PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2024, **26**, 5802

Quantum chemistry meets high-resolution spectroscopy for characterizing the molecular bricks of life in the gas-phase†

Vincenzo Barone

Received 24th October 2023,
Accepted 28th November 2023

DOI: 10.1039/d3cp05169b

rsc.li/pccp

Computation of accurate geometrical structures and spectroscopic properties of large flexible molecules in the gas-phase is tackled at an affordable cost using a general exploration/exploitation strategy. The most distinctive feature of the approach is the careful selection of different quantum chemical models for energies, geometries and vibrational frequencies with the aim of maximizing the accuracy of the overall description while retaining a reasonable cost for all the steps. In particular, a composite wave-function method is used for energies, whereas a double-hybrid functional (with the addition of core–valence correlation) is employed for geometries and harmonic frequencies and a cheaper hybrid functional for anharmonic contributions. A thorough benchmark based on a wide range of prototypical molecular bricks of life shows that the proposed strategy is close to the accuracy of state-of-the-art composite wave-function methods, and is applicable to much larger systems. A freely available web-utility post-processes the geometries optimized by standard electronic structure codes paving the way toward the accurate yet not prohibitively expensive study of medium- to large-sized molecules by experimentally-oriented researchers.

1 Introduction

Accurate determination of molecular structures is a fundamental pursuit in several fields of molecular sciences and gas-phase data are the mandatory prerequisite for disentangling the role of intrinsic stereo-electronic and environmental effects in tuning processes occurring in condensed phases. In this framework high-resolution molecular spectroscopy plays a central role since it provides accurate information on intrinsic molecular features[1–5] and allows for the unequivocal identification of chemical species in gaseous samples of unknown composition[4,6–8] also in hostile environments, *e.g.* in interstellar space[9,10] or the atmospheres of exoplanets.[11]

Recently, rotational spectroscopy has been extended to the investigation of solid thermolabile molecules (like most molecular bricks of life) thanks to the introduction of the laser ablation (LA) technique.[12–17] Furthermore, supersonic-jet expansion has resulted in great simplification of rotational spectra due to the cooling of molecules produced in the LA step to low rotational temperatures. Finally, chirp-pulse microwave (MW) spectrometers[18,19] combined with fast-mixing nozzles permit the investigation of non-covalent interactions in large systems.[20–23]

Unfortunately, direct interpretation of the spectroscopic signals in structural and dynamic terms is seldom straightforward, as for prototypical molecular bricks of life (see Fig. 1). In this respect, molecular simulations can play an invaluable role, provided that they are able to couple accuracy and feasibility.[5,24,25] One of the most effective strategies to reach this goal is based on an integrated computational approach that employs quantum chemical (QC) models of increasing accuracy in the different steps of an exploration/exploitation workflow. The main steps of this strategy[26–29] can be summarized as follows:

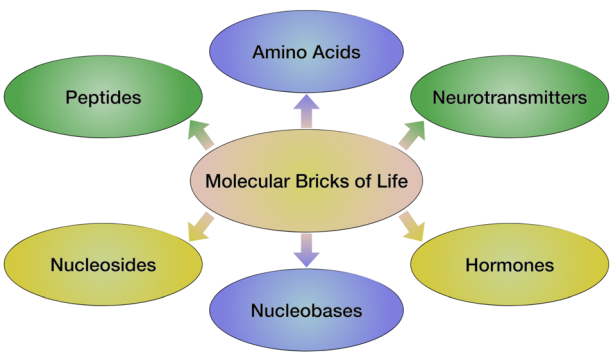

Fig. 1 The main families of molecular bricks of life analyzed in the present paper.

Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy.
E-mail: vincenzo.barone@sns.it

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3cp05169b

(1) Unsupervised perception of the molecular system with the aim of disentangling hard and soft degrees of freedom and, possibly, identifying the most suitable fragmentation patterns.[30]

(2) Exploration of the PES governed by soft degrees of freedom using a fast semi-empirical method,[31] guided by knowledge-based and evolutionary algorithms[26,32] and followed by the refinement of the structures of the most stable minima[29] and, possibly, by the analysis of relaxation paths between pairs of adjacent minima.[28]

(3) Determination of accurate geometries and force fields for the most stable structures not involved in fast relaxation processes.[33,34]

(4) Evaluation of accurate electronic energies and properties for the final panel of low-energy minima.[35,36]

(5) Computation of relative populations and spectroscopic parameters under the experimental conditions of interest employing the quantities obtained in steps 3 and 4.[37–41]

In step 1, the Proxima software[30] for molecular perception is used to produce automatically a full list of stereo-isomers (and/or tautomers) and to identify soft degrees of freedom. In this step, special attention is paid to pseudo-rotation coordinates for describing the puckering of rings not involving π-electron conjugation[42,43] and the combinations of dihedral angles around each bond not belonging to cycles and with a bond order lower than a pre-defined threshold.

The main aspects of all the other steps will be analyzed in the following sections, taking into account the latest achievements of contemporary computational chemistry, and, in particular, the availability of reduced-cost QC methods for the accurate evaluation of electronic energies for large molecules.[44–49] Unfortunately, equilibrium geometries and vibrational frequencies can benefit only marginally from these developments, which have not yet been extended to analytical energy derivatives. However, the intrinsic errors affecting the standard methods employed in the current spectroscopic studies (typically second-order Møller–Plesset perturbation theory (MP2)[50] or hybrid density functionals)[51–54] hamper any *a priori* prediction of the spectroscopic outcome. Therefore, the current practice is to resort to *a posteriori* interpretations in terms of the agreement between experimental and computed spectroscopic parameters, irrespective of the computed stability of the selected species.[55] Improved results can be obtained by combining different QC methods and correcting the DFT geometrical parameters with bond-specific scaling factors[56] derived from a large database of accurate molecular structures,[57,58] or with reference to suitable fragments, whose accurate geometries are already known.[59] While both approaches have been combined with remarkable success in the so-called Nano-LEGO model,[57,58,60] the use of a large number of parameters remains quite unsatisfactory and, above all, suitable fragments are not always available.[35,61] Based on these premises, it is the purpose of this paper to show that unbiased prediction and interpretation of high-resolution spectra can be obtained by means of the Pisa Composite Scheme (PCS),[33–36] which consistently improves the accuracy of current approaches for molecules containing up to a few dozen atoms, while retaining a black-box nature and the use of reasonable computational resources.

2 Methods and computational tools

2.1 Exploration of rugged potential energy surfaces

A knowledge-based systematic search of soft degrees of freedom[32] can be optionally performed in order to obtain a panel of guess structures (*e.g.* the 3^n staggered conformers generated by rotations around n non-terminal single bonds, which are not part of cycles), whose geometries are then optimized using the fast GFN2-XTB semi-empirical method.[31] Then, custom implementation of the island model evolutionary algorithm (IM-EA)[62] is employed to produce iteratively other candidates with maximum diversity of soft degrees of freedom. To this end, different genetic operators are applied to disjointed regions (islands) of the potential energy surface (PES) governed by soft degrees of freedom, which are mixed only at predefined intervals by a dedicated operator (migration). In order to further increase the coverage of the PES, several runs with different initial populations are performed for each molecular system. Further details can be found in ref. 26 and 29.

Alternative strategies involving universal machine learning (ML) potentials to search for low-energy structures have been proposed[63] and are possibly competitive with the evolutionary algorithm sketched above, but a comparison of different exploration strategies is beyond the scope of the present paper. In any case, at the end of the whole exploration, low-energy structures within a pre-defined energy range are selected by discarding too similar structures and then performing single point energy evaluations at the B3LYP/6-31+G* level,[64] also including Grimme's D3BJ dispersion corrections.[65] This computational model (hereafter referred to as B3/SVP) is used only for the selection of an initial panel of structures to be next refined at higher levels and has been chosen because it couples a remarkable computational efficiency with the prediction of reasonable anharmonic contributions (*vide infra*).

In the next step, structures lying within a smaller energy range are optimized at the same level and the surviving ones define the panel of candidates for the final structural refinement, which is performed employing the revDSD-PBEP86-D3BJ double-hybrid functional[66] (hereafter rDSD), for which effective analytical gradients and Hessians are available.[67] Several studies have shown that this functional provides excellent geometrical structures,[57] dipole moments,[68] spectroscopic parameters,[69] non-covalent intermolecular interactions,[70,71] and conformational landscapes.[72–74] In the present paper, the rDSD functional is employed in conjunction with a modified cc-pVTZ-F12[75] basis set in which d functions on first-row atoms are removed and the two f functions on second- and third-row atoms are replaced by a single f function taken from the cc-pVTZ basis set.[76] The resulting basis set (referred to as 3F12$^-$ in the following) has dimensions comparable with those of the jun-cc-pVTZ basis set employed systematically in previous studies,[77] but it delivers results much closer to those of augmented quadruple-zeta basis sets,[33] which are, in turn, close to the complete basis set (CBS) limit.[78]

The integrated strategy sketched above minimizes the number of expensive geometry optimizations with a negligible reduction in accuracy in the final results. The different energy thresholds depend on the system and the spectroscopic

technique of interest. In general terms, a conservative limit for the relative stability of detectable conformers is around 900 cm^{-1} (which corresponds to a relative population of about 1% at room temperature, where kT/hc = 207 cm^{-1}).[28,79] As a consequence, the typical thresholds for the acceptance of semi-empirical structures, B3/SVP geometry optimizations and final rDSD/3F12$^-$ refinement are 2500, 1500 and 1000 cm^{-1}, respectively.

As mentioned in the introduction, relaxation toward more stable structures can take place whenever the energy barriers ruling this process are sufficiently low, with a typical threshold being about 400 cm^{-1}.[80–82] With the aim of unraveling those fast relaxations, the paths connecting adjacent minima are analyzed systematically at the rDSD/3F12$^-$ level and the structures of the most significant transition states (TSs) estimated in this way are refined by full geometry optimizations at the same level.

Finally, dipole moments, quadrupolar coupling constants, and other one-electron properties are computed with good accuracy at the rDSD/3F12$^-$ level, whereas electronic energies are refined by the composite wave-function method described in the next subsection. Noted is that, while rDSD/3F12$^-$ geometries are sufficient for semi-quantitative purposes, the models developed for obtaining the improved structures needed for computing accurate rotational constants will be analyzed after the discussion of relative stabilities.

2.2 Accurate relative stabilities

The composite methods developed by our group during recent years[15,25,36,77,83,84] start from frozen core (fc) energies computed by second order Møller–Plesset perturbation theory (MP2)[50] at the complete basis set (CBS) limit (E^{V2}), which are then integrated by post-MP2 valence contributions (ΔE^{V}) and MP2 core–valence (CV) correlation (ΔE^{CV2}) in the framework of the additivity approximation:

$$E = E^{V2} + \Delta E^{V} + \Delta E^{CV2} \quad (1)$$

This formulation permits the use of different QC models to estimate the ΔE^{V} contribution, and different basis sets (possibly also CBS extrapolations) for computing E^{V2} and ΔE^{V}, whereas ΔE^{CV2} is always evaluated using the following equation:

$$\Delta E^{CV2} = E(\text{MP2(ae)/3wC}) - E(\text{MP2(fc)/3wC}) \quad (2)$$

where ae means all-electrons and 3wC is the correlation consistent cc-pwCVTZ basis set.[85]

Explicitly correlated (F12) approaches[86,87] are employed in the new version of the Pisa composite scheme (PCS/F12) introduced in the present paper to evaluate E^{V2} and ΔE^{V} in conjunction with the cc-pVnZ-F12 family of basis sets[75] (hereafter nF12). Thanks to the balanced optimization of these basis sets for Hartree–Fock (HF) and MP2-F12[86] computations and to the inclusion of the complementary auxiliary basis set (CABS) in the HF step,[88,89] the CBS limit of the whole MP2-F12 energy (including the HF contribution) can be estimated accurately using the two-point extrapolation proposed by Helgaker:[90]

$$E^{V2} = \frac{4^3 E(\text{fc-MP2-F12/4F12}) - 3^3 E(\text{fc-MP2-F12/3F12})}{4^3 - 3^3}, \quad (3)$$

For systems not involving strong non-dynamical correlation, the coupled cluster anzatz including single, double and (perturbatively) triple excitations (CCSD(T))[91] can be confidently employed to evaluate post-MP2 contributions.[25,77] The CCSD(F12*) variant[92] provides very reliable post-MP2 contributions from single and double excitations, but, unfortunately, the straightforward addition of a perturbative estimate of triple excitations does not benefit from the explicitly-correlated contribution. However, a size-consistent estimate of this contribution (T+) has been recently proposed[93] and is employed in the PCS/F12 approach. Therefore, the ΔE^{V} contribution is obtained by a two-point extrapolation[90] of CCSD(F12*)(T+) energies in conjunction with the 2F12$^+$ and 3F12$^+$ basis sets, with the + superscript indicating the addition of a single set of f or g functions (taken from the cc-pVTZ or cc-pVQZ basis set) to the standard 2F12 or 3F12 basis set only for third-row atoms:

$$\Delta E^{V} = \frac{3^3 \Delta E^{CC}(\text{3F12}^+) - 2^3 \Delta E^{CC}(\text{2F12}^+)}{3^3 - 2^3} \quad (4)$$

with

$$\Delta E^{CC} = E(\text{CCSD(F12*)(T+)}) - E(\text{MP2-F12}) \quad (5)$$

Noted is that replacement of the nF12 basis sets with their nF12$^+$ counterparts improves significantly the results for molecules containing third-row atoms. Furthermore, the cardinal numbers of the basis sets and the exponent in the extrapolation eqn (2) and (3) could be in principle optimized.[94] However, several tests have shown that the results delivered by the standard choices (maximum angular momentum for the cardinal number of the basis set and exponent of 3)[90] provide accurate results, possibly because of the good saturation of the valence space in F12 basis sets.[75] The density fitting (DF) approximation was used throughout the HF and post-HF calculations employing the aug-cc-pV(n+1)Z-RI-JK[95] and aug-cc-pwCV(n+1)Z-RI[96] fitting basis sets, respectively. Slater-type f_{12} correlation factors (fitted with 6 Gaussians each) with exponents of 0.9, 1.0, and 1.1 were employed for the 2F12$^+$, 3F12 (or 3F12$^+$), and 4F12 basis sets, respectively.[97]

The choice of the computational level for the different terms included in the new PCS/F12 model has been based on the requirement of reasonable resources for each of them with specific reference to molecules containing up to about 20 atoms. In particular, the cost of CCSD(F12*)(T+)/2F12$^+$ and MP2-F12/4F12 computations is negligible with respect to their CCSD(F12*)(T+)/3F12$^+$ counterparts, which are, in turn, fully feasible on a standard workstation for molecules as large as guanine.[35] For larger systems, containing up to about 50 atoms, single point CCSD(F12*)(T+)/3F12$^+$ energy evaluations can be routinely performed thanks to efficient parallelization and implementation of frozen natural orbital (FNO), natural auxiliary functions (NAF) and/or related reduced-scaling techniques in

several computer codes.[44–49] The results of several benchmark studies have shown that FNO (and related) approximations produce results very close to those obtained from full computations, while reducing the computer time by about five times.[36]

All the PCS/F12 computations have been performed with the MRCC[98] code.

