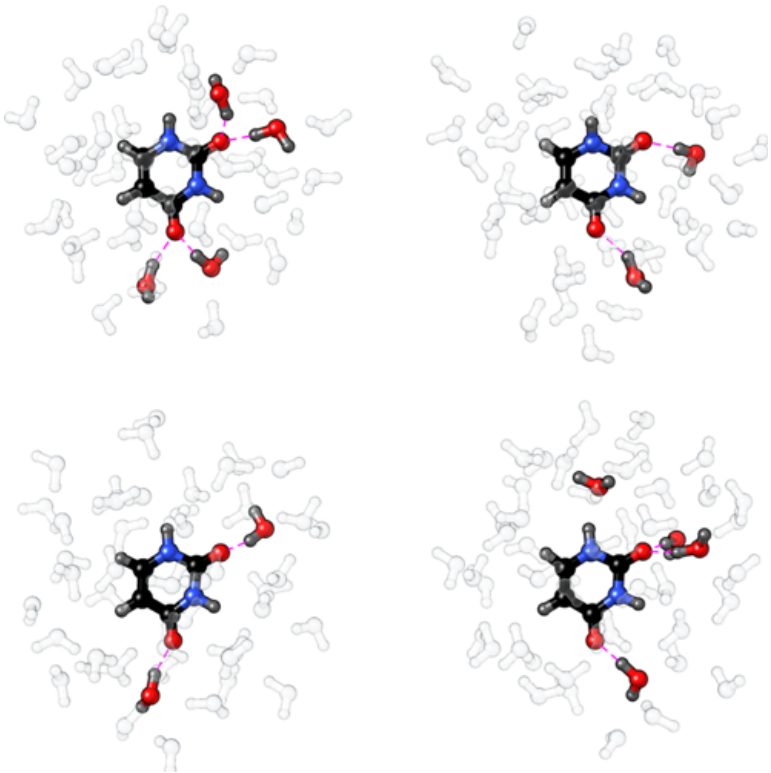
Even if all theoretical issues could be solved, brute-force simulations with state-of-the-art QM models are still unpractical. In such cases, the free energy perturbation (FEP) method [194] opens the way to further extend the limits of current approaches. In short, a molecular dynamics (MD) trajectory is first carried out at an affordable (yet sufficiently reliable) level that allows keeping the computational cost under control. The configurational sampling is then used in conjunction with perturbation theory to obtain the final physical-chemical properties of interest at a high QM (or QM/MM) level. In this approach, the perturbation Hamiltonian is nothing else that the difference between the target high-level (H_{HL}) and reference low-level (H_{LL}) Hamiltonians. For a selected value of the solute's mass-weighted Cartesian coordinates $\mathbf{q}$, the free energy at high level then reads:

$$G_{HL}(\mathbf{q}) = G_{LL}(\mathbf{q}) - \frac{1}{\beta} ln \langle e^{-\beta \Delta V(\mathbf{q};\mathbf{s})} \rangle_{LL} \quad , \tag{12}$$

where $\Delta V = V_{HL} - V_{LL}$ is the corresponding difference in potential energy; $\mathbf{s}$ is used to emphasize the explicit dependence of the potential energy difference on the solvent coordinates and the average is computed on the sampling issuing from equilibrium configurations obtained at the low level of theory. Derivatives of Eq. (12) with respect to the mass-weighted solute's coordinates allow for getting the FEP corrected free energy gradient and Hessian [195]. In the FEP framework we are thus led back to the problem of obtaining accurate intermolecular interactions. After several years of continuing developments, accurate solvent-solvent interactions have been made available for the most widely employed solvents (water being, of course, the paradigmatic example).

Accurate solute-solvent interactions must be, instead, developed on a case-by-case basis and in the framework of either QM/MM or MM/MM approaches. Furthermore, two-body interactions are not always sufficient and the inclusion of three-body contributions involves additional problems concerning both the most effective functional form and the number of reference structures to be fitted.

In the conventional variational multi-scale procedures (hereafter VAR) [189, 190, 191], the unperturbed properties (energies, dipole moments, etc.) of the solute are provided by QM calculations, whereas MD simulations provide a description of the classical (mainly electrostatic) contribution generated by the environment, which modifies the solute Hamiltonian. In the perturbed matrix method (PMM) [196, 197], instead of including environmental effects in the solute Hamiltonian operator, they are treated by means of first-order perturbation of a reference solute Hamiltonian (without any iterative process). The output of the MD-PMM procedure is the trajectory of the eigenvectors and eigenvalues of the perturbed Hamiltonian matrix that can be used to evaluate every (perturbed) QM property of the solute interacting with its environment. It is also possible to combine the VAR and PMM approaches by defining a limited number of reference configurations evaluated by the former method, with fluctuations around each of them described by the latter (much faster) method [198]. The reference configurations can be selected, after a MD run, by means of a suitable clustering of the snapshot and a proper definition of the center of each cluster. It is thus possible to fasten



significantly the use of Eq. (12) because only a few configurations are fully computed at high level of theory, whereas fluctuation effects are estimated by perturbative corrections to an effective one-electron (DFT) Hamiltonian with a negligible additional cost. In this connection, it is also advantageous to employ a solvent model enforcing non-periodic boundary conditions (NPBC) in place of their more conventional periodic counterparts (PBC). The GLOB model [199, 200] is a very refined NPBC model, which has been parametrized for a large panel of solvents [201] and includes an effective confinement potential coupled to a polarizable continuum model (PCM) [202] for describing the missing bulk effects.

