ANNUAL REVIEWS

Annual Review of Physical Chemistry

Gas-Phase Computational Spectroscopy: The Challenge of the Molecular Bricks of Life

Vincenzo Barone[1] and Cristina Puzzarini[2]

[1]SMART Laboratory, Scuola Normale Superiore, Pisa, Italy; email: vincenzo.barone@sns.it

[2]Dipartimento di Chimica "Giacomo Ciamician," Universitá di Bologna, Bologna, Italy; email: cristina.puzzarini@unibo.it

Annu. Rev. Phys. Chem. 2023. 74:29–52

First published as a Review in Advance on November 22, 2022

The *Annual Review of Physical Chemistry* is online at physchem.annualreviews.org

https://doi.org/10.1146/annurev-physchem-082720-103845

Keywords

molecular bricks of life, computational spectroscopy, rotational spectra, vibrational spectra, flexibility, anharmonicity

Abstract

Gas-phase molecular spectroscopy is a natural playground for accurate quantum-chemical computations. However, the molecular bricks of life (e.g., DNA bases or amino acids) are challenging systems because of the unfavorable scaling of quantum-chemical models with the molecular size (active electrons) and/or the presence of large-amplitude internal motions. From the theoretical point of view, both aspects prevent the brute-force use of very accurate but very expensive state-of-the-art quantum-chemical methodologies. From the experimental point of view, both features lead to congested gas-phase spectra, whose assignment and interpretation are not at all straightforward. Based on these premises, this review focuses on the current status and perspectives of the fully a priori prediction of the spectral signatures of medium-sized molecules (containing up to two dozen atoms) in the gas phase with special reference to rotational and vibrational spectroscopies of some representative molecular bricks of life.

1. INTRODUCTION

Rotational and vibrational spectroscopies are at the heart of our knowledge of the structural and dynamic properties of molecular systems because of their ability to unveil different physicochemical characteristics in a noninvasive way. In particular, high-resolution molecular spectroscopy in the gas phase (the only field considered in the present review) provides accurate information on intrinsic molecular features without any perturbation from noninnocent environments (1–5). Furthermore, these techniques allow for the unequivocal identification of chemical species in hostile environments, e.g., the interstellar space (6, 7) and planetary atmospheres (8), but also in gaseous samples of unknown composition (4, 9–11).

Recently, rotational spectroscopy has been extended to the investigation of solid compounds by the introduction of the laser ablation technique