The relative stability of low-energy minima is not related directly to the electronic energy differences (ΔE), but, rather, to the corresponding relative enthalpies at 0 K (ΔH_0°) or free energies (ΔG°) at a temperature depending on the experimental conditions. In this connection, quantitative comparison with experiments requires going beyond the current practice of using scaled harmonic frequencies[99–101] by incorporating explicit anharmonic contributions. To this end, equilibrium rotational constants (B^{eq}), harmonic frequencies (ω_i) and Coriolis couplings ($\zeta_{ij,\tau}$) are computed at the rDSD/3F12$^-$ level,[67] whereas cubic (f_{ijk}) and semi-diagonal quartic (f_{iijj}) force constants are obtained from finite differences of analytical B3/SVP Hessians.[37] Then anharmonic zero point energies (ZPEs) are estimated in the framework of second order vibrational perturbation theory (VPT2) by means of an analytical and resonance-free expression:[102,103]

$$\mathrm{ZPE} = \sum_{i=1}^{N}\frac{\omega_i}{2} + \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{f_{iijj}}{32} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\left[\frac{f_{ijk}{}^2}{48(\omega_i+\omega_j+\omega_k)} + \frac{f_{iik}f_{jjk}}{32\omega_k}\right] - \sum_{\tau=x,y,z}\frac{B_\tau^{\mathrm{eq}}}{4}\left[1 - \sum_{i=1}^{N1}\sum_{j=i+1}^{N}\{\zeta_{ij,\tau}\}^2\frac{(\omega_i-\omega_j)^2}{\omega_i\omega_j}\right] \quad (6)$$

2.3 Accurate molecular structures

Derivation of eqn (1) w.r.t. Cartesian coordinates leads to energy gradients, which can be employed for geometry optimizations by composite methods.[104,105] However, nearly identical results[106] are obtained applying the additivity approximation directly to geometrical parameters (r) optimized separately at different levels:

$$r = r^{\mathrm{V2}} + \Delta r^{\mathrm{V}} + \Delta r^{\mathrm{CV2}} \quad (7)$$

with

$$\Delta r^{\mathrm{CV2}} = r(\mathrm{MP2(ae)/3wC}) - r(\mathrm{MP2(fc)/3wC}) \quad (8)$$

Thanks to the F12 ansatz, accurate geometries are obtained neglecting the CBS extrapolation and employing the 2F12$^+$ basis set for both r^{V2} and Δr^{V} contributions. Furthermore the simpler CCSD(F12b)(T) approach[87] (whose analytical gradients are available) provides fully reliable results. Under these circumstances eqn (7) becomes

$$r^{\mathrm{PCS/F12}} = r(\mathrm{CCSD(F12b)(T)/2F12^+}) + \Delta r^{\mathrm{CV2}} \quad (9)$$

All the geometry optimizations required by the PCS/F12 model have been performed with the MOLPRO program.[107]

An accurate reduced-cost version of the PCS model (referred to as PCS/DFT) has been recently introduced,[33] in which the Δr^{CV2} contribution is retained, but the coupled cluster model is replaced by the much cheaper rDSD double hybrid functional[66] and the 3F12 basis set is employed in both the r^{V2} and Δr^{V} terms:

$$r^{\mathrm{PCS/DFT}} = r(r\mathrm{DSD/3F12}) + \Delta r^{\mathrm{CV2}} \quad (10)$$

It is noteworthy that the same Δr^{CV2} correction leads to accurate results for both PCS/F12 and PCS/DFT models without employing in the latter case any empirical parameter in addition to those present in the underlying functional.[33] However, while the cost of the Δr^{CV2} correction is negligible in the framework of the PCS/F12 model, it increases by about three times the cost of PCS/DFT computations. A first saving of computer resources can be obtained replacing the 3F12 basis set by its smaller 3F12$^-$ counterpart with negligible reduction of accuracy.[34] The way toward further progress is paved by the observation that in all the examined cases the Δr^{CV2} contribution to valence and dihedral angles can be safely neglected, whereas this is not the case for bond lengths. Then, bonded atoms can be easily identified employing covalent radii (r^{cov}, taken for instance from ref. 108) and Pauling bond orders P_{ij}:[109]

$$P_{ij} = \exp\{(r_i^{\mathrm{cov}} + r_j^{\mathrm{cov}} - r_{ij})/0.3\} \quad (11)$$

In eqn (11) r_{ij} is the interatomic distance in Å and two atoms are considered bonded if P_{ij} is larger than 0.3 (which corresponds to a distance 0.35 Å longer than the sum of the covalent radii). While this approximation does not reduce *per se* the required computational resources, it paves the way toward a reduced-cost model since the Δr^{CV2} correction to bond lengths is well approximated by a simple one-parameter function of the corresponding rDSD bond lengths and the principal quantum numbers (n) of the involved atoms.[34] Here a further step is taken by employing the same covalent radii already used in eqn (11) in place of the rDSD/3F12$^-$ bond lengths in order to obtain a method-independent estimate of the CV contributions (Δr^{CVB}), which is then supplemented by a method-dependent correction (Δr^{VB}) of small inaccuracies in the treatment of valence electrons:

$$r^{\mathrm{PCS/Bonds}} = r(\mathrm{rDSD/3F12^-}) + \sum_{ij}^{\mathrm{bonded}}\Delta r_{ij}^{\mathrm{B}} \quad (12)$$

with

$$\Delta r_{ij}^{\mathrm{B}} = \Delta r_{ij}^{\mathrm{CVB}} + \Delta r_{ij}^{\mathrm{VB}} \quad (13)$$

and

$$\Delta r_{ij}^{\mathrm{CVB}} = -k_{\mathrm{CV}}\sqrt{n_i n_j - 1}\left(r_i^{\mathrm{cov}} + r_j^{\mathrm{cov}}\right) \quad (14)$$

Noted is that eqn (14) leads rightly to vanishing corrections for bonds between first-row atoms.

In the case of the rDSD/3F12$^-$ model, Δr^{VB} correction is needed only for counterbalancing slight errors for delocalization (Δ^{del}) and hyperconjugation (Δ^{hyp}):

$$\Delta r^{\mathrm{VB}} = \Delta r^{\mathrm{del}} + \Delta r^{\mathrm{hyp}} \quad (15)$$

The following simple expressions have the correct behaviour and involve just one additional parameter:

$$\Delta r_{ij}^{\mathrm{del}} = \Delta r_{ij}^{\mathrm{CVB}} \left[\sqrt{|P_{ij}-2|}-1\right][1-\delta(i,O)-\delta(j,O)+\delta(i,O)\delta(j,O)] \quad (16)$$

and

$$\Delta r_{ij}^{\mathrm{hyp}} = -k_{\mathrm{hyp}}(P_{ij}-1)^2[\delta(i,C)\delta(j,F)+\delta(i,F)\delta(j,C)] \quad (17)$$

The new PCS/bond model employs Kronecker δs to tune the inclusion of the corrective terms for selected bonds and just two empirical parameters (k_{CV} = 0.0011 and k_{hyp} = 0.025). In the next sections it will be shown that accurate equilibrium geometries and rotational constants can be obtained for saturated and unsaturated molecules containing H, C, N, O, F and S atoms with computer times 5 to 10 times shorter than those required by the PCS/DFT model and orders of magnitude shorter than those required by PCS/F12. All the geometry optimizations required by the PCS/DFT and PCS/bond models have been performed with the Gaussian package,[110] whereas the computation of the final PCS/bond geometries has been implemented using a website (**https://www.skies-village.it/proxima/pcsbonds/**), which, starting from a given set of atomic numbers and Cartesian coordinates, generates interactively the corresponding PCS geometry and a 2D representation of the molecular structure.[111]

2.4 Top-down and bottom-up strategies

Experimental rotational constants represent very specific and accurate fingerprints of molecular structures, which, depending on the number of available experimental data, can be employed in two different joint theory-spectroscopy strategies, known as top-down and bottom-up approaches, respectively.[3]

Indeed, a generic rotational constant is inversely proportional to the corresponding principal inertia moment, which, in turn, only depends on molecular geometry and isotopic masses. However, direct structural determinations from experimental rotational constants are hampered by the limited amount of data with respect to the number of geometrical parameters and by the appropriate account of vibrational effects. The first limitation can be overcome by considering for a given molecule different isotopic species. However, determination of the couplings between rotations and vibrations needed to go from vibrationally averaged to equilibrium rotational constants is practically impossible for polyatomic molecules. Therefore, in the top-down approach the so called semi-experimental (SE) equilibrium geometry[112] is obtained by a least-squares fit of the experimental rotational constants for the vibrational ground state (B^0) of different isotopologues corrected by computed electronic (ΔB^{el}) and vibrational (ΔB^{vib}) contributions:[113]

$$B_\tau^0 = B_\tau^{\mathrm{eq}} + \Delta B_\tau^{\mathrm{el}} + \Delta B_\tau^{\mathrm{vib}} \quad (18)$$

with

$$\Delta B_\tau^{\mathrm{el}} = -B_\tau^{\mathrm{eq}} \frac{m}{M_{\mathrm{p}}} g_{\tau\tau} \quad (19)$$

where m is the electron mass, M_{p} the proton mass and the rotational g tensor (expressed in eqn (19) in units of the nuclear magneton) can be safely computed by hybrid density functionals using London orbitals.[34,59] On the other hand, the computation of ΔB^{vib} requires harmonic frequencies (ω_i), Coriolis couplings ($\zeta_{ij,\tau}$) and semi-diagonal third derivatives of the energy with respect to normal modes (f_{iij}):

$$\Delta B_\tau^{\mathrm{vib}} = -2(B_\tau^{\mathrm{eq}})^2 \left[\sum_{\eta=x,y,z} \frac{3(a_{i,\tau\eta})^2}{4\omega_i I_\eta^{\mathrm{eq}}} + \sum_{i,j=1}^{N} \frac{(\zeta_{ij,\tau})^2(\omega_i-\omega_j)^2}{\omega_i\omega_j(\omega_i+\omega_j)} + \pi\sqrt{\frac{c}{h}} \sum_{i,j=1}^{N} \frac{f_{iij}a_{j,\tau\tau}}{\omega_j^{3/2}}\right] \quad (20)$$

where $a_{i,\tau\eta}$ is the derivative of the τ, η component of the inertia moment with respect to normal mode i. Thanks to the lack of any resonance and the implementation of very effective VPT2 engines,[39,114] these computations can be performed without particular problems for the relatively small molecules for which sufficient experimental data are available.

Whenever the lack of experimental information is too extensive, 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 rotational constants. In general terms, the optimal level of accuracy associated with predicted rotational constants should be close to 0.1% (1 MHz for a constant of 1 GHz), which roughly corresponds to errors smaller than 0.001 Å for typical bond lengths and 0.002 radians (0.1 degrees) for typical valence angles.[115] This target accuracy can be surely obtained by expensive composite schemes incorporating high excitation orders in the correlation treatment.[116] Nevertheless, the PCS/F12 model described above and its 'cheap' predecessors[77,84] are able to draw closer to this high accuracy limit.[15,36] However, for the quite large flexible molecules of interest in the present context, even this computational level becomes too expensive, and cheaper alternatives are needed. At the same time, the magnitude of vibrational corrections is typically between 0.1% and 1.0% that of the corresponding equilibrium rotational constant. Since the accuracy of ΔB_{vib} contributions provided by global hybrid density functionals in conjunction with medium-size basis sets (*e.g.* the B3/SVP model previously defined) falls well within 5%,[59] the ensuing average error of 0.05% on the rotational constants is more than acceptable. As we will see, equilibrium rotational constants obtained by the low-cost PCS/bond variant described in the previous section in conjunction with B3/SVP vibrational corrections draw closer to the goal accuracy of 0.1% on ground vibrational state rotational constants.

3 Results and discussion

3.1 Validation

Before considering representative molecular bricks of life, a panel of medium-sized molecules has been analyzed with the

aim of validating the performance of the PCS model and determining the weight of its different contributions.

As is well known, atomization energies represent the most demanding thermochemical quantities and, actually, the so-called chemical accuracy (4 kJ mol^{-1}) is considered a satisfactory target. The first validation of PCS/F12 has been performed with respect to the W4/11 set of highly accurate atomization energies.[117] Concerning valence contributions, the mean unsigned error (MUE) and root mean square deviation (RMSD) are below 1 kJ mol^{-1} (0.7 and 0.9, respectively) and the maximum error (MAX) is below 4.0 kJ mol^{-1}. Since the average value of the contribution of higher excitations (CCSDTQ5–CCSD(T)) is 1.2 kJ mol^{-1}, the accuracy of PCS/F12 results is well within the intrinsic error bar of composite methods including up to CCSD(T) contributions. The situation is more involved for the CV contribution, due to the intrinsic error (around 3 kJ mol^{-1}) related to its evaluation at the MP2 level. While further work is needed to devise a more accurate approximation not becoming the rate determining step of the whole PCS/F12 model, already the present version provides reliable results for reaction energies including tautomerization equilibria.[36]

The accuracy of the PCS/F12 model for conformational equilibria has been assessed with reference to the conformational landscape of the prototypical glycine amino acid (see Table 1), which shows eight different energy minima.[118]

A comparison with the state-of-the-art results of ref. 119 (including up to CCSDT(Q) contributions) confirms the accuracy of the PCS/F12 model, with the MAX and MUE between the two models being 17 and 8 cm^{-1} (0.2 and 0.1 kJ mol^{-1}), respectively. This quantitative agreement is not unexpected since conformational equilibria are quite well described using low-level QC methods: as a matter of fact, rDSD/3F12$^-$ relative energies are already usually sufficient for semi-quantitative analyses of this kind of problems.[118]

Next, the accuracy of PCS/bond molecular structures is examined for the panel of medium-sized semi-rigid molecules shown in Fig. 2, whose computed and experimental rotational constants are collected in Table 2. When needed, the difference between rDSD/3F12$^-$ and PCS/bond equilibrium rotational constants will be labeled ΔB^{B}. It is quite apparent that the rDSD/3F12$^-$ model already produces respectable results, but systematically overestimates the experimental ground state rotational constants. Among the different corrections to the rDSD/3F12$^-$ equilibrium rotational constants, ΔB^{vib} plays the leading role, reducing systematically the computed values by up to 1%. This effect is partially counterbalanced by the ΔB^{B} contribution, which, however, never exceeds 0.3% of the corresponding B^{eq} value. Finally, the ΔB^{el} contribution is smaller than the target accuracy of the proposed computational approach for all the molecules considered in the present study except H_2CS. As such, in the following this contribution will be neglected.