Although the integrated VAR/PMM/GLOB (hereafter VPG) model has not been yet employed to compute structures in solution, its application to several kinds of spectroscopies is very promising [198, 203]. As an example, Figure 10 shows the reference structures describing an aqueous solution of uracil, while Figure 11 compares the corresponding UV-vis spectrum computed by the above integrated procedure with that obtained using the PCM model and a full variational approach. The limits of a continuum description of hydrogen-bonding solvents are well evident, whereas the remarkable agreement between the standard VAR procedure employing 500 different snapshots and the VPG approach perturbing just 4 reference structures (with a reduction of two orders of magnitude in the computer time) points out the reliability and robustness of the latter approach and suggests the interest of its use also for structural investigations. Analogously, a similar model can be employed for investigating the effect of a solid environment in the absence of too strong delocalization.

3.4 Solid State and Catalysis

Although the main focus of this review is on non-covalent interactions in small complexes in the gas-phase as a prerequisite for a deeper understanding of more complex situations (thus resulting from the interplay of different effects), some comments for solid state and catalysis are however deserved. As a matter of fact, the detailed analysis of the interactions between molecules and ions in crystals plays an increasingly important role in modern solid-state chemistry, and in particular crystal engineering, where the derivation of predictive structure-property relationships is a crucial step toward a genuine "engineering" of crystals. This was indeed anticipated some years ago by Desiraju, who described the crystal engineering of molecular solids as the "understanding of intermolecular interactions in the context of crystal packing and the utilization of such understanding in the design of new solids with desired physical and chemical properties" [204]. Utilization and design require understanding as an essential prerequisite, especially in the context of crystal packing. Recent years have seen an explosive increase of the role of different non-canonical interactions such as triel, tetrel/carbon, pnicogen/pnictogen, chalcogen, halogen and aerogen ?bonds? also in the fields of material science and catalysis.

While the methodologies described in the previous sections can provide reliable

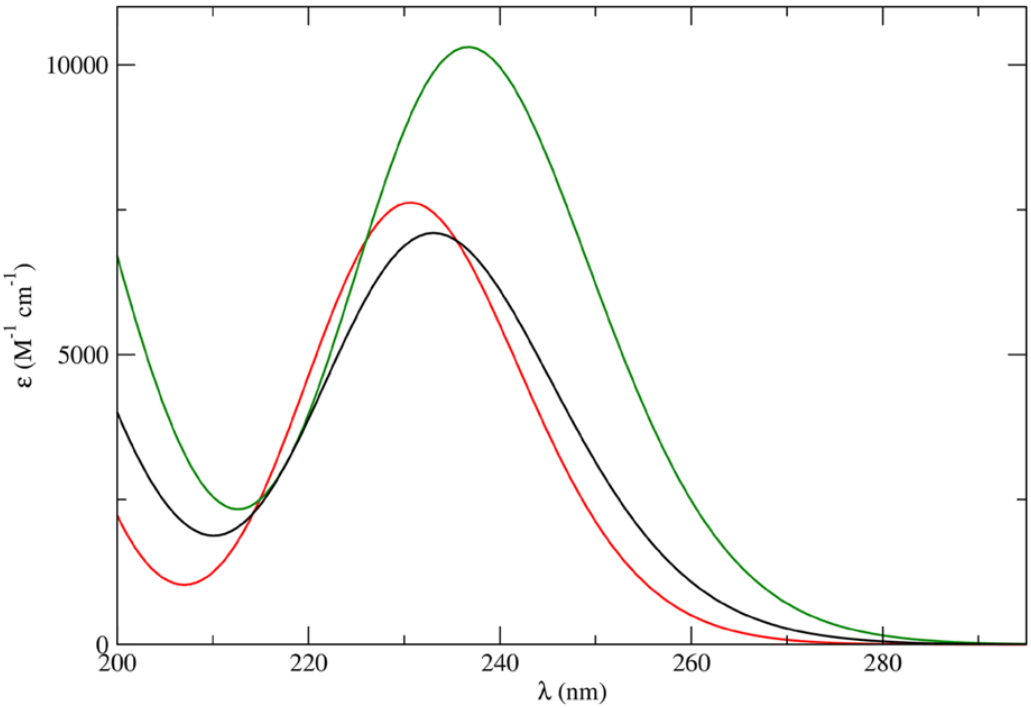


Figure 11. UV spectrum of uracil in water obtained using the PCM (green curve), VAR (black curve), and VPG (red curve) approaches.

structural and energetic information on different kinds of intermolecular interactions, when dealing with materials and/or heterogeneous catalysts, these must be extended to larger clusters in order to investigate possible non-additive effects. The unfavorable scaling of state-of-the-art quantum-chemical methods with the system dimension requires additional considerations. In this connection, at least two different routes can be envisaged. A first possibility is offered by methods enforcing periodic boundary conditions, especially those based on localized functions (e.g. the CRYSTAL code [205, 206]). These have been extended to include both advanced DFT approaches and, more recently, powerful post-Hartree-Fock techniques, thus paving the route toward very reliable yet effective results. However, the development of general tools allowing systematic studies is still in progress. Therefore, at this stage, it is not possible to provide a review on the performances of these approaches. A second way is provided by the extension of composite methods to multilayer models in which the sophistication (hence the cost) of the approach decreases with the distance from the local environment of the 'active center'. This strategy resembles that discussed in the previous section in connection with processes occurring in solution, but here the situation is comparatively simpler because the problem of statistical averaging of several nearly isoenergetic configurations is much less important. On the other hand, the dimensions of the cluster used to model the solid surface can become quite large and the partition into different regions (to be treated at different levels) less evident, not to speak of the presence of larger boundary effects whenever covalent bonds are broken. These methods are

particularly appealing when the interest is not on the overall properties of the solid, but –rather– on the chemisorption and reactivity of small molecules for processes broadly related to catalysis. Since a survey of the rich literature on this subject is outside the scope of the present review, in the following, we will only provide the reader with a flavor of the topic by means of a specific example taken from a recent work in the field of heterogeneous catalysis.