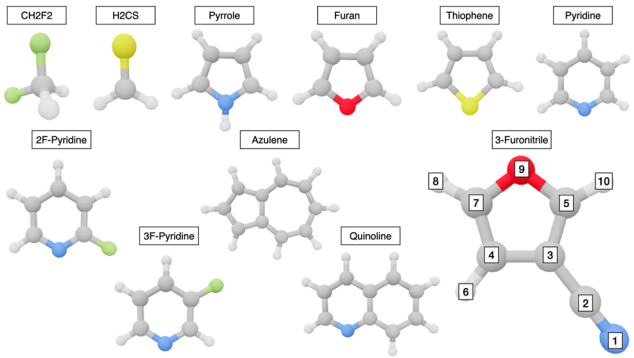

Fig. 2 Semi-rigid molecules of the validation set.

In summary, the PCS/bond results always fulfill the target accuracy of 0.1%, with this confirming that a double-hybrid functional in conjunction with a sufficiently large basis set does not require a huge number of bond-specific corrections and, even more importantly, any correction to valence and dihedral angles.

An intuitive picture of the performance of the model is offered through graphical representation of the results based on normal distributions defined by:

$$\rho(x) = N_{\mathrm{c}} \mathrm{e}^{-0.5\left(\frac{x-\Delta_{\mathrm{av}}}{\Delta_{\mathrm{std}}}\right)^2} \quad (21)$$

where Δ_{av} and Δ_{std} are the relative unsigned mean error and standard deviation, respectively, whereas N_{c} is a suitable normalization constant. The results shown in Fig. 3 confirm the remarkable accuracy and robustness of the PCS/bond model.

The rotational constants delivered by the PCS/F12 and PCS/bond variants have been compared for the two most stable conformers of glycine (see Fig. 4), together with the two low-energy forms of cycloserine, and the three tautomers of creatinine (see Fig. 5), for which experimental rotational constants are available and PCS/F12 computations are feasible. The results collected in Table 3 show that the quality of the PCS/bond results is even marginally better than that of their PCS/F12 counterparts, except for cycloserine, where PCS/F12 rotational constants are exceptionally accurate.

A comparison between the B3/SVP and the more accurate (but much more expensive) rDSD/3F12$^-$ vibrational corrections for glycine confirms the reliability of the cheaper approach except in the case of glycine II: as a matter of fact the very flat potential energy surface around this conformer (see the section on amino acids) enforces more stringent constraints on the level of

Table 1 Relative electronic energies of the eight conformers of glycine computed using different methods. The conformer nomenclature is taken from ref. 118 and all the values are in cm^{-1}

Conformer	Reference[a]	PCS/F12	rDSD/3F12$^-$
I (ttt)	0	0	0
II (ccc)	226	226	209
I′ (gtt)	435	426	443
III (tct)	602	619	585
III′ (gct)	928	936	919
Ic (ttc)	1680	1689	1680
IIIc (tcc)	2040	2056	2023
I′c (gtc)	2123	2123	2132
MAX		17	17
MUE		8	11

[a] From ref. 119.

Table 2 Computed and experimental rotational constants (in MHz) of prototypical semi-rigid molecules; rDSD stands for rDSD/3F12$^-$ and PCS for PCS/bond

Molecule	Axis	B^{eq} (rDSD)	ΔB^{B} (ΔB^{el})	ΔB^{vib}	B^0 (PCS)[a]	B^0 (exp.)[b]
CH_2F_2	*a*	49 423.3	285.8(−2.1)	−514.8	49 194.3	49 142.9
	b	10 583.7	67.7(−0.2)	−61.0	10 590.4	10 604.8
	c	9251.0	60.4(−0.2)	−69.4	9242.0	9249.8
H_2CS	*a*	293 949.7	595.1(−843.4)	−1958.7	291 742.7	291 613.3
	b	17 755.9	21.0(−1.3)	−74.0	17 701.6	17 699.0
	c	16 744.5	20.6(−0.2)	−108.2	16 656.7	16 652.5
Pyrrole	*a*	9174.2	27.9(−0.5)	−73.7	9128.4	9130.6
	b	9042.5	24.8(−0.3)	−69.8	8997.5	9001.3
	c	4554.0	13.1(0.2)	−36.7	4530.4	4532.1
Furan	*a*	9493.0	32.6(−0.5)	−78.7	9446.9	9447.1
	b	9287.1	25.8(−0.5)	−65.4	9247.5	9246.7
	c	4694.5	14.5(0.1)	−38.3	4670.7	4670.8
Thiophene	*a*	8077.0	27.3(0.4)	−59.5	8044.8	8041.8
	b	5434.1	18.1(0.2)	−29.6	5422.6	5418.1
	c	3248.5	10.9(−0.1)	−21.7	3237.7	3235.8
Pyridine	*a*	6067.4	15.7(−0.3)	−44.7	6038.4	6039.3
	b	5832.9	17.2(−0.4)	−37.1	5813.0	5804.9
	c	2973.6	8.6(0.1)	−21.3	2960.9	2959.2
2F-Pyridine	*a*	5901.6	16.4(−0.3)	−43.8	5874.2	5870.9
	b	2704.9	8.1(−0.1)	−15.8	2697.2	2700.0
	c	1854.8	5.4(0.0)	−12.1	1848.1	1849.2
3F-Pyridine	*a*	5860.8	16.8(−0.3)	−42.6	5835.0	5829.7
	b	2642.1	8.3(−0.1)	−14.8	2635.6	2637.5
	c	1821.1	5.6(0.0)	−11.5	1815.2	1815.7
3-Furonitrile	*a*	9340.4	30.3	−75.1	9295.6	9296.5
	b	1939.9	6.3	−6.1	1940.1	1940.3
	c	1606.3	5.2	−7.0	1604.5	1604.6
Azulene	*a*	2856.4	8.9	−20.1	2843.2	2842.0
	b	1259.2	3.5	−7.7	1255.0	1254.8
	c	873.9	2.6	−5.4	871.1	870.7
Quinoline	*a*	3159.0	9.5	−22.4	3146.1	3145.4
	b	1275.8	3.7	−7.9	1271.6	1271.6
	c	908.8	2.6	−5.7	905.7	905.7
MAX%		0.801			0.140	
MUE%		0.397			0.047	

[a] Not including ΔB^{el} except for H_2CS. [b] The experimental ground state rotational constants have been rounded to one decimal place.

anharmonic contributions. However, systematic benchmarks[59,118] have shown that this situation is quite unusual for the molecules of interest in the present study.

The relative stability of different forms of cycloserine is very similar at the rDSD/3F12$^-$ and PCS/F12 level (195 and 205 cm^{-1}, respectively) and the same applies to the relative stability of the EI and A1 forms of creatinine with respect to their most stable ZI counterpart (28 and 680 cm^{-1} at the RDSD/3F12$^-$ level to be compared with 20 and 645 cm^{-1} at the PCS/F12 level). While rDSD/3F12$^-$ electronic energies are never used in the final evaluation of the relative populations of low energy minima, the above results give further support to their use in the last steps of PES explorations.

A direct comparison between experimental and computed geometrical parameters would allow an unbiased analysis of the performance of the PCS/bond model, but, unfortunately, accurate experimental structures are not available for the quite large molecules, which are the main targets of the present study. However, the very recent determination of the semi-experimental equilibrium structure of 3-furonitrile[122] (see Fig. 2) permits a detailed comparison for a heteroaromatic molecule containing a quasi-linear substituent. The results collected in Table 4 show that the accuracy of all the geometrical parameters is in the expected range and rivals that of the much heavier 'cheap' composite wave-function method,[71,77] which is not applicable to larger molecules. Actually, the comparison is biased toward the 'cheap' results, since the semi-experimental equilibrium geometry was determined freezing bond lengths and valence angles involving hydrogen atoms at their 'cheap' values. Indeed, the maximum and mean relative unsigned errors (MAX% and MUE%, respectively) with respect to semi-experimental rotational constants suggest that in this

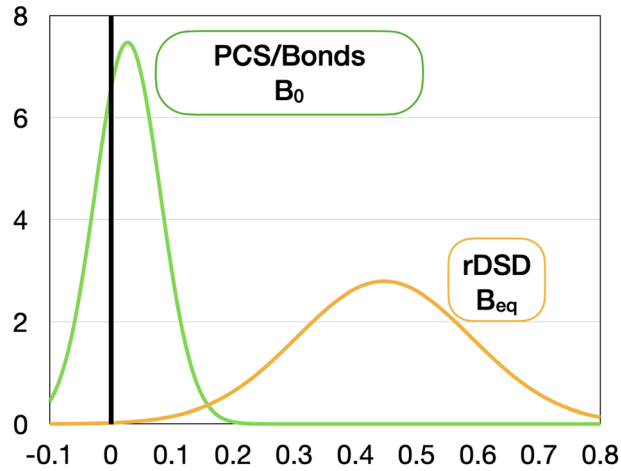

Fig. 3 Error statistics for the semi-rigid molecules of the validation set.

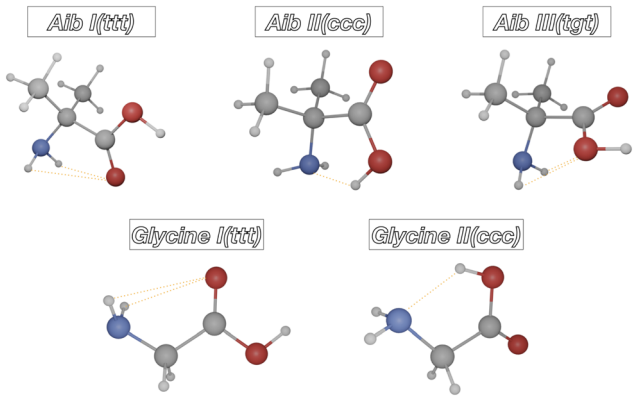

Fig. 4 Molecular structure of the glycine and aminoisobutyric acid (Aib) conformers detected in the rotational spectra.

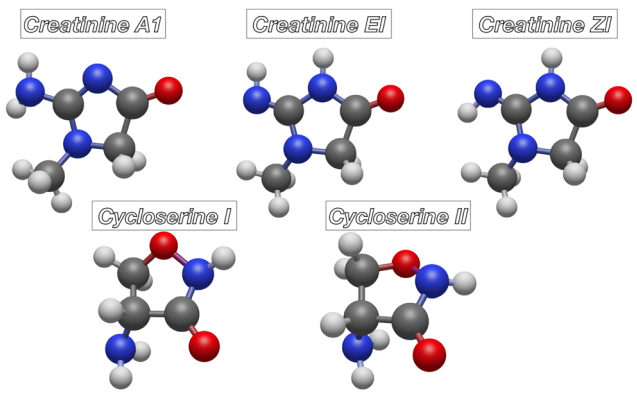

Fig. 5 Molecular structures of the cycloserine isomers and creatinine tautomers detected in the rotational spectra.

specific case the PCS/bond model might be even more accurate than its 'cheap' counterpart.

Several molecules of current biological and medicinal interest are quite flexible, with this feature increasing the difficulty

Table 3 Ground state rotational constants (in MHz) for the conformers of glycine, together with the forms of cycloserine and tautomers of creatinine detected using rotational spectroscopy (see Fig. 4 and 5 for the molecular structures)

Species	Parameter	$B^0_{exp.}$ [a]	ΔB_{vib} [b]	PCS/F12	PCS/bond
Glycine I (ttt)	B_a	10 341.5	−77.1 (−76.2)	10 319.5	10 316.1
	B_b	3876.2	−29.5 (−31.9)	3871.6	3869.3
	B_c	2912.4	−22.7 (−22.7)	2920.3	2907.6
Glycine II (ccc)	B_a	10 130.2	−44.0 (−60.4)	10 144.9	10 128.5
	B_b	4071.5	−39.5 (−37.1)	4056.1	4059.3
	B_c	3007.5	−37.4 (−34.8)	2993.2	2998.8
MAX%				0.48	0.30
MUE%				0.27	0.20
Cycloserine I	B_a	3685.0	−19.0	3683.1	3678.0
	B_b	3160.5	−28.3	3158.7	3163.4
	B_c	1815.5	−15.6	1815.0	1814.1
Cycloserine II	B_a	3649.0	−38.6	3647.5	3642.6
	B_b	3168.6	−22.4	3167.8	3171.0
	B_c	1993.7	−19.6	1992.7	1988.7
MAX%				0.06	0.25
MUE%				0.05	0.15
Creatinine ZI	B_a	3842.4	−38.5	3820.2	3825.6
	B_b	1825.4	−11.5	1819.3	1824.3
	B_c	1261.0	−10.4	1259.6	1259.5
Creatinine EI	B_a	3890.2	−41.1	3879.1	3881.4
	B_b	1810.2	−11.9	1804.7	1810.4
	B_c	1258.1	−11.2	1256.9	1256.9
Creatinine A1	B_a	3840.9	−29.3	3816.5	3828.0
	B_b	1831.5	−14.0	1826.6	1830.8
	B_c	1266.9	−11.9	1268.1	1266.3
MAX%				0.64	0.44
MUE%				0.30	0.15

[a] The experimental ground state rotational constants (taken from ref. 120, 55 and 121 for glycine, cycloserine and creatinine, respectively) have been rounded to one decimal place. [b] B3/SVP and (in parenthesis) rDSD/3F12$^-$.