The composite model recently proposed for providing chemisorption energies within chemical accuracy [207] strongly resembles the ChS approach discussed in the preceding sections. The approach starts from a DFT geometry optimization and harmonic frequency evaluation for a proper "supercell", thereby enforcing periodic boundary conditions (PBC) in conjunction with a gradient corrected functional and a suitable basis set. Subsequently, a cluster model is extracted from the periodic computation and terminated with hydrogen atoms (in the case of covalent solids like zeolites or Metal-Organic Frameworks, MOFs) or with point charges or full-ion effective potentials (in the case of ionic crystals). The geometry of this cluster model is then optimized at both the MP2 level in conjunction with a triple-zeta basis set and using DFT, with the difference between the two results being used for correcting the original DFT/PBC values. To examine the effect of convergence with the cluster size, the energy is evaluated for a larger cluster. If the hybrid scheme worked properly and the smaller cluster model was chosen correctly, the results for the two clusters plus the DFT/PBC description of the environment should provide coincident results. As a final step, at the equilibrium structure derived, CCSD(T) single-point calculations (using a triple-zeta quality basis set) of energies and, possibly, of properties are carried out for the smallest cluster models and corrected for the extrapolation to the CBS limit at the MP2 level. In view of providing an example for this approach, some results taken from ref. [207], are collected in Table 5. From these results, it is quite apparent that, although further developments are surely needed, chemical accuracy is becoming a reasonable target also in the field of heterogeneous catalysis.

4 Concluding remarks

In this review, we hope to have shown that state-of-the-art QC computations combined with rotational spectroscopy allow for fully characterizing the intermolecular interactions taking place in molecular complexes from both structural and energetic points of view. By means of selected examples, it has been shown that it is possible to accurately investigate the geometrical modifications occurring upon formation of molecular adducts as well as to understand the nature of the interaction ruling their stability. Particular emphasis has been put on hydrogen bonds, which are ubiquitous in nature, and –e.g.– govern molecular conformations and thus biochemical functionality. Indeed, the structures of proteins and other biological molecules are determined by a delicate balance between several molecular interactions. Starting from model systems interacting with only one water molecule, we have then addressed microsolvated clusters, and we finally

Table 5. Comparison of chemisorption energies (in kJ mol^{-1}) obtained using DFT/PBC and composite methods with experiment for some representative systems.[a]

System	DFT/PBC	composite scheme	experiment	cc-exp[b]
CO/Mg(001)	-22.1	-21.2	-20.6	-0.6
CH4/Mg(001)	-14.8	-14.0	-15.0	1.0
H2/MOF-5	-9.6	-8.0	-8.1	0.1
CO-Mg-MOF-74	-41.5	-43.3	-43.8	0.5
CO_2/Mg-MOF74	-41.5	-48.8	-46.2	-2.6
CH_4/H-CHA	-34.7	-25.3	-27.2	1.9
C_2H_6/H-CHA	-45.8	-36.2	-33.5	2.7
C_3H_8/H-CHA	-57.3	-46.7	-43.8	2.9

[a] Results taken from ref. [207].

[b] Energy difference between composite scheme results and experiment.

extended the problem to condensed phase. Attention has also been paid to emerging classes of non-covalent interactions, such as the pnicogen and chalcogen bonds.

The semi-experimental approach, which requires a joint effort by quantum chemistry and rotational spectroscopy, has been extended to accurately determine the intermolecular parameters and to investigate the intramolecular modifications. On the other hand, a QC protocol to evaluate interaction energies with great precision has been established and benchmarked. Both of them have been found suitable for further extending the size of the systems amenable to a rigorous and accurate characterization.

Thanks to all these developments the long-standing goal of fulfilling the chemical accuracy (i.e. uncertainties of about 4 kJ mol^{-1} for energetic quantities and 0.005 bohr or radiants for structural parameters) and interpretability has been reached for small- and medium-sized non-covalent complexes in the gas phase. Reaching the same goal for condensed phases (solutions and solids) is considerably more complex, but promising approaches have been developing, and a flavor is provided in the last part of this review.

acknowledgement

Support from the Italian MIUR (PRIN 2015, Grant Number 2015F59J3R; PRIN 2017, Grant Number 2017A4XRCA) is acknowledged. This work has also been supported by the University of Bologna (RFO funds). The authors also thank Prof. Jens-Uwe Grabow for useful discussions.

References

[1] de Vries M S and Hobza P 2007 *Ann. Rev. Phys. Chem.* **58** 585–612
[2] Roscioli J R and Pratt D W 2003 *Proc. Nat. Acad. Sci. USA* **100** 13752–13754
[3] Perez C, Neill J L, Muckle M T, Zaleski D P, Peña I, Lopez J C, Alonso J L and Pate B H 2015 *Angew. Chem. Int. Ed.* **54** 979–982
[4] López J C, Sánchez R, Blanco S and Alonso J L 2015 *Phys. Chem. Chem. Phys.* **17**(3) 2054–2066
[5] Wang W, Donini O, Reyes C M and Kollman P A 2001 *Annu. Rev. Biophys. Biomol. Struct.* **30** 211–243
[6] Riley K E and Hobza P 2011 *WIREs Comput. Mol. Sci.* **1** 3–17
[7] Tang W, Johnston S, Iggo J A, Berry N G, Phelan M, Lian L, Bacsa J and Xiao J 2013 *Angew. Chem. Int. Ed.* **52** 1668–1672
[8] Wheeler S E, Seguin T J, Guan Y and Doney A C 2016 *Acc. Chem. Res.* **49** 1061–1069
[9] Vallavoju N and Sivaguru J 2014 *Chem. Soc. Rev.* **43**(12) 4084–4101
[10] Guo X, Liao Q, Manley E F, Wu Z, Wang Y, Wang W, Yang T, Shin Y E, Cheng X, Liang Y, Chen L X, Baeg K J, Marks T J and Guo X 2016 *Chem. Mater.* **28** 2449–2460
[11] Georgakilas V, Tiwari J N, Kemp K C, Perman J A, Bourlinos A B, Kim K S and Zboril R 2016 *Chem. Rev.* **116** 5464–5519
[12] Li W, Spada L, Tasinato N, Rampino S, Evangelisti L, Gualandi A, Cozzi P G, Melandri S, Barone V and Puzzarini C 2018 *Angew. Chem. Int. Ed.* **57** 13853–13857
[13] Obenchain D A, Spada L, Alessandrini S, Rampino S, Herbers S, Tasinato N, Mendolicchio M, Kraus P, Gauss J, Puzzarini C and et al 2018 *Angew. Chem. Int. Ed.* **57** 15822–15826
[14] Wang J, Spada L, Chen J, Gao S, Alessandrini S, Feng G, Puzzarini C, Gou Q, Grabow J U and Barone V 2019 *Angew. Chem. Int. Ed.* **58** 13935–13941
[15] Li W, Li M, Jin Y, Gou Q, Grabow J U and Feng G 2019 *J. Chem. Phys.* **151** 164307
[16] Kraus P, Obenchain D A and Frank I 2018 *J. Phys. Chem. A* **122** 1077–1087
[17] Scheiner S 2013 *Acc. Chem. Res.* **46** 280–288
[18] Bauzá A, Ramis R and Frontera A 2014 *J. Phys. Chem. A* **118** 2827–2834
[19] Mahmudov K T, Kopylovich M N, Guedes da Silva M F C and Pombeiro A J L 2017 *Dalton Trans.* **46**(31) 10121–10138
[20] Vogel L, Wonner P and Huber S M 2019 *Angew. Chem. Int. Ed.* **58** 1880–1891
[21] Mak C H 2016 *J. Phys. Chem. B* **120** 6010–6020
[22] Hobza P and Sponer J 2002 *J. Am. Chem. Soc.* **124** 11802–11808
[23] Sinnokrot M O and Sherrill C D 2004 *J. Am. Chem. Soc.* **126** 7690–7697
[24] Puzzarini C and Biczysko M 2015 *J. Phys. Chem. A* **119** 5386–5395
[25] Juanes M, Saragi R T, Caminati W and Lesarri A 2019 *Chem. Eur. J.* **25**
[26] Blanco S, Pinacho P and López J C 2016 *Angew. Chem. Int. Ed.* **55** 9331–9335
[27] Nochebuena J, Cuautli C and Ireta J 2017 *Phys. Chem. Chem. Phys.* **19**(23) 15256–15263
[28] Rossi M, Fang W and Michaelides A 2015 *J. Phys. Chem. Lett.* **6** 4233–4238
[29] Alonso J L and López J C 2015 Microwave spectroscopy of biomolecular building blocks (*Top. Curr. Chem.* vol 364) (Springer Berlin Heidelberg) pp 335–401
[30] Becucci M and Melandri S 2016 *Chem. Rev.* **116** 5014–5037
[31] Benz S, López-Andarias J, Mareda J, Sakai N and Matile S 2017 *Angew. Chem. Int. Ed.* **56** 812–815
[32] Fanfrlík J, Prada A, Padelkova Z, Pecina A, Machacek J, Lepsik M, Holub J, Ruzicka A, Hnyk D and Hobza P 2014 *Angew. Chem. Int. Ed.* **53** 10139–10142
[33] Chen L, Xiang J, Zhao Y and Yan Q 2018 *J. Am. Chem. Soc.* **140** 7079–7082
[34] Gleiter R, Haberhauer G, Werz D B, Rominger F and Bleiholder C 2018 *Chem. Rev.* **118** 2010–2041
[35] Gordy W and Cook R L 1984 *Microwave Molecular Spectra, 3rd Edition* (New York: Wiley)
[36] Allen W D, East A L L and Császár A G 1993 (Kluwer, Dordrecht) p 343–373