of predicting accurate geometrical parameters and vibrational corrections to rotational constants. In order to illustrate this point, the three medium-sized drugs shown in Fig. 6 (isoamil-acetate, aspirine and vitamin C) have been investigated. Inspection of Table 5 shows that the trends are the same as those discussed for semi-rigid molecules, but the errors increase significantly. However, also in this case the PCS/bond results represent a significant improvement over the underlying rDSD/3F12$^-$ results, which, in turn, are already much more accurate than the current standards in MW studies of biomolecule building blocks without any significant increase in the required computational resources.[123–126]

3.2 Nucleobases

Most molecular bricks of life undergo conformational or tautomeric equilibria, which, as already mentioned in the validation subsection, increase the challenges involved in the

Table 4 Computed and experimental rotational constants (in MHz) of 3-furonitrile. The atom numbering is shown in Fig. 2

Param.	Exp.[a]	Cheap[b]	rDSD/3F12^{-} [c]	PCS/bond[c]
N1–C2	1.1581	1.1570	1.1619	1.1594
C2–C3	1.4210	1.4213	1.4219	1.4196
C3–C4	1.4370	1.4363	1.4384	1.4361
C3–C5	1.3610	1.3602	1.3646	1.3619
C4–C7	1.3516	1.3509	1.3548	1.3533
C7–O9	1.3628	1.3618	1.3650	1.3633
C4–H6	1.0741	1.0741	1.0771	1.0759
C7–H8	1.0732	1.0732	1.0761	1.0749
C5–H10	1.0737	1.0737	1.0765	1.0753
≺(N1C2C3)	179.42	179.36	179.32	179.32
≺(C2C3C4)	127.40	127.44	127.40	127.40
≺(C2C3C5)	126.20	126.22	126.18	126.18
≺(C3C4C7)	105.47	105.46	105.55	105.55
≺(C4C7O9)	110.81	110.79	110.70	110.70
≺(C3C4H6)	127.08	127.08	127.40	127.40
≺(C4C7H8)	133.35	133.35	133.42	133.42
≺(O9C5H10)	117.11	117.11	117.13	117.13
≺(C3C9H10)	132.66	132.66	132.78	132.78
B_a^{eq}	9371.6	9381.4	9340.4	9370.7
B_b^{eq}	1946.4	1948.3	1939.9	1946.2
B_c^{eq}	1611.6	1613.2	1606.3	1611.5
MAX%		0.11	0.36	0.01
MUE%		0.10	0.34	0.01

[a] From ref. 122 with CH bond lengths, together with C3C4H6, C4C7H8, O9C5H10, and C3C9H10 valence angles fixed at the 'cheap' computed values and ΔB^{vib} computed in this work at the B3/SVP level. [b] For the definition and expected accuracy of the 'cheap' model see ref. 60 and 71. [c] This work.

Table 5 Rotational constants and vibrational corrections (in MHz) for the representative drugs; rDSD stands for rDSD/3F12^{-} and PCS for PCS/bond

Axis	ΔB^{vib} [a]	B^0 (exp.)[b]	B^0 (rDSD)	B^0 (PCS)
Isoamyl-acetate				
a	−28.9	3280.9	3273.3	3287.4
b	−8.6	713.0	706.8	709.1
c	−8.0	690.1	685.8	688.0
Aspirine				
a	−7.6	1156.1	1152.1	1155.4
b	−4.7	762.6	758.0	760.3
c	−3.3	509.0	506.9	507.9
Vitamin C (I)				
a	−14.4	1562.5	1553.9	1559.9
b	−5.7	715.2	711.8	714.0
c	−4.3	524.3	521.6	523.3
Vitamin C (II)				
a	−21.1	1472.7	1463.6	1469.4
b	−3.5	678.1	673.9	675.9
c	−4.0	575.0	571.6	573.1
Vitamin C (III)				
a	−8.8	1455.7	1448.8	1454.0
b	−8.4	760.5	754.1	756.4
c	−5.9	569.5	570.6	572.4
MAX%			0.87	0.56
MUE%			0.53	0.28

[a] At the B3/SVP level. [b] The experimental ground state rotational constants (taken from ref. 123–125 for isoamyl-acetate, aspirine and vitamin C, respectively) have been rounded to one decimal place.

prediction of accurate molecular structures. In particular, nucleobases are semi-rigid molecules, but are affected by tautomeric equilibria involving different π-delocalization patterns, which are quite demanding for QC computations.

The number of possible tautomers (N_T) of a given species is $N_T = N_S!/[N_H!(N_S - N_H)!]$, where N_S is the number of tautomeric sites and N_H is the number of labile protons. For nucleobases, the so called "canonical" (keto and amino) forms predominate over their "minor" enol and imino counterparts under physiological conditions. In the case of uracil, 2-thiouracil, thymine and adenine (see Fig. 7) the "canonical" tautomer is significantly more stable than all the "minor" ones also in the gas phase, and the results collected in Table 6 show that the experimental rotational constants are reproduced very well by PCS/bond computations (MAX% = 0.17 and MUE% = 0.05).

The situation is more involved for cytosine and guanine.[127] In particular, cytosine has two endo (N1 and N3) and two exo (O7 and N8) tautomeric sites and two labile protons, so that $N_S = 4$, $N_H = 2$ and $N_T = 6$. These tautomers can be classified as keto-amino (KA and KA1), enol-amino (EA), keto-imino (KI), and enol-imino (EI and EI1).

Furthermore, each enol and imino group shows, together with the most stable form, an additional rotamer (labeled with a *c* subscript). The focus of the discussion will be on the five most stable species (namely KA, EA, EAc, KI, and KIc) shown in Fig. 8, which are the only ones having non negligible

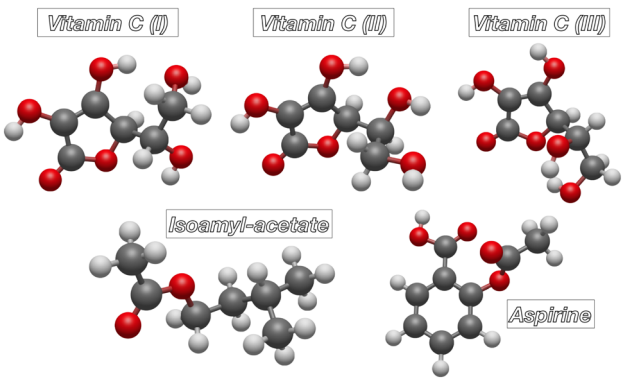

Fig. 6 Representative medium-sized drugs.

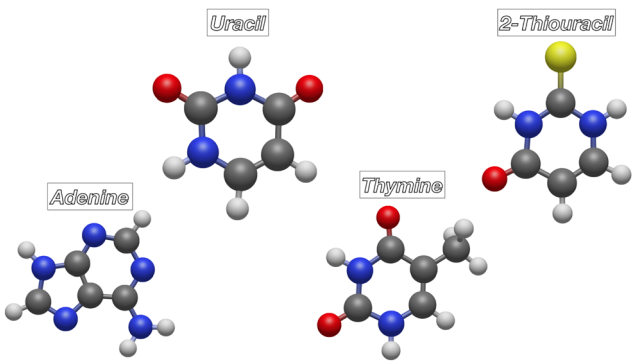

Fig. 7 Nucleobases with a single dominant tautomeric form.

Table 6 Computed and experimental rotational constants (in MHz) of uracil, 2-thiouracil, adenine and thymine; rDSD stands for rDSD/3F12$^-$ and PCS for PCS/bond

Axis	B^{eq} (rDSD)	ΔB^{B}	ΔB^{vib}	B^0 (PCS)	B^0 (exp.)[a]
Uracil					
a	3899.7	13.3	−27.7	3885.4	3883.9
b	2025.8	7.4	−10.6	2022.6	2023.7
c	1333.2	4.8	−7.4	1330.6	1330.9
2-Thiouracil					
a	3568.3	11.8	−22.9	3556.8	3555.1
b	1316.3	5.0	−6.4	1314.2	1315.0
c	961.6	3.3	−4.9	960.0	960.0
Adenine					
a	2380.9	8.7	−13.7	2375.9	2371.9
b	1576.2	5.4	−8.7	1572.9	1573.4
c	948.5	3.4	−5.8	946.1	946.3
Thymine					
a	3212.2	10.8	−23.4	3199.6	3201.2
b	1407.2	5.2	−7.7	1404.7	1404.8
c	984.5	3.5	−5.6	982.4	983.2
MAX%	0.41			0.17	
MUE%	0.23			0.05	

[a] The experimental ground state rotational constants (taken from ref. 128, 129, 130, and 131 for uracil, 2-thiouracil, adenine, and thymine, respectively) have been rounded to one decimal place.

populations in the gas phase according to all the available theoretical and experimental studies.[51,61,132]

The relative electronic energies of those tautomers and rotamers with respect to the KA species computed at the PCS/F12 level are 250, 288, 542 and 1148 cm^{-1}, respectively, whereas the relative ΔH_0° are 245, 243, 643, and 1205 cm^{-1}, respectively. Irrespective of the employed basis set, the relative stability of the EA species with respect to its KA counterpart is underestimated by about 150 cm^{-1} at the rDSD level and overestimated by about 350 cm^{-1} at the MP2 level. On the other hand, the relative stability of different rotamers and the energy difference between the KI and KA species are well reproduced by both rDSD and MP2 computations. Finally, the main effect of ZPE is to increase the relative stability of the KA species with respect to all the other tautomers (and rotamers).

Inspection of Table 7 confirms the remarkable accuracy of PCS/bond equilibrium rotational constants and the need to include vibrational corrections (at least at the B3/SVP level) for appropriate comparison with experimental data. For instance, the experimental PCS/bonds) differences between the rotational constants of EA and EAc rotamers are ΔB_a = −62.3 (−61.2) MHz, ΔB_b = 17.3 (16.7) MHz, and ΔB_c = 0.4 (0.3) MHz, with the error being one order of magnitude smaller than that obtained at the MP2 level.[51] This finding is indeed remarkable for assignment purposes since quadrupole coupling constants cannot help in this connection in view of their very similar value in EA and EAc for all the ^{14}N nuclei.[51]

Guanine has 4 endo (N1, N3, N7, N9) and 2 exo (O=C and NH_2) tautomeric sites and 3 labile protons, so that N_S = 6, N_H = 3, and N_T = 20. Among those tautomers, there are 10 amino and

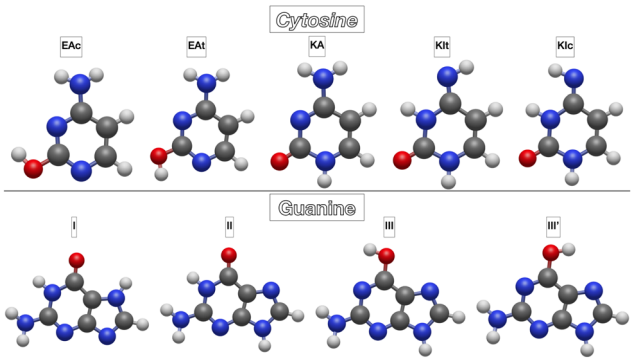

Fig. 8 Most stable tautomers of cytosine and guanine.

10 imino species, but the imino species will not be considered in the following since they are much less stable than their amino counterparts. Two keto-amino and one enol-amino (EA) tautomers are possible for each of the two non-equivalent structures of the imidazole ring (N7H and N9H), with both *trans* and *cis* (labeled with a *c* subscript) conformations of the hydroxyl group corresponding to energy minima for enol-amino tautomers. The different species will be identified by the same two letter code employed for cytosine followed by one (for EA forms) or two (for the KA forms) numbers indicating the positions of the other two acidic hydrogen atoms.

All the methods agree in confirming that only four species (KA17 (I), KA19 (II), EA9 (III) and EA_c 9 (III′)) should have non-negligible populations in the gas phase.[133] The relative electronic energies of these species obtained at the PCS/F12 level (0, 234, 276 and 368 cm^{-1}) are virtually identical (maximum difference of 8 cm^{-1}) to their W1–F12 counterparts.[134] The

Table 7 Comparison between experimental and computed rotational constants (in MHz) for the five tautomers and rotamers of cytosine detected in MW spectra; rDSD stands for rDSD/3F12$^-$ and PCS for PCS/bond

Species	Axis	B^{eq} (rDSD)	ΔB^{B}	ΔB^{vib}	B^0 (PCS)	B^0 (exp.)[a]
EA	*a*	3967.9	11.7	−27.6	3952.0	3951.8
	b	2013.9	5.9	−11.0	2008.8	2009.0
	c	1336.5	4.0	−8.6	1331.9	1332.5
EAc	*a*	3902.8	11.9	−23.9	3890.8	3889.5
	b	2032.2	5.9	−12.6	2025.5	2026.3
	c	1337.1	3.9	−8.8	1332.2	1332.9
KA	*a*	3888.8	12.7	−27.4	3874.1	3871.5
	b	2028.8	6.1	−9.4	2025.5	2025.0
	c	1333.7	4.1	−7.9	1329.9	1330.3
KI	*a*	3865.7	11.9	−29.4	3848.2	3848.2
	b	2030.2	6.2	−11.0	2025.4	2026.3
	c	1331.1	4.1	−7.7	1327.5	1328.0
KIc	*a*	3879.6	12.1	−28.4	3863.3	3861.3
	b	2015.2	6.2	−11.4	2010.0	2011.4
	c	1326.3	4.1	−7.7	1322.7	1323.2
MAX%		0.47			0.07	
MUE%		0.30			0.04	

[a] The experimental ground state rotational constants taken from ref. 51 have been rounded to one decimal place.

main effect of ZPEs is to increase the relative stability of all the other species with respect to the KA17 (I) rotamer by about 80 cm^{-1}. However, scaled harmonic[134] and anharmonic (present work) relative ZPEs show non negligible differences (maximum 60 and average 40 cm^{-1}), with these results confirming the importance of refined vibrational contributions for obtaining accurate thermochemical data. In any case, the computed trends (larger population of KA tautomers with respect to EA tautomers) are in full agreement with their experimental estimates.[133,135]

The computed rotational constants of the four most stable forms of guanine are compared in Table 8 to their experimental counterparts. Already the rDSD/3F12$^-$ results are quite accurate, and correction of bond lengths by the PCS/bond approach further improves the accuracy, which becomes fully quantitative when vibrational corrections are also included. As a matter of fact, the final PCS/bond MUE% and MAX% (0.03% and 0.04%) are comparable with those delivered by the most sophisticated (and much more expensive) wave-function composite methods for small semi-rigid molecules.[115,116]

3.3 Amino acids

Contrary to nucleobases, the low-energy structures contributing to the overall properties of amino-acids do not involve bond pattern changes (with the possible exception of histidine), but require extensive sampling of conformational PESs.

The 'soft' dihedral angles governing the conformational landscape of α-amino acids belong either to the backbone (ϕ' = LP–N–C^α–C′ and ψ = N–C^α–C′–O(H) dihedral angles) or to the side-chain (χ_1 = N–C^α–C^β–X^γ, χ_2 = C^α–C^β–C^γ–Y^δ, *etc.*) where X, Y are generic substituents and LP is the nitrogen lone-pair perpendicular to the plane defined by the two amine hydrogens and the C^α atom. Only planar (or nearly planar) conformations are always allowed for the carboxy moiety (ω = C^α–C′–O–H ≈ 0° or 180°), with $\omega \approx 0°$ being preferred, unless the hydroxyl hydrogen is involved in strong hydrogen bonds with other electronegative atoms. The *c, g, s, t* labels are used to indicate the *cis*, *gauche*, skew, and *trans* conformations determined by the above dihedral angles in the following order: ϕ, ψ, χ_1, χ_2, … In the case of proline, the ψ and ω dihedral angles retain the same definitions, whereas the puckering of the pyrrolidine ring can be described by two pseudorotation coordinates, the puckering amplitude (α) and the phase angle (τ).[42,43]

A systematic exploration of the conformational landscape of α-amino acids by the integrated strategy sketched in Section 2.1 provides several low-energy structures, most of which are stabilized by hydrogen bonds of type I (bifurcated $NH_2\cdots O{=}C$, $\phi' \approx 180°$, $\psi \approx 180°$, $\omega \approx 180°$, ttt) and II (N$\cdots$HO, $\phi' \approx 0°$, $\psi \approx 0°$, $\omega \approx 0°$, ccc).[16] Low-energy conformers of type III ($\phi' \approx 180°$, $\psi \approx 0°$, $\omega \approx 180°$, tct) have also been found, but they usually relax to more stable I conformers overcoming the very small energy barriers governing rotation around the ψ dihedral angle.

In the following, one very rigid (cyclopropylglycine, Ac3c) and one extremely flexible (proline, Pro) amino acid will be analyzed in some detail, together with 4-fluorothreonine (4FT), which well illustrates the need (and power) for integrated computational strategies.

In the case of cyclopropylglycine (see Fig. 9) the conjugation between the three-membered ring and the carbonyl substituent strongly hinders any rotation around the ψ dihedral angle outside planar conformations. As a consequence, conformers I, II and III have C_s symmetry and lie in quite deep energy wells making Ac3c an ideal target for obtaining accurate structural parameters and vibrational corrections to rotational constants. In fact, the error bar of the PCS/bond rotational constants

Table 8 Comparison between experimental and computed rotational constants (in MHz) for the four tautomers and rotamers of guanine detected in MW spectra; rDSD stands for rDSD/3F12$^-$ and PCS for PCS/bond

Species	Axis	B^{eq} (rDSD)	ΔB^B	ΔB^{vib}	B^0 (PCS)	B^0 (exp.)[a]
KA17 (I)	*a*	1927.4	6.7	−11.6	1922.5	1922.2
	b	1124.6	4.2	−6.7	1122.1	1121.7
	c	710.8	2.5	−4.2	709.1	709.0
KA19 (II)	*a*	1927.7	6.7	−12.0	1922.4	1922.3
	b	1119.3	4.1	−6.5	1116.9	1116.7
	c	708.6	2.6	−4.2	707.0	706.7
EA9 (III)	*a*	1923.4	6.6	−13.3	1916.7	1916.1
	b	1134.8	4.2	−6.0	1132.9	1132.4
	c	713.9	2.6	−4.2	712.3	712.2
EA_c9 (III′)	*a*	1931.2	6.6	−13.8	1924.0	1923.5
	b	1138.6	4.2	−6.1	1136.7	1136.0
	c	716.5	2.6	−4.3	714.8	714.7
MAX%		0.40			0.04	
MUE%		0.27			0.03	

[a] The experimental ground state rotational constants taken from ref. 133 have been rounded to one decimal place.

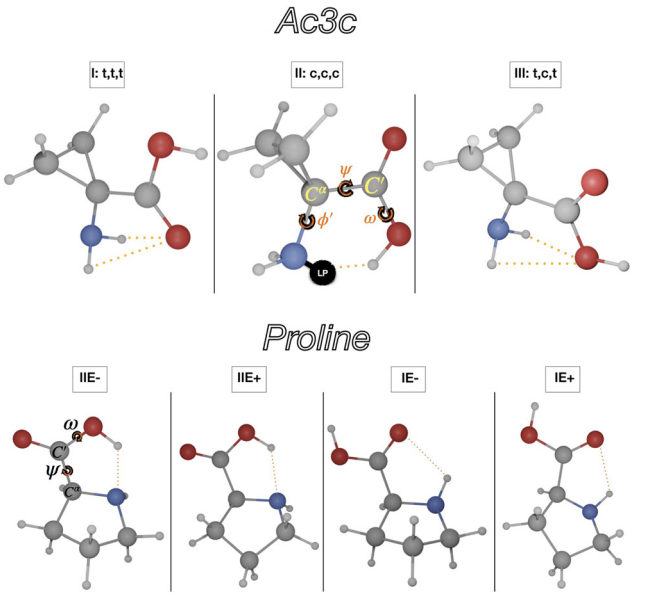

Fig. 9 Most of the stable conformers of cyclopropylglycine (Ac3c) and proline.

reported in Table 9 is much lower than the difference between the rotational constants of the various conformers, with this allowing a full *a priori* assignment of experimental spectra.

In the case of aminoisubutyric acid (Aib, see Fig. 4) the presence of two methyl groups at C^{α} should again hinder the relaxation of conformer III to conformer I, and this expectation is confirmed by QC computations. However, the lack of conjugative contributions results in a larger degree of freedom to the ψ dihedral angle around the fully planar arrangement. As a result, conformer I retains C_s symmetry, but this is not the case for conformers II and III, in which ψ becomes -15° and -56°, respectively. Actually, the latter structure is better indicated as III′ since the bifurcated hydrogen bridge of the reference structure becomes a single hydrogen bond. On the other hand, the very small barrier connecting the two equivalent non-planar conformers of type II leads to a much larger error of the computed rotational constants due to the difficulty in describing vibrational corrections for large amplitude motions.

The results collected in Table 9 confirm the accuracy of rDSD/3F12$^-$ relative energies and show the expected reversal of the relative stability of conformers II and III, with inclusion of anharmonic ZPEs leading to an essentially equal stability of conformers I and II in Aib.

Replacement of both methyl groups with hydrogen atoms leads to glycine (see Fig. 4), which has a much larger conformational freedom due to the negligible steric hindrance of the substituents at C^{α}. As a result, conformer III has not been detected in MW studies (due to its easy relaxation to the most stable I counterpart) and conformer II loses the C_s symmetry again. As already anticipated, the other consequence of the increased flexibility is a worsening of the agreement between computed and experimental rotational constants, partially due to the reduced accuracy of vibrational corrections computed in the framework of the VPT2 model (*cf.* Tables 3 and 9). However, even under those circumstances, PCS/bond computations are largely sufficient to guide and interpret the experimental results.

The next prototypical system is proline (see Fig. 9), whose conformational PES shows several low-energy conformers of type I, II, and III. However, all the species of type III are too unstable to be detected in MW studies, whereas four structures (two of type I and two of type II) are found in the range of 600 cm^{-1}. All those species show envelope structures of the pyrrolidine ring with either *exo*- or *endo*-like placements of the carboxy moiety (referred to as E^+ and E^-, respectively) and have been detected in the MW spectra.[138,139] PCS/F12 computations show that the species of type II are significantly more stable than their counterparts of type I, with ZPE contributions slightly reducing the difference (see Table 10). As expected, the presence of the large-amplitude puckering of the pyrrolidine ring increases the errors of the computed rotational constants with respect to those obtained for Ac3c, Aib, and, even Gly (*cf.* Tables 3, 9 and 10). In particular, the B_a rotational constants are significantly underestimated (in some cases by more than 1%), whereas the B_b and B_c rotational constants are overestimated by comparable amounts. Furthermore, this is the first molecule, whose rDSD/3F12$^-$ equilibrium rotational constants and their PCS/bond ground state counterparts show comparable errors with respect to the experiment. However, even in those circumstances, the agreement between computed and experimental results remains sufficiently good to permit the unequivocal assignment of the detected species, which is further confirmed by quadrupolar coupling constants.[140]

Fluorinated α-amino-acids have been attracting increasing attention because their introduction in specific domains can be used for tuning the stability, folding and biological activity of proteins.[141] In this framework 4FT plays a central role since it is the only fluorinated amino acid found in nature.[142] The identification of the most stable conformers of 4FT, potentially observable in rotational spectroscopy experiments, started from a knowledge-based step in which 21 conformers were generated

Table 9 Rotational constants (in MHz), relative energies, ΔE, and relative enthalpies at 0 K, ΔH_0° (in cm^{-1}) for the conformers of Ac3c and Aib detected in rotational spectra; rDSD stands for rDSD/3F12$^-$, whereas PCS stands for PCS/bond in the case of rotational constants and for PCS/F12 in the case of energies

Conformer	Param.	B^{eq} (rDSD)	ΔB^{B}	ΔB^{vib}	B^0 (PCS)	B^0 (exp.)[a]
Ac3c						
I(ttt)	*a*	3993.1	15.3	−33.3	3975.1	3973.8
	b	2664.9	9.7	−22.9	2651.7	2656.3
	c	1843.5	6.9	−14.3	1836.1	1838.4
II(ccc)	ΔE	269			246	
	ΔH_0°	325			313	
	a	3965.6	14.6	−26.4	3953.8	3951.8
	b	2691.1	10.0	−27.6	2673.5	2678.6
	c	1850.9	6.9	−15.0	1842.8	1844.9
III(tct)	ΔE	133			138	
	ΔH_0°	150			155	
	a	4043.4	14.9	−31.3	4027.0	4026.9
	b	2628.5	10.3	−24.1	2614.7	2618.6
	c	1836.1	6.8	−14.1	1828.8	1831.1
MAX%		0.49			0.39	
MUE%		0.37			0.14	
Aib						
I(ttt)	*a*	3401.9	12.7	−32.1	3382.5	3383.8
	b	2382.4	9.0	−23.0	2368.4	2372.0
	c	2019.3	7.6	−13.7	2013.2	2015.3
II(ccc)	ΔE	−113			−118	
	ΔH_0°	−3			−8	
	a	3350.2	12.5	−32.9	3329.8	3330.2
	b	2442.4	9.2	−25.8	2425.8	2415.4
	c	2032.0	7.8	−15.0	2024.8	2020.4
III′(tgt)	ΔE	264			269	
	ΔH_0°	291			296	
	a	3402.2	12.6	−31.8	3383.0	3385.2
	b	2403.6	9.0	−24.1	2388.5	2390.7
	c	2004.2	7.6	−13.6	1998.2	2000.7
MAX%		1.12			0.43	
MUE%		0.52			0.14	

[a] The experimental rotational constants for Ac3c and Aib are taken from ref. 136 and 137, respectively, and have been rounded to one decimal place.

Table 10 Rotational constants (in MHz), relative energies, ΔE, and relative enthalpies at 0 K, ΔH_0° (in cm^{-1}) for the conformers detected in the rotational spectrum of proline; rDSD stands for rDSD/3F12$^-$, whereas PCS stands for PCS/bond in the case of rotational constants and for PCS/F12 in the case of energies

Conformer	Param.	B^{eq} (rDSD)	ΔB^B	ΔB^{vib}	B^0 (PCS)	B^0 (exp.)[a]
IIE$^-$	*a*	3718.4	15.1	−46.9	3686.6	3673.9
	b	1681.8	5.6	−7.1	1680.3	1688.4
	c	1404.8	4.6	−6.0	1403.4	1407.4
IIE$^+$	ΔE	215			200	
	ΔH_0°	210			195	
	a	3992.8	16.1	−46.6	3962.3	3923.6
	b	1592.7	5.3	−13.6	1584.4	1605.9
	c	1271.7	4.3	−9.7	1266.3	1279.8
IE$^-$	ΔE	607			580	
	ΔH_0°	462			435	
	a	3913.3	16.0	−44.3	3885.0	3857.2
	b	1577.1	5.2	−16.5	1565.8	1590.5
	c	1372.2	4.5	−14.7	1362.0	1377.5
IE$^+$	ΔE	591			561	
	ΔH_0°	474			444	
	a	4032.3	16.7	−44.6	4004.4	4004.0
	b	1568.5	5.2	−13.6	1560.1	1567.3
	c	1284.3	4.2	−12.9	1275.6	1281.5
MAX%		1.76			1.55	
MUE%		0.72			0.74	

[a] The experimental rotational constants are taken from ref. 139 and have been rounded to one and two decimal places, respectively.

by fluorination of the terminal methyl group in the 7 conformers detected for threonine (Thr).[118,143] Next, additional low-energy conformers were detected using the IM-EA algorithm sketched in Section 2.1. The final panel of 12 conformers lying in the range of 750 cm^{-1} was submitted to full geometry optimizations at the B3/SVP and next rDSD/3F12$^-$ levels. Refinement of the relative electronic energies at the PCS/F12 level does not provide any significant change at least concerning the general trends, whereas inclusion of ZPE (leading to relative enthalpies at 0 K) induces significant differences in the stability order predicted by electronic energies, with a general destabilization of type II conformers.

In summary, 12 conformers should be detectable in the rotational spectra (see Fig. 10). Characterization of such a large number of conformers in rotational spectra is not an easy task, but the very accurate rotational constants obtained at the PCS/bond level provide an invaluable aid to the assignment of the very congested spectrum, with the results summarized in Table 11 pointing once again to a good agreement between theory and experiment.