[37] Puzzarini C, Stanton J F and Gauss J 2010 *Int. Rev. Phys. Chem.* **29** 273–367
[38] Barone V (ed) 2011 *Computational Strategies for Spectroscopy: from Small Molecules to Nano Systems* (John Wiley & Sons, Inc.)
[39] Puzzarini C 2013 *Phys. Chem. Chem. Phys.* **15**(18) 6595–6607
[40] Barone V 2016 *WIREs Comput. Mol. Sci.* **6** 86–110
[41] Puzzarini C, Bloino J, Tasinato N and Barone V 2019 *Chem. Rev.* **119** 8131–8191
[42] Aliev M R and Watson J K G 1985 Chapter 1 - higher-order effects in the vibration–rotation spectra of semirigid molecules *Molecular Spectroscopy: Modern Research* vol 3 ed Rao K N (Academic Press) pp 1–67 ISBN 978-0-12-580643-5
[43] Watson J K G 1968 *J. Chem. Phys.* **48** 4517–4524
[44] Caminati W and Grabow J U 2009 *Frontiers of Molecular Spectroscopy* (Elsevier) chap Microwave Spectroscopy: Molecular Systems
[45] Pate B H, Evangelisti L, Caminati W, Xu Y, Thomas J, Patterson D, Pérez C and Schnell M 2018 *Chiral Analysis: Advances in Spectroscopy, Chromatography, and Emerging Methods, 2nd ed.* (Elsevier) chap Quantitative chiral analysis by molecular rotational spectroscopy
[46] Grabow J U 2011 *Handbook of High Resolution Spectroscopy* (New York: Wiley) chap Fourier Transform Microwave Spectroscopy Measurement and Instrumentation, pp 723–800
[47] Balle T J and Flygare W H 1981 *Rev. Sci. Instrum.* **52** 33–45
[48] Grabow J, Stahl W and Dreizler H 1996 *Rev. Sci. Instrum.* **67** 4072–4084
[49] Brown G G, Dian B C, Douglass K O, Geyer S M, Shipman S T and Pate B H 2008 *Rev. Sci. Instrum.* **79** 053103
[50] Pérez C, Lobsiger S, Seifert N A, Zaleski D P, Temelso B, Shields G C, Kisiel Z and Pate B H 2013 *Chem. Phys. Lett.* **571** 1–15 ISSN 0009-2614
[51] Legon A C and Rego C A 1990 *J. Chem. Soc., Faraday Trans.* **86**(11) 1915–1921
[52] Emilsson T, Klots T D, Ruoff R S and Gutowsky H S 1990 *J. Chem. Phys.* **93** 6971–6976
[53] Alonso J L, Cocinero E J, Lesarri A, Sanz M E and López J C 2006 *Angew. Che. Inte. Ed.* **45** 3471–3474
[54] Vaquero V, Sanz M E, Peña I, Mata S, Cabezas C, López J C and Alonso J L 2014 *J. Phys. Chem. A* **118** 2584–2590
[55] Suma K, Sumiyoshi Y and Endo Y 2006 *Science* **311** 1278–1281
[56] Ohshima Y, Sato K, Sumiyoshi Y and Endo Y 2005 *J. Am. Chem. Soc.* **127** 1108–1109
[57] Piccardo M, Penocchio E, Puzzarini C, Biczysko M and Barone V 2015 *J. Phys. Chem. A* **119** 2058–2082
[58] Piccardo M, Bloino J and Barone V 2015 *Int. J. Quantum Chem.* **115** 948–982
[59] Puzzarini C and Barone V 2011 *Phys. Chem. Chem. Phys.* **13** 7189–7197
[60] Shavitt I 1977 *The method of configuration interaction* (Plenum Press, New York)
[61] Shavitt I and Bartlett R J 2009 *Many-body methods in chemistry and physics: MBPT and coupled-cluster theory* (Cambridge University Press)
[62] Raghavachari K, Trucks G W, Pople J A and Head-Gordon M 1989 *Chem. Phys. Lett.* **157** 479–483 ISSN 0009-2614
[63] Noga J and Bartlett R J 1987 *J. Chem. Phys.* **86** 7041–7050
[64] Scuseria G E and Schaefer III H F 1988 *Chem. Phys. Lett.* **152** 382–386
[65] Kállay M and Surján P R 2001 *J. Chem. Phys.* **115** 2945–2954
[66] Bomble Y J, Stanton J F, Kállay M and Gauss J 2005 *J. Chem. Phys.* **123** 054101
[67] Kállay M and Gauss J 2005 *J. Chem. Phys.* **123** 214105
[68] Heckert M, Kállay M, Tew D P, Klopper W and Gauss J 2006 *J. Chem. Phys.* **125** 044108
[69] Heckert M, Kállay M and Gauss J 2005 *Mol. Phys.* **103** 2109–2115
[70] Allen W D, East A L L and Csaszaŕ A G 1993 *Structures and Conformations of Non-Rigid Molecules* (Dordrecht: Kluwer) p 343
[71] Császár A G, Allen W D and Schaefer III H F 1998 *J. Chem. Phys.* **108** 9751–9764
[72] Tajti A, Szalay P G, Császár A G, Kállay M, Gauss J, Valeev E F, Flowers B A, Vázquez J and