Systematic studies of other α-amino acids based on the nano-LEGO[57,58,118] or PCS strategies confirmed that the backbone of species containing simple non-polar side-chains (*e.g.*, alanine or valine) shows the same low-energy conformers of type I and II found for glycine, albeit with the reduced sym-

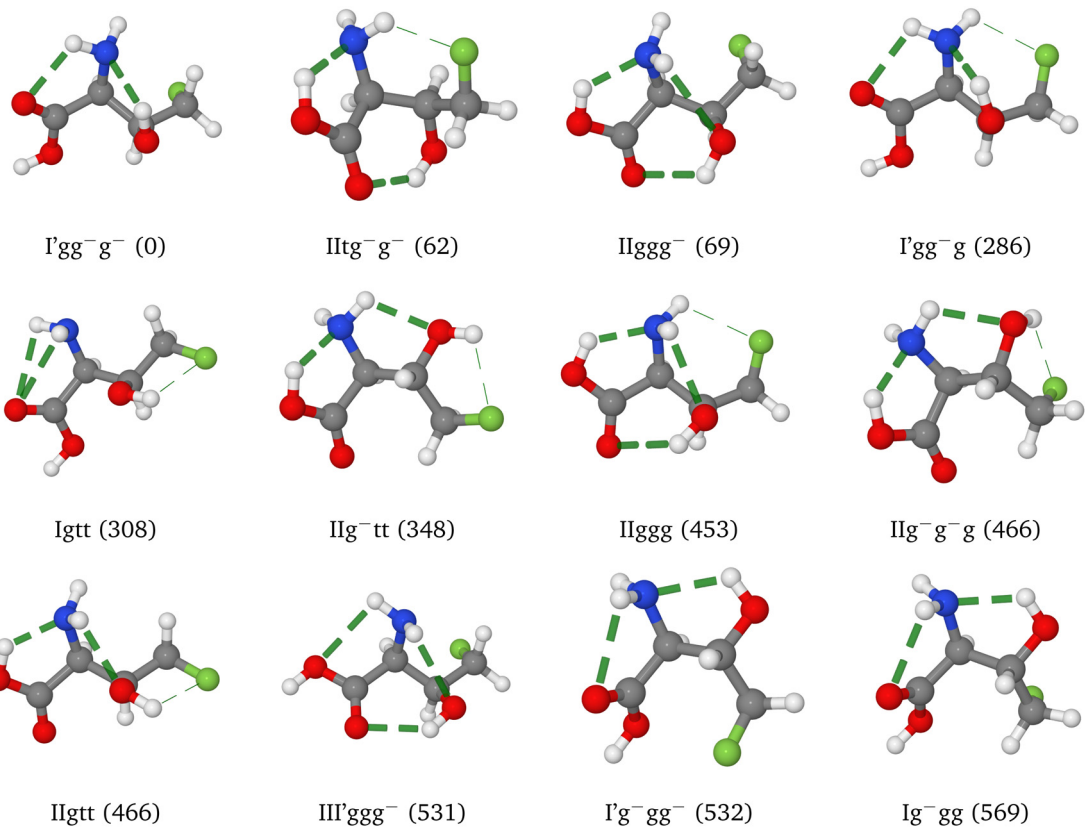

Fig. 10 Structures and intra-molecular hydrogen-bonds of the final panel of 12 low-energy conformers of 4FT. Hydrogen bonds involving N or O atoms are made evident by thick broken lines, whereas those involving F atoms by thin broken lines. The relative enthalpies at 0 K are given in parenthesis.

Table 11 Rotational constants (in MHz) for the detected conformers of 4FT; rDSD stands for rDSD/3F12$^-$ and PCS for PCS/bond

Conformer	Param.	B^{eq} (rDSD)	ΔB^{B}	ΔB^{vib}	B^0 (PCS)	B^0 (exp.)[a]
I′gg$^-$g$^-$	*a*	2498.2	14.0	−21.3	2490.9	2486.9
	b	1034.5	4.6	−9.3	1029.8	1034.1
	c	994.8	4.2	−8.5	990.5	993.0
IItg$^-$g$^-$	*a*	1912.2	12.2	−13.5	1910.9	1906.1
	b	1433.4	4.8	−14.0	1424.2	1434.0
	c	1121.8	5.2	−11.6	1115.4	1120.4
IIggg$^-$	*a*	2611.7	14.0	−24.6	2601.1	2597.9
	b	1074.4	4.7	−11.0	1068.1	1065.8
	c	948.4	4.1	−9.8	942.7	950.2
I′gg$^-$g	*a*	2831.0	14.0	−26.1	2818.9	2822.2
	b	1047.8	4.5	−11.0	1041.3	1045.5
	c	949.9	4.3	−7.6	946.6	948.9
Igtt	*a*	2977.5	13.3	−27.2	2963.6	2966.8
	b	960.0	5.1	−6.6	958.5	959.1
	c	863.2	4.6	−6.0	861.8	861.5
IIg$^-$tt	*a*	2370.0	10.9	−21.4	2359.5	2372.2
	b	1144.0	6.3	−9.7	1140.6	1139.8
	c	850.6	4.3	−7.8	847.1	846.1
IIggg	*a*	2903.8	14.1	−28.2	2889.7	2888.2
	b	1075.9	4.5	−11.9	1068.5	1071.5
	c	918.0	4.1	−6.4	915.7	919.1
IIg$^-$g$^-$g	*a*	2310.1	12.2	−20.1	2302.2	2307.5
	b	1246.5	6.0	−11.7	1240.8	1242.2
	c	924.0	4.0	−9.0	919.0	920.7
IIgtt	*a*	3046.6	12.5	−29.8	3029.3	3036.7
	b	937.1	4.9	−7.3	934.7	935.3
	c	845.4	4.4	−6.3	843.5	843.7
III′ggg$^-$	*a*	2523.3	14.3	−20.1	2517.5	2522.3
	b	1056.4	4.4	−11.2	1049.6	1051.4
	c	932.1	4.3	−7.8	928.6	931.8
Ig$^-$gg$^-$	*a*	1849.1	10.1	−17.4	1841.8	1846.7
	b	1552.7	5.2	−10.5	1547.4	1551.6
	c	997.9	4.2	−9.0	993.1	995.1
Ig$^-$gg	*a*	2077.1	11.1	−15.7	2072.5	2086.4
	b	1364.4	5.3	−14.0	1355.7	1355.8
	c	963.6	3.7	−9.4	957.9	956.6
MAX%		0.81			0.79	
MUE%		0.29			0.25	

[a] The experimental rotational constants taken from ref. 144 have been rounded to one decimal place.

metry related to the presence of a chiral C$^\alpha$ atom.[145,146] On the other hand, polar side-chains (as found, *e.g.*, in serine, cysteine, threonine, aspartic acid, or asparagine) give access to backbone-(side-chain) hydrogen-bonds, with this strongly increasing the number of low-energy conformers.[118] Additional interactions between backbone polar hydrogen atoms and side-chain π-systems are possible for amino acids containing aromatic moieties (like, *e.g.*, phenylalanine and tyrosine).[147] This diversified landscape is further enriched in amino acids showing specific features, like tautomerism (histidine) or heteroaromatic structures with non-equivalent rings (tryptophan).[140]

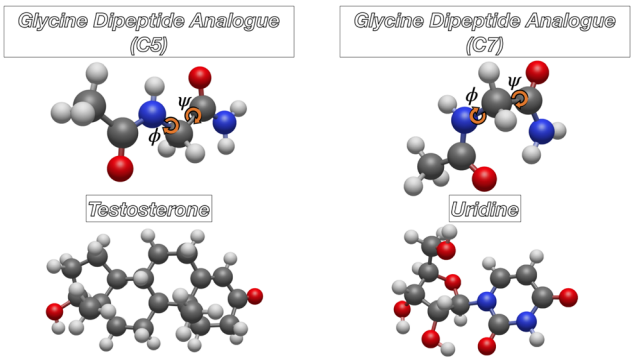

Fig. 11 Glycine dipeptide analogue, uridine nucleoside, and testosterone hormone.

Table 12 Rotational constants (B^0) and vibrational corrections (ΔB_i^{vib}) for dipeptides, nucleosides and hormones. All the values are given in MHz

Axis	ΔB^{vib}[a]	B^0(exp.)[b]	B^0(rDSD/3F12$^-$)	B^0(PCS/bond)
GlyDA C7				
a	−23.9	4421.3	4367.7	4385.6
b	−18.2	1214.2	1205.6	1209.0
c	−13.5	1081.3	1073.1	1075.9
GlyDA C5				
a	−42.3	5268.9	5205.0	5226.2
b	−7.4	1012.0	1003.7	1007.0
c	−5.1	857.2	851.6	854.5
MAX%			1.21	0.81
MUE%			0.89	0.56
AlaDA C7$_{eq}$				
a	−0.9	2598.8	2578.5	2588.8
b	−13.3	1125.7	1117.1	1120.6
c	−11.4	914.5	908.4	911.3
AlaDA C5				
a	−8.4	3100.3	3073.1	3085.1
b	−7.2	955.1	946.0	949.2
c	−7.2	812.7	809.3	812.2
MAX%			0.95	0.62
MUE%			0.74	0.39
Uridine				
a	−9.1	886.0	880.7	883.9
b	−2.9	335.6	334.3	335.1
c	−2.0	270.1	268.7	270.2
MAX%			0.60	0.24
MUE%			0.50	0.14
Testosterone				
a	−8.1	785.3	780.7	783.8
b	−1.6	168.7	168.0	168.6
c	−1.4	153.8	153.2	153.8
MAX%			0.58	0.19
MUE%			0.46	0.08

[a] At the B3/SVP level. [b] The experimental ground state rotational constants (taken from ref. 15, 150, 151 and 152 for glycine dipeptide analogue, alanine dipeptide analogue, uridine, and testosterone, respectively) have been rounded to one decimal place.

3.4 Other molecular bricks

The smallest flexible systems showing the peptide linkage (CO-NH) are the so-called dipeptide analogues, which actually contain two of these moieties because both the amino and carboxyl terminations are replaced by amide groups. Several studies have shown that substituents (*e.g.*, methyl groups) at the N-terminus have a negligible effect, whereas this is not the case for the C-terminus.[148,149] Therefore, CH_3–CO–NH–CHR–CO–NH_2 is the smallest representative model (see Fig. 11).

In order to focus the attention on the backbone, in the following only the glycine and alanine dipeptide analogues are explicitly considered, whose soft degrees of freedom are the ϕ ($C'NC^{\alpha}C'$) and ψ ($NC^{\alpha}C'N$) dihedral angles. In general terms, the conformational flexibility of the dipeptide analogues is smaller than that of the corresponding amino acids, with only the C_5 ($\phi \approx 180^\circ$, $\psi \approx 180^\circ$) and C_7 ($|\phi| \approx 90^\circ$, $|\psi| \approx 60^\circ$, with ϕ and ψ of opposite sign) conformers being populated in the gas phase (see Fig. 11). For chiral residues like alanine, two different situations, namely C_7^{eq} and C_7^{ax} (with eq and ax standing for equatorial and axial, respectively), are possible; however, only the first one is experimentally accessible. Both C_5 and C_7 conformers of glycine[15] and alanine (C_7^{eq})[150] dipeptide analogues have been characterized accurately by rotational spectroscopy with the agreement between PCS/bond and experimental rotational constants being in the expected range for flexible molecules (see Table 12).

Dipeptide analogues containing more complex side chains have been investigated experimentally, but they do not add any new feature to the general trends since the detected conformers reduce to only the C_7^{eq} structure because of the increased backbone rigidity (*e.g.* proline dipeptide analogue[153]) or the formation of additional hydrogen bridges between the backbone and the side chain (*e.g.* serine dipeptide analogue[154]).

Accurate computational studies of nucleosides and hormones have not yet been performed, but the preliminary results for the uridine nucleoside and the testosterone hormone (see Fig. 11) reported in Table 12 suggest that the agreement between PCS/bond and experiment is on par with that obtained for other flexible systems. In the case of uridine, the PCS/bond rotational constants are an order of magnitude more accurate (MAX% and MUE% reduced from 2.4% to 0.24% and 1.9% to 0.15%, respectively) than their MP2 counterparts employed in the original experimental study,[151] while requiring a comparable computational effort. This improvement permits the unbiased assignment of the detected conformer without any need for additional spectroscopic information (*e.g.*, quadrupole couplings) or chemical intuition. As a matter of fact, the rotational constants of the most stable conformer perfectly match the experimental rotational constants (and the same applies to quadrupole couplings). Even lower MAX% (0.19%) and MUE% (0.08%) are obtained for testosterone, which is the largest molecule (49 atoms) considered in the present paper. While there are no ambiguities in the assignment of the rotational spectrum of this molecule,[152] the situation could be different for other hormones (*e.g.*, β-estradiol[155]) involving soft degrees of freedom, which are at present under investigation in our laboratory.

The last family of compounds considered in the present study is that of neurotransmitters, which are chemical compounds released by neurons at the synapse, where they bind to a receptor in order to transmit a signal to the target cell. Tryptamine (Try) and serotonin (5-hydroxytryptamine, Srt) have been chosen for illustrative purposes due to their average dimension and high flexibility (see Fig. 12). The structure of both compounds is closely related to that of the tryptophan amino acid, which has been recently studied in detail.[140] As a

Table 13 Rotational constants (B^0) and vibrational corrections (ΔB_i^{vib}) for neurotransmitters. All the values are given in MHz

Axis	ΔB^{vib} [a]	B^0 (exp.)[b]	B^0 (rDSD/3F12$^-$)	B^0 (PCS/bond)
Tryptamine I (gg$^-$g)				
a	−11.9	1730.2	1722.2	1728.2
b	−4.9	681.9	680.0	681.7
c	−3.8	551.5	549.9	551.3
Tryptamine II (tg$^-$g)				
a	−9.1	1709.4	1704.7	1710.7
b	−5.5	681.9	678.7	680.4
c	−3.8	550.8	548.7	550.1
MAX%			0.47	0.22
MUE%			0.36	0.10
Serotonin I (ggg)				
a	−5.7	1163.2	1157.8	1161.5
b	−7.1	650.6	648.9	650.7
c	−3.2	450.1	448.5	449.7
Serotonin II (gg$^-$g)				
a	−5.6	1286.5	1283.9	1288.2
b	−5.2	571.8	568.8	570.3
c	−3.7	435.6	433.6	434.7
Serotonin III (tg$^-$g)				
a	−5.8	1267.0	1263.8	1268.0
b	−3.5	574.6	573.1	574.6
c	−3.3	436.0	434.1	435.2
MAX%			0.52	0.26
MUE%			0.36	0.12

[a] At the B3/SVP level. [b] The experimental rotational constants (taken from ref. 157 and 156 for tryptamine, and serotonin, respectively) have been rounded to one decimal place.