Stanton J F 2004 *J. Chem. Phys.* **121** 11599–11613
[73] Harding M E, Vázquez J, Ruscic B, Wilson A K, Gauss J and Stanton J F 2008 *J. Chem. Phys.* **128** 114111
[74] Puzzarini C, Biczysko M, Barone V, Largo L, Peña I, Cabezas C and Alonso J L 2014 *J. Phys. Chem. Lett.* **5** 534–540
[75] Ganyecz Á, Kállay M and Csontos J 2017 *J. Chem. Theory Comput.* **13** 4193–4204
[76] Boese A D, Oren M, Atasoylu O, Martin J M L, Kállay M and Gauss J 2004 *J. Chem. Phys.* **120** 4129–4141
[77] Karton A, Rabinovich E, Martin J M L and Ruscic B 2006 *J. Chem. Phys.* **125** 144108
[78] Dunning Jr T H 1989 *J. Chem. Phys.* **90** 1007–1023
[79] Widmark P O, Malmqvist P A and Roos B 1990 *Theoret. Chim. Acta* **77** 291–306
[80] Widmark P O, Persson B and Roos B 1991 *Theoret. Chim. Acta* **79** 419–432
[81] Jensen F 2013 *WIREs Comput. Mol. Sci.* **3** 273–295
[82] Feller D 1993 *J. Chem. Phys.* **98** 7059–7071
[83] Helgaker T, Klopper W, Koch H and Noga J 1997 *J. Chem. Phys.* **106** 9639
[84] Feller D and Peterson K A 1999 *J. Chem. Phys.* **110** 8384–8396
[85] Dunning T H, Peterson K A and Wilson A K 2001 *J. Chem. Phys.* **114** 9244–9253
[86] Kendall R, Dunning Jr T and Harrison R 1992 *J. Chem. Phys.* **96** 6769–6806
[87] Papajak E, Leverentz H R, Zheng J and Truhlar D G 2009 *J. Chem. Theory Comput.* **5** 1197–1202
[88] Papajak E, Zheng J, Xu X, Leverentz H R and Truhlar D G 2011 *J. Chem. Theory Comput.* **7** 3027–3034
[89] Woon D E and Dunning Jr T H 1995 *J. Chem. Phys.* **103** 4572–4585
[90] Peterson K A and Dunning Jr T H 2002 *J. Chem. Phys.* **117** 10548–10560
[91] Stanton J F, Gauss J, Harding M E and Szalay P G 2016 CFOUR. a quantum chemical program package with contributions from A. A. Auer, R. J. Bartlett et al. and the integral packages MOLECULE (J. Almlöf and P. R. Taylor), PROPS (P. R. Taylor), ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de
[92] Thorwirth S, Mück L A, Gauss J, Tamassia F, Lattanzi V and McCarthy M C 2011 *J. Phys. Chem. Lett.* **2** 1228–1231
[93] Barone V, Biczysko M, Bloino J and Puzzarini C 2013 *Phys. Chem. Chem. Phys.* **15** 1358–1363
[94] Barone V, Biczysko M, Bloino J and Puzzarini C 2014 *J. Chem. Phys.* **141** 034107
[95] Barone V, Biczysko M, Bloino J, Cimino P, Penocchio E and Puzzarini C 2015 *J. Chem. Theory Comput.* **11** 4342–4363
[96] Barone V, Biczysko M, Bloino J and Puzzarini C 2013 *Phys. Chem. Chem. Phys.* **15**(25) 10094–10111
[97] Alessandrini S and Puzzarini C 2016 *J. Phys. Chem. A* **120** 5257–5263
[98] Martin J M and Taylor P R 1996 *Chem. Phys. Lett.* **248** 336 – 344 ISSN 0009-2614
[99] Demaison J, Margulès L and Boggs J E 2003 *Struct. Chem.* **14** 159–174
[100] Puzzarini C 2009 *J. Phys. Chem. A* **113** 14530–14535
[101] Møller C and Plesset M S 1934 *Phys. Rev.* **46** 618–622
[102] Puzzarini C, Biczysko M, Barone V, Peña I, Cabezas C and Alonso J L 2013 *Phys. Chem. Chem. Phys.* **15**(39) 16965–16975
[103] Spada L, Tasinato N, Vazart F, Barone V, Caminati W and Puzzarini C 2017 *Chem. Eur. J.* **23** 4876–4883 ISSN 1521-3765
[104] Grimme S 2006 *J. Chem. Phys.* **124** 034108
[105] Grimme S 2011 *WIREs Comput. Mol. Sci.* **1** 211–228 ISSN 1759-0884
[106] Grimme S, Antony J, Ehrlich S and Krieg H 2010 *J. Chem. Phys.* **132** 154104
[107] Grimme S, Ehrlich S and Goerigk L 2011 *J. Comput. Chem.* **32** 1456–1465 ISSN 1096-987X
[108] Puzzarini C and Barone V 2018 *Acc. Chem. Res.* **51** 548–556
[109] Puzzarini C, Tasinato N, Bloino J, Spada L and Barone V 2019 *Phys. Chem. Chem. Phys.* **21**

3431–3439
[110] Boys S and Bernardi F 1970 *Mol. Phys.* **19** 553–566
[111] Bomble Y J, Vázquez J, Kállay M, Michauk C, Szalay P G, Császár A G, Gauss J and Stanton J F 2006 *J. Chem. Phys.* **125** 064108
[112] Schuurman M S, Muir S R, Allen W D and Schaefer H F 2004 *J. Chem. Phys.* **120** 11586–11599
[113] Dixon D A, Feller D and Peterson K A 2012 *Chapter One - A Practical Guide to Reliable First Principles Computational Thermochemistry Predictions Across the Periodic Table* vol 8 (Elsevier) pp 1 – 28
[114] Puzzarini C 2011 *Phys. Chem. Chem. Phys.* **13** 21319–21327
[115] Puzzarini C, Penocchio E, Biczysko M and Barone V 2014 *J. Phys. Chem. A* **118** 6648–6656
[116] Puzzarini C 2015 *J. Phys. Chem. A* **119** 11614–11622
[117] Kodrycka M and Patkowski K 2019 *J. Chem. Phys.* **151** 070901
[118] Schurkus H F, Luenser A and Ochsenfeld C 2017 *J. Chem. Phys.* **146** 211106
[119] Nagy P R and Kállay M 2019 *J. Chem. Theory Comput.* **in press** DOI:10.1021/acs.jctc.9b00511
[120] Glendening E D, Landis C R and Weinhold F 2012 *WIREs Comput. Mol. Sci.* **2** 1–42
[121] Jeziorski B, Moszynski R and Szalewicz K 1994 *Chem. Rev.* **94** 1887–1930
[122] Parker T M, Burns L A, Parrish R M, Ryno A G and Sherrill C D 2014 *J. Chem. Phys.* **140** 094106
[123] Weinhold F and Landis C R 2012 (Wiley, New Jersey)
[124] Reed A E, Curtiss L A and Weinhold F 1988 *Chem. Rev.* **88** 899–926
[125] Glendening E D 2005 *J. Phys. Chem. A* **109** 11936–11940
[126] Karalti O, Alfé D, Gillan M J and Jordan K D 2012 *Phys. Chem. Chem. Phys.* **14** 7846–7853
[127] Tamijani A A, Salam A and de Lara-Castells M P 2016 *J. Phys. Chem. C* **120** 18126–18139
[128] Patkowski K *WIREs Comput. Mol. Sci.* **in press; DOI: 10.1002/wcms.1452** e1452
[129] Parrish R M, Hohenstein E G and Sherrill C D 2013 *J. Chem. Phys.* **139** 174102
[130] Stone A J and Misquitta A J 2009 *Chem. Phys. Lett.* **473** 201 – 205 ISSN 0009-2614
[131] Řezáč J, Dubecký M, Jurečka P and Hobza P 2015 *Phys. Chem. Chem. Phys.* **17**(29) 19268–19277
[132] Spada L, Tasinato N, Bosi G, Vazart F, Barone V and Puzzarini C 2017 *J. Mol. Spectrosc.* **337** 90–95
[133] Řezáč J and Hobza P 2013 *J. Chem. Theory Comput.* **9** 2151–2155
[134] De Jong W A, Harrison R J and Dixon D A 2001 *J. Chem. Phys.* **114** 48–53
[135] Tsuzuki S, Uchimaru T, Matsumura K, Mikami M and Tanabe K 2000 *Chem. Phys. Lett.* **319** 547–554
[136] Tkatchenko A, DiStasio R A, Head-Gordon M and Scheffler M 2009 *J. Chem. Phys.* **131** 094106
[137] Tan S Y S, Wylie L, Begic I, Tran D and Izgorodina E I 2017 *Phys. Chem. Chem. Phys.* **19**(42) 28936–28942 URL `http://dx.doi.org/10.1039/C7CP04391K`
[138] Řezáč J and Hobza P 2016 *Chem. Rev.* **116** 5038–5071
[139] Řezáč J, Riley K E and Hobza P 2012 *J. Chem. Theory Comput.* **8** 4285–4292
[140] Oliveira V and Kraka E 2017 *J. Phys. Chem. A* **121** 9544–9556
[141] Kesharwani M K, Manna D, Sylvetsky N and Martin J M L 2018 *J. Phys. Chem. A* **122** 2184–2197
[142] Sylvetsky N, Kesharwani M K and Martin J M L 2017 *J. Chem. Phys.* **147** 134106
[143] Pulay P, Meyer W and Boggs J E 1978 *J. Chem. Phys.* **68** 5077–5085
[144] Demaison J, Boggs J E and Császár A G (eds) 2011 *Equilibrium Molecular Structures: From Spectroscopy to Quantum Chemistry* (Boca Raton, FL, US: CRC Press, Taylor & Francis Group)
[145] Mendolicchio M, Penocchio E, Licari D, Tasinato N and Barone V 2017 *J. Chem. Theory Comp.* **13** 3060–3075
[146] Penocchio E, Piccardo M and Barone V 2015 *J. Chem. Theory Comput.* **11** 4689–4707
[147] Penocchio E, Mendolicchio M, Tasinato N and Barone V 2016 *Can. J. Chem.* **94** 1065–1076
[148] For information and download, see https://smart.sns.it/molecules/
[149] Blanco S, López J C, Lesarri A and Alonso J L 2006 *J. Am. Chem. Soc.* **128** 12111–12121

[150] Alessandrini S, Gauss J and Puzzarini C 2018 *J. Chem. Theory Comput.* **14** 5360–5371
[151] Senent M L, Puzzarini C, Dominguez-Gomez R, Carvajal M and Hochlaf M 2014 *J. Chem. Phys.* **140** 124302
[152] Demaison J, Margulés L and Rudolph H 2010 *J. Mol. Struct.* **978** 229–233
[153] Zwier T S 1996 *Ann. Rev. Phys. Chem.* **47** 205–241
[154] Bernhard D, Fatima M, Poblotzki A, Steber A L, Perez C, Suhm M A, Schnell M and Gerhards M 2019 *Phys. Chem. Chem. Phys.* **21**(29) 16032–16046
[155] Watt M M, Collins M S and Johnson D W 2013 *Acc. Chem. Res.* **46** 955–966
[156] Zahn S, Frank R, Hey-Hawkins E and Kirchner B 2011 *Chem. Eur. J.* **17** 6034–6038
[157] Scheiner S 2011 *J. Chem. Phys.* **134** 094315
[158] Bauzá A, Mooibroek T J and Frontera A 2015 *Chem. Commun.* **51**(8) 1491–1493
[159] Beno B R, Yeung K S, Bartberger M D, Pennington L D and Meanwell N A 2015 *J. Med. Chem.* **58** 4383–4438
[160] Pascoe D J, Ling K B and Cockroft S L 2017 *J. Am. Chem. Soc.* **139** 15160–15167
[161] Sundaralingam M and Sekharudu Y 1989 *Science* **244** 1333–1337 ISSN 0036-8075
[162] Levy Y and Onuchic J N 2006 *Ann. Rev. Biophys. Biomol. Struct.* **35** 389–415
[163] Mallamace F, Corsaro C, Mallamace D, Baglioni P, Stanley H E and Chen S H 2011 *J. Phys. Chem. B* **115** 14280–14294
[164] Wyttenbach T, Paizs B, Barran P, Breci L, Liu D, Suhai S, Wysocki V H and Bowers M T 2003 *J. Am. Chem. Soc.* **125** 13768–13775
[165] Caminati W, López J C, Blanco S, Mata S and Alonso J L 2010 *Phys. Chem. Chem. Phys.* **12**(35) 10230–10234
[166] Levy D H 1980 *Annu. Rev. Phys. Chem.* **31** 197–225
[167] Pérez C, Zaleski D P, Seifert N A, Temelso B, Shields G C, Kisiel Z and Pate B H 2014 *Angew. Chem. Int. Ed.* **53** 14368–14372
[168] Auer B, Kumar R, Schmidt J R and Skinner J L 2007 *Proc. Nat. Acad. Sci. USA* **104** 14215–14220
[169] Stokely K, Mazza M G, Stanley H E and Franzese G 2010 *Proc. Nat. Acad. Sci. USA* **107** 1301–1306
[170] Kuo J L and Klein M L 2004 *J. Chem. Phys.* **120** 4690–4695
[171] López J C, Alonso J L, Peña I and Vaquero V 2010 *Phys. Chem. Chem. Phys.* **12**(42) 14128–14134
[172] Bartlett G, Choudhary A, Raines R and Woolfson D N
[173] Horowitz S and Trievel R C 2012 *J. Biol. Chem.* **287** 41576–41582
[174] Gruenloh C J, Carney J R, Arrington C A, Zwier T S, Fredericks S Y and Jordan K D 1997 *Science* **276** 1678–1681
[175] Janzen C, Spangenberg D, Roth W and Kleinermanns K 1999 *J. Chem. Phys.* **110** 9898–9907
[176] Zwier T S 1996 *Annual Review of Physical Chemistry* **47** 205–241
[177] Mata S, Cortijo V, Caminati W, Alonso J L, Sanz M E, López J C and Blanco S 2010 *J. Phys. Chem. A* **114** 11393–11398
[178] Alonso J L, Peña I, Eugenia Sanz M, Vaquero V, Mata S, Cabezas C and López J C 2013 *Chem. Commun.* **49**(33) 3443–3445
[179] Pinacho P, Krin A, Pérez C, Zinn S, López J C, Blanco S and Schnell M 2018 *Phys. Chem. Chem. Phys.* **20**(23) 15635–15640
[180] Gruet S, Pérez C, Steber A L and Schnell M 2018 *Phys. Chem. Chem. Phys.* **20**(8) 5545–5552
[181] Steber A L, Pérez C, Temelso B, Shields G C, Rijs A M, Pate B H, Kisiel Z and Schnell M 2017 *J. Phys. Chem. Lett.* **8** 5744–5750
[182] Jeffrey G A 1997 *Introduction to Hydrogen Bonding* (Oxford University Press, Oxford)
[183] Kisiel Z, Pietrewicz B A, Desyatnyk O, Pszczolkowski L, Struniewicz I and Sadlej J 2003 *J. Chem. Phys.* **119** 5907–5917
[184] Liu K, Brown M G, Cruzan J D and Saykally R J 1996 *Science* **271** 62–64
[185] Kraitchman J 1953 *Am. J. Phys.* **21** 17
[186] Morzan U N, Alonso de Armiño D J, Foglia N O, Ramírez F, González Lebrero M C, Scherlis