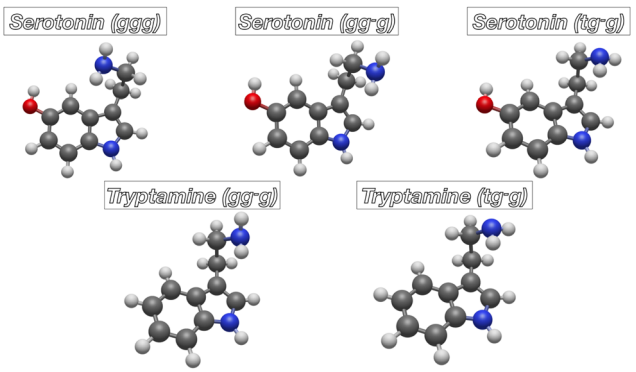

Fig. 12 Molecular structures of prototypical neurotransmitters.

consequence, the same notation employed for amino acids can be extended to the three dihedral angles (ϕ', χ_1, χ_2) ruling the conformational behaviour of the considered neurotransmitters. Of course, ψ and ω dihedral angles are now lacking, whereas the additional dihedral angle involving the hydroxyl group of serotonin is frozen in *syn* or *anti* orientations, with the former one being preferred for all the low-energy conformers.[156]

In analogy with tryptophan, the preferred conformation of the χ_2 dihedral angle (governing the position of the indole ring) is close to 90° (broadly referred to as g) and the more stable structures are characterized by the interaction of one amine hydrogen with the π-system of the phenyl ($\chi_1 \approx 60°$) or pyrrole ($\chi_1 \approx -60°$) ring of indole. These interactions are possible for two orientations of the amine lone pair corresponding to $\phi' \approx 60°$ or 180°. The ggg and gg$^-$g conformers have comparable stability in serotonin (ΔE = 11 cm^{-1} at the rDSD/3F12$^-$ level), whereas only the gg$^-$g conformer has been detected in the rotational spectrum of tryptamine.[157] On the other hand, gg$^-$g and tg$^-$g conformers have been detected for both neurotransmitters, with the gg$^-$g conformer being slightly more stable according to rDSD/3F12$^-$ computations (128 and 120 cm^{-1} for Try and Ser, respectively).

The rotational constants of the conformers detected in MW spectra are collected in Table 13. The remarkable agreement between experimental and PCS/bond values confirm that double-hybrid functionals also deliver extremely accurate structural parameters for flexible systems, but unbiased comparison with experiment requires the inclusion of vibrational corrections and (to a lower extent) core–valence correlation contributions.

4. Conclusions

The main target of the present study was the computation of accurate structures, thermochemical, and spectroscopic parameters of molecular bricks of life in the framework of a general strategy applicable to medium- to large-sized molecules. The main outcome is that the PCS family of methods permits the unbiased computation of accurate spectroscopic parameters for molecules containing a few dozen atoms. Above these dimensions the computation of vibrational corrections can become the bottleneck of the whole procedure and inexpensive approximate approaches have been introduced to deal with this problem.[34] While further work is surely deserved in this connection, even the largest molecule considered in the present paper (testosterone, which has 49 atoms) could be studied without problems employing the VPT2 implementation available in the Gaussian software.

As already mentioned, most molecules of biological interest are highly flexible and the search for stable structures on the corresponding rugged potential energy surfaces has taken great advantage of the development of multi-level QC methods driven by ML algorithms.[27] However, the underlying problem of large amplitude modes limits the accuracy of any perturbative approach to vibrational corrections. The challenges related to this issue depend on the strength of the couplings of the large amplitude modes among themselves and with the other small amplitude modes. Whenever these couplings are small enough (*e.g.*, isolated torsions or ring deformations), the large amplitude modes can be removed from the perturbative treatment and their contribution taken into account by different one-dimensional discrete variable representations.[158] The details and successful applications of this procedure can be found, for example, in ref. 159–161. In this connection, the development of an effective VPT2 engine based on generalized internal coordinates[40] is particularly promising since the inter-mode couplings are much reduced with respect to the traditional implementation in Cartesian coordinates. However, the general case of strongly coupled large amplitude modes calls for new ideas and implementations.

Even taking these remarks into consideration, the reduced computational cost and the user-friendly implementation of the new model pave the way toward widespread application, which could also be used by non-specialists in the assignment and interpretation of experimental results.

Author contributions

V. Barone is the only author.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

The author thanks all his coworkers at the Scuola Normale Superiorew (Luigi Crisci, Silvia Di Grande, Federico Lazzari and Marco Mendolicchio) for the enlightening discussions and their help with several technical issues.

Notes and references

1 V. Barone, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2016, **6**, 86–110.
2 T. Sugiki, N. Kobayashi and T. Fujiwara, *Comput. Struct. Biotechnol. J.*, 2017, **15**, 328–329.
3 C. Puzzarini and V. Barone, *Acc. Chem. Res.*, 2018, **51**, 548–556.
4 J. L. Lane, *Frontiers and Advances in Molecular Spectroscopy*, Elsevier, Amsterdam, 2018.
5 C. Puzzarini, J. Bloino, N. Tasinato and V. Barone, *Chem. Rev.*, 2019, **119**, 8131–8191.
6 Y. He, L. Tang, X. Wu, X. Hou and Y. Lee, *Appl. Spectrosc. Rev.*, 2007, **42**, 119–138.
7 J. Lindon, G. E. Tranter and D. Koppenaal, *Encyclopedia of Spectroscopy and Spectrometry*, Academic Press, Oxford, 2017.

8 M.-A. Martin-Drumel, M. C. McCarthy, D. Patterson, B. A. McGuire and K. N. Crabtree, *J. Chem. Phys.*, 2016, **144**, 124202.
9 M. Melosso, L. Bizzocchi, H. Gazzeh, F. Tonolo, J.-C. Guillemin, S. Alessandrini, V. M. Rivilla, L. Dore, V. Barone and C. Puzzarini, *Chem. Commun.*, 2022, **58**, 2750–2753.
10 V. M. Rivilla, L. Colzi, I. Jiménez-Serra, J. Martn-Pintado, A. Megías, M. Melosso, L. Bizzocchi, A. López-Gallifa, A. Martínez-Henares and S. Massalkhi, *Astrophys. J., Lett.*, 2022, **929**, L11.
11 J. P. Gardner, J. C. Mather, M. Clampin, R. Doyon, M. A. Greenhouse, H. B. Hammel, J. B. Hutchings, P. Jakobsen, S. J. Lilly, K. S. Long, J. I. Lunine, M. J. McCaughrean, M. Mountain, J. Nella, G. H. Rieke, M. J. Rieke, H.-W. Rix, E. P. Smith, G. Sonneborn, M. Stiavelli, H. S. Stockman, R. A. Windhorst and G. S. Wright, *Space Sci. Rev.*, 2006, **123**, 485–606.
12 A. Lesarri, S. Mata, J. C. López and J. L. Alonso, *Rev. Sci. Instrum.*, 2003, **74**, 4799–4804.
13 S. Blanco, A. Lesarri, J. C. López and J. L. Alonso, *J. Am. Chem. Soc.*, 2004, **126**, 11675–11683.
14 M. E. Sanz, S. Blanco, J. C. López and J. L. Alonso, *Angew. Chem., Int. Ed.*, 2008, **47**, 6216–6220.
15 C. Puzzarini, M. Biczysko, V. Barone, L. Largo, I. Peña, C. Cabezas and J. L. Alonso, *J. Phys. Chem. Lett.*, 2014, **5**, 534–540.
16 J. L. Alonso and J. C. López, *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, Springer, 2015, pp. 335–401.
17 I. León, E. R. Alonso, S. Mata, C. Cabezas and J. L. Alonso, *Angew. Chem., Int. Ed.*, 2019, **58**, 16002–16007.
18 G. G. Brown, B. C. Dian, K. O. Douglass, S. M. Geyer, S. T. Shipman and B. H. Pate, *Rev. Sci. Instrum.*, 2008, **79**, 053103.
19 A. L. Steber, J. L. Neill, D. P. Zaleski, B. H. Pate, A. Lesarri, R. G. Bird, V. Vaquero-Vara and D. W. Pratt, *Faraday Discuss.*, 2011, **150**, 227–242.
20 C. Pérez, J. L. Neill, M. T. Muckle, D. P. Zaleski, I. Peña, J. C. López, J. L. Alonso and B. H. Pate, *Angew. Chem., Int. Ed.*, 2015, **54**, 979–982.
21 A. L. Steber, C. Pérez, B. Temelso, G. C. Shields, A. M. Rijs, B. H. Pate, Z. Kisiel and M. Schnell, *J. Phys. Chem. Lett.*, 2017, **8**, 5744–5750.
22 C. Puzzarini, L. Spada, S. Alessandrini and V. Barone, *J. Phys.: Condens. Matter*, 2020, **32**, 343002.
23 F. Xie, M. Fusè, S. Hazrah, A. W. Jaeger, V. Barone and Y. Xu, *Angew. Chem., Int. Ed.*, 2020, **59**, 22427–22430.
24 V. Barone, *Computational Strategies for Spectroscopy: from Small Molecules to Nano Systems*, John Wiley & Sons, Hoboken, 2012.
25 V. Barone, J. Lupi, Z. Salta and N. Tasinato, *J. Chem. Theory Comput.*, 2021, **17**, 4913–4928.
26 G. Mancini, M. Fusè, F. Lazzari, B. Chandramouli and V. Barone, *J. Chem. Phys.*, 2020, **153**, 124110.
27 V. Barone, C. Puzzarini and G. Mancini, *Phys. Chem. Chem. Phys.*, 2021, **23**, 17079–17096.
28 I. León, M. Fusè, E. R. Alonso, S. Mata, G. Mancini, C. Puzzarini, E. R. Alonso and V. Barone, *J. Chem. Phys.*, 2022, **157**, 074107.
29 G. Mancini, M. Fusè, F. Lazzari and V. Barone, *Digital Discovery*, 2022, **1**, 10539–10547.
30 F. Lazzari, A. Salvadori, G. Mancini and V. Barone, *J. Chem. Inf. Model.*, 2020, **60**, 2668–2672.
31 C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
32 D. Ferro-Costas, I. Mosquera-Lois and A. Fernandez-Ramos, *J. Cheminf.*, 2021, **13**, 100.
33 V. Barone, *J. Phys. Chem. Lett.*, 2023, **14**, 5883–5890.
34 V. Barone, *J. Chem. Phys.*, 2023, **159**, 081102.
35 V. Barone, S. Di Grande, F. Lazzari and M. Mendolicchio, *J. Phys. Chem. A*, 2023, **127**, 6771–6778.
36 V. Barone, L. Crisci and S. Di Grande, *J. Chem. Theory Comput.*, 2023, **19**, 7273–7286.
37 V. Barone, *J. Chem. Phys.*, 2005, **122**, 014108.
38 A. M. Rosnik and W. F. Polik, *Mol. Phys.*, 2014, **112**, 261–300.
39 P. R. Franke, J. F. Stanton and G. E. Douberly, *J. Phys. Chem. A*, 2021, **125**, 1301–1324.
40 M. Mendolicchio, J. Bloino and V. Barone, *J. Chem. Theory Comput.*, 2022, **18**, 7603–7619.
41 S. Grimme, *Chem. – Eur. J.*, 2012, **18**, 9955–9964.
42 D. Cremer and J. A. Pople, *J. Am. Chem. Soc.*, 1975, **97**, 1354–1358.
43 L. Paoloni, S. Rampino and V. Barone, *J. Chem. Theory Comput.*, 2019, **15**, 4280–4294.
44 T. Janowski and P. Pulay, *J. Chem. Theory Comput.*, 2008, **4**, 1585–1592.
45 E. Deumens, V. F. Lotrich, A. Perera, M. J. Ponton, B. A. Sanders and R. J. Bartlett, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 895–901.
46 V. M. Anisimov, G. H. Bauer, K. Chadalavada, R. M. Olson, J. W. Glenski, W. T. C. Kramer, E. Aprà and K. Kowalski, *J. Chem. Theory Comput.*, 2014, **10**, 4307–4316.
47 I. A. Kaliman and A. Krylov, *J. Comput. Chem.*, 2017, **38**, 842–853.
48 H. Kruse and J. Šponer, *J. Phys. Chem. A*, 2019, **123**, 9209–9222.
49 L. Gyevi-Nagy, M. Kállay and P. R. Nagy, *J. Chem. Theory Comput.*, 2021, **17**, 860.
50 C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–622.
51 J. l Alonso, V. Vaquero, I. Peña, J. C. Lopez, S. Mata and W. Caminati, *Angew. Chem., Int. Ed.*, 2013, **52**, 2331–2334.
52 S. V. M. Caliebe, P. Pinacho and M. Schnell, *J. Phys. Chem. Lett.*, 2022, **127**, 11913–11917.
53 A. Vats and A. Pathak, *Mon. Not. R. Astron. Soc.*, 2022, **517**, 5780–5790.
54 A. Vats, S. Srivastav, A. Pandey and A. Pathak, *Phys. Chem. Chem. Phys.*, 2023, **25**, 19066–19072.
55 E. R. Alonso, M. Fusé, I. León, C. Puzzarini, J. L. Alonso and V. Barone, *J. Phys. Chem. A*, 2021, **125**, 2121–2129.
56 E. Penocchio, M. Piccardo and V. Barone, *J. Chem. Theory Comput.*, 2015, **11**, 4689–4707.