D A and Estrin D A 2018 *Chem. Rev.* **118** 4071–4113
[187] Pagliai M, Mancini G, Carnimeo I, De Mitri N and Barone V 2017 *J. Comput. Chem.* **38** 319–335
[188] Levitt M 2014 *Angew. Chem. Int. Ed.* **53** 10006–10018
[189] Karplus M 2014 *Angew. Chem. Int. Ed.* **53** 9992–10005
[190] Warshel A 2014 *Angew. Chem. Int. Ed.* **53** 10020–10031
[191] Chung L W, Sameera W M C, Ramozzi R, Page A J, Hatanaka M, Petrova G P, Harris T V, Li X, Ke Z, Liu F and et al 2015 *Chem. Rev.* **115** 5678–5796
[192] Cisneros G A, Karttunen M, Ren P and Sagui C 2014 *Chem. Rev.* **114** 779–814
[193] Carnimeo I, Cappelli C and Barone V 2015 *J. Comput. Chem.* **36** 2271–2290
[194] Chipot C and Pohonile A 2007 *Free energy calculations* vol 86 (Springer, New York)
[195] Bistafa C, Kitamura Y, Martins-Costa M T C, Nagaoka M and Ruiz-López M F 2019 *J. Chem. Theory Comput.* **15** 4615–4622
[196] Zanetti-Polzi L, Del Galdo S, Daidone I, D'Abramo M, Barone V, Aschi M and Amadei A 2018 *Phys. Chem. Chem. Phys.* **20**(37) 24369–24378
[197] Carrillo-Parramon O, Del Galdo S, Aschi M, Mancini G, Amadei A and Barone V 2017 *J. Chem. Theory Comput.* **13** 5506–5514
[198] Del Galdo S, Chandramouli B, Mancini G and Barone V 2019 *J. Chem. Theory Comput.* **15** 3170–3184
[199] Brancato G, Rega N and Barone V 2008 *J. Chem. Phys.* **128** 144501
[200] Mancini G, Brancato G and Barone V 2014 *J. Chem. Theory Comput.* **10** 1150–1163
[201] Mancini G, Brancato G, Chandramouli B and Barone V 2015 *Chem. Phys. Lett.* **625** 186–192 ISSN 0009-2614
[202] Mennucci B 2012 *WIREs Comput. Mol. Sci.* **2** 386–404
[203] Del Galdo S, Mancini G, Daidone I, Polzi L Z, Amadei A and Barone V 2018 *J. Comput. Chem.* **39** 1747–1756
[204] Desiraju G R 1989 *Crystal Engineering: The Design of Organic Solids* (Elsevier, Amsterdam)
[205] Erba A, Baima J, Bush I, Orlando R and Dovesi R 2017 *J. Chem. Theory Comput.* **13** 5019–5027
[206] Dovesi R, Erba A, Orlando R, Zicovich-Wilson C M, Civalleri B, Maschio L, Rérat M, Casassa S, Baima J, Salustro S and Kirtman B 2018 *WIREs Comput. Mol. Sci.* **8** e1360
[207] Sauer J 2019 *Acc. Chem. Res.* **52** 3502–3510