57 G. Ceselin, V. Barone and N. Tasinato, *J. Chem. Theory Comput.*, 2021, **17**, 7290–7311.
58 V. Barone, G. Ceselin, F. Lazzari and N. Tasinato, *J. Phys. Chem. A*, 2023, **127**, 5183–5192.
59 M. Piccardo, E. Penocchio, C. Puzzarini, M. Biczysko and V. Barone, *J. Phys. Chem. A*, 2015, **119**, 2058–2082.
60 A. Melli, V. Barone and C. Puzzarini, *J. Phys. Chem. A*, 2021, **125**, 2989–2998.
61 V. Barone, *J. Chem. Theory Comput.*, 2023, **19**, 4970–4981.
62 *Evolutionary Computation*, ed. D. B. Fogel, T. Bäck and Z. Michalewicz, Institute of Physics Publishing, Bristol, Philadelphia, 2000.
63 Z. Liu, T. Zubatiuk, A. Roitberg and O. Isayev, *J. Chem. Inf. Model.*, 2022, **62**, 5373–5382.
64 A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100.
65 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
66 G. Santra, N. Sylvetsky and J. M. Martin, *J. Phys. Chem. A*, 2019, **123**, 5129–5143.
67 M. Biczysko, P. Panek, G. Scalmani, J. Bloino and V. Barone, *J. Chem. Theory Comput.*, 2010, **6**, 2115–2125.
68 D. Hait and M. Head-Gordon, *J. Chem. Theory Comput.*, 2018, **14**, 1969–1981.
69 V. Barone, G. Ceselin, M. Fusè and N. Tasinato, *Front. Chem.*, 2020, **8**, 584203.
70 K. Kriz, M. Novacek and J. Rezac, *J. Chem. Theory Comput.*, 2021, **17**, 1548–1561.
71 J. Lupi, S. Alessandrini, C. Puzzarini and V. Barone, *J. Chem. Theory Comput.*, 2021, **17**, 6974–6992.
72 Y. K. Kang and H. S. Park, *Chem. Phys. Lett.*, 2014, **600**, 112–117.
73 M. K. Kesharwani, A. Karton and J. M. Martin, *J. Chem. Theory Comput.*, 2016, **12**, 444–454.
74 Y. K. Kang and H. S. Park, *Chem. Phys. Lett.*, 2018, **702**, 69–75.
75 K. A. Peterson, T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2008, **128**, 084102.
76 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.
77 S. Alessandrini, V. Barone and C. Puzzarini, *J. Chem. Theory Comput.*, 2020, **16**, 988–1006.
78 N. Mehta and J. M. L. Martin, *J. Chem. Theory Comput.*, 2022, **11**, 5978–5991.
79 E. R. Alonso, I. León and J. L. Alonso, *Intra- and Intermolecular Interactions Between Non-Covalently Bonded Species*, Elsevier, 2020, pp. 93–141.
80 R. S. Ruoff, T. D. Klots, T. Emilsson and H. S. Gutowsky, *J. Chem. Phys.*, 1990, **93**, 3142–3150.
81 P. D. Godfrey and R. D. Brown, *J. Am. Chem. Soc.*, 1998, **120**, 10724–10732.
82 G. M. Florio, R. A. Christie, K. D. Jordan and T. S. Zwier, *J. Am. Chem. Soc.*, 2002, **124**, 10236–10247.
83 C. Puzzarini and V. Barone, *Phys. Chem. Chem. Phys.*, 2011, **13**, 7189–7197.
84 S. Di Grande, M. Kàllay and V. Barone, *J. Comput. Chem.*, 2023, **44**, 2149–2157.
85 K. A. Peterson and T. H. Dunning Jr, *J. Chem. Phys.*, 2002, **117**, 10548–10560.
86 H.-J. Werner, T. B. Adler and F. R. Manby, *J. Chem. Phys.*, 2007, **126**, 164102.
87 G. Knizia, T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2009, **130**, 054104.
88 K. E. Yousaf and K. A. Peterson, *J. Chem. Phys.*, 2008, **129**, 184108.
89 K. E. Yousaf and K. A. Peterson, *Chem. Phys. Lett.*, 2009, **476**, 303–307.
90 T. Helgaker, W. Klopper, H. Koch and J. Noga, *J. Chem. Phys.*, 1997, **106**, 9639–9646.
91 K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chem. Phys. Lett.*, 1989, **157**, 479–483.
92 C. Hattig, D. P. Tew and A. Kohn, *J. Chem. Phys.*, 2010, **132**, 231102.
93 M. Kállay, R. A. Horvath, L. Gyevi-Nagy and P. R. Nagy, *J. Chem. Phys.*, 2021, **155**, 034107.
94 A. Karton and J. M. L. Martin, *J. Chem. Phys.*, 2012, **136**, 124114.
95 F. Weigend, *J. Comput. Chem.*, 2008, **29**, 167–175.
96 C. Hattig, D. P. Tew and A. Kohn, *Phys. Chem. Chem. Phys.*, 2005, **7**, 59–66.
97 D. P. Tew and W. Klopper, *J. Chem. Phys.*, 2005, **123**, 074101.
98 M. Kállay, P. R. Nagy, Z. Rolik, D. Mester, G. Samu, J. Csontos, J. Csónka, B. P. Szabó, L. Gyevi-Nagy, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei and B. Hégely, *MRCC, a Quantum Chemical Program Suite*, 2018.
99 M. K. Kesharwani, B. Brauer and J. M. L. Martin, *J. Phys. Chem. A*, 2015, **119**, 1701–1714.
100 B. Chen and L. Radom, *J. Chem. Theory Comput.*, 2016, **12**, 3774–3780.
101 J. C. Zapata Trujillo and L. K. McKemmish, *J. Phys. Chem. A*, 2023, **127**, 1715–1735.
102 M. S. Schuurman, W. D. Allen and H. F. Schaefer III, *J. Comput. Chem.*, 2005, **26**, 1106–1112.
103 J. Bloino, M. Biczysko and V. Barone, *J. Chem. Theory Comput.*, 2012, **8**, 1015–1036.
104 M. Heckert, M. Kállay and J. Gauss, *Mol. Phys.*, 2005, **103**, 2109–2115.
105 M. Heckert, M. Kállay, D. P. Tew, W. Klopper and J. Gauss, *J. Chem. Phys.*, 2006, **125**, 044108.
106 C. Puzzarini, *J. Phys. Chem. A*, 2009, **113**, 14530–14535.
107 H.-J. Werner, P. J. Knowles, F. R. Manby, J. A. Black, K. Doll, A. HeÃŸelmann, D. Kats, A. Köhn, T. Korona, D. A. Kreplin, Q. Ma, T. F. Miller, A. Mitrushchenkov, K. A. Peterson, I. Polyak, G. Rauhut and M. Sibaev, *J. Chem. Phys.*, 2020, **152**, 144107.
108 B. Cordero, V. Gomez, A. E. Platero-Prats, M. Revés, J. Echevarria, E. Cremades, F. Barragan and S. Alvarez, *Dalton Trans.*, 2008, 2832–2838.
109 L. Pauling, *J. Am. Chem. Soc.*, 1947, **69**, 542–553.
110 M. J. Frisch, W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone,

A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16 Revision C.01*, Gaussian Inc., Wallingford CT, 2016.

111 V. Barone and F. Lazzari, *J. Phys. Chem. A*, 2023, **127**, DOI: **10.1021/acs.jpca.3c06649**.

112 J. Demaison, *Mol. Phys.*, 2007, **105**, 3109–3138.

113 M. Mendolicchio, E. Penocchio, D. Licari, N. Tasinato and V. Barone, *J. Chem. Theory Comput.*, 2017, **13**, 3060–3075.

114 Q. Yang, M. Mendolicchio, V. Barone and J. Bloino, *Front. Astron. Space Sci.*, 2021, **8**, 665232.

115 C. Puzzarini and J. F. Stanton, *Phys. Chem. Chem. Phys.*, 2023, **25**, 1421–1429.

116 C. Puzzarini, M. Heckert and J. Gauss, *J. Chem. Phys.*, 2008, **128**, 194108.

117 A. Karton, E. Rabinovich, J. M. L. Martin and B. Ruscic, *J. Chem. Phys.*, 2011, **125**, 144108.

118 V. Barone, M. Fusè, F. Lazzari and G. Mancini, *J. Chem. Theory Comput.*, 2023, **19**, 1243–1260.

119 A. B. Nacsa and G. Csako, *Phys. Chem. Chem. Phys.*, 2021, **23**, 9663–9671.

120 F. J. Lovas, Y. Kawashima, J. U. Grabow, R. D. Suenram, G. T. Fraser and E. Hirota, *Astrophys. J.*, 1995, **455**, L201–L204.

121 I. Léon, N. Tasinato, L. Spada, E. R. Alonso, S. Mata, A. Balbi, C. Puzzarini, J. L. Alonso and V. Barone, *ChemPlusChem*, 2021, **86**, 1374–1386.

122 M. Melosso, S. Alessandrini, L. Spada, A. Melli, X. Wang, Y. Zheng, C. Duan, J. Li, W. Du, Q. Gou, L. Bizzocchi, L. Dore, V. Barone and C. Puzzarini, *Phys. Chem. Chem. Phys.*, 2023, **25**, 31281–31291.

123 L. Sutikdja, D. Jelisavac, W. Stahl and I. Kleiner, *Mol. Phys.*, 2012, **110**, 2883–2893.

124 C. Cabezas, J. L. Alonso, J. C. López and S. Mata, *Angew. Chem., Int. Ed.*, 2012, **51**, 1375–1378.

125 I. Peña, A. M. Daly, C. Cabezas, S. Mata, C. Bermudez, A. Nino, J. C. López, J.-U. Grabow and J. L. Alonso, *J. Phys. Chem. Lett.*, 2013, **4**, 65–69.

126 S. Grimme and M. Steinmetz, *Phys. Chem. Chem. Phys.*, 2013, **15**, 16031–16042.

127 A. Kaczor, I. D. Reva, L. M. Proniewicz and R. Fausto, *J. Phys. Chem. A*, 2006, **110**, 2360–2370.

128 V. Vaquero, M. E. Sanz, J. C. Lopez and J. L. Alonso, *J. Phys. Chem. A*, 2007, **111**, 3443–3445.

129 C. Puzzarini, M. Biczysko, V. Barone, I. Peña, C. Cabezas and J. L. Alonso, *Phys. Chem. Chem. Phys.*, 2013, **15**, 16965–16975.

130 R. D. Brown, P. D. Godfrey, D. McNaughton and A. P. Pierlot, *Chem. Phys. Lett.*, 1988, **156**, 61–63.

131 J. C. Lopez, I. Peña, V. Vaquero, M. E. Sanz and J. L. Alonso, *J. Chem. Phys.*, 2007, **126**, 191103.

132 G. Bazso, G. Tarczay, G. Fogarasi and P. G. Szalay, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6799–6807.

133 J. L. Alonso, I. Peña, J. C. Lopez and V. Vaquero, *Angew. Chem., Int. Ed.*, 2009, **48**, 6141–6143.

134 A. Karton, *J. Phys. Chem. A*, 2019, **123**, 6720–6732.

135 M. Y. Choi and R. E. Miller, *J. Am. Chem. Soc.*, 2006, **128**, 7320–7328.

136 A. I. Jimenéz, V. Vaquero, C. Cabezas, J. C. López, C. Cativela and J. L. Alonso, *J. Am. Chem. Soc.*, 2011, **133**, 10621–10628.

137 J. L. Alonso, I. Peña, J. C. López, E. R. Alonso and V. Vaquero, *Chem. – Eur. J.*, 2019, **25**, 2288–2294.

138 A. Lesarri, S. Mata, E. J. Cocinero, S. Blanco, J. C. López and J. L. Alonso, *Angew. Chem., Int. Ed.*, 2002, **41**, 4673–4676.

139 S. Mata, V. Vaquero, C. Cabezas, C. Pérez, J. C. López and J. L. Alonso, *Phys. Chem. Chem. Phys.*, 2009, **11**, 4141–4144.

140 V. Barone, L. M. Uribe Grajales, S. Di Grande, F. Lazzari and M. Mendolicchio, *J. Phys. Chem. A*, 2023, **127**, 7534–7543.

141 J. Moschner, V. Stulberg, R. Fernandes, S. Huhmann, J. Leppkes and B. Koksch, *Chem. Rev.*, 2019, **119**, 10718–10801.

142 C. D. Murphy, C. Schaffrath and D. OHagan, *Chemosphere*, 2003, **52**, 455–461.

143 J. L. Alonso, M. Pérez, M. E. Sanz and S. Blanco, *Phys. Chem. Chem. Phys.*, 2009, **11**, 617–627.

144 V. Barone, M. Fusè, R. Aguado, S. Potenti, I. Leon, E. R. Alonso, S. Mata, F. Lazzari, G. Mancini, L. Spada, A. Gualandi, P. G. Cozzi, C. Puzzarini, J. L. Alonso and V. Barone, *Chem. – Eur. J.*, 2023, e202203990.

145 V. Barone, S. Di Grande and C. Puzzarini, *Molecules*, 2023, **28**, 913.

146 V. Barone, *Phys. Chem. Chem. Phys.*, 2023, **25**, 22768–22774.

147 V. Barone and M. Fusè, *J. Phys. Chem. A*, 2023, **127**, 3648–3657.

148 W. Yu, X. Xu, H. Li, R. Pang, K. Fang and Z. Lin, *J. Comput. Chem.*, 2009, **30**, 2105–2121.

149 A. Bhattacharya and E. R. Bernstein, *J. Phys. Chem. A*, 2011, **115**, 10679–10688.

150 C. Cabezas, M. Varela, V. Cortijo, A. L. Jiménez, L. Peña, L. Daly, A. M. López, J. C. Cativela and J. L. Alonso, *Phys. Chem. Chem. Phys.*, 2013, **15**, 2580–2585.

151 I. Peña, C. Cabezas and J. L. Alonso, *Angew. Chem., Int. Ed.*, 2015, **54**, 2991–2994.

152 I. León, E. R. Alonso, S. Mata and J. L. Alonso, *J. Phys. Chem. Lett.*, 2021, **12**, 6983–6987.

153 C. Cabezas, M. Varela and J. L. Alonso, *ChemPhysChem*, 2013, **14**, 2539–2543.

154 C. Cabezas, M. A. T. Robben, A. M. Rijs, I. Peña and J. L. Alonso, *Phys. Chem. Chem. Phys.*, 2015, **17**, 20274–20280.

155 S. Zinn and M. Schnell, *ChemPhysChem*, 2018, **19**, 2915–2920.
156 C. Cabezas, M. Varela, I. Peña, J. C. López and J. L. Alonso, *Phys. Chem. Chem. Phys.*, 2012, **14**, 13618–13623.
157 W. Caminati, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2806–2809.
158 D. Skouteris, D. Calderini and V. Barone, *J. Chem. Theory Comput.*, 2016, **12**, 1011–1018.
159 C. Puzzarini, N. Tasinato, J. Bloino, L. Spada and V. Barone, *Phys. Chem. Chem. Phys.*, 2019, **21**, 3615–3625.
160 A. Baiardi, J. Bloino and V. Barone, *J. Chem. Theory Comput.*, 2017, **13**, 2804–2822.
161 G. Ceselin, Z. Salta, J. Bloino, N. Tasinato and V. Barone, *J. Phys. Chem. A*, 2022, **126**, 2373–2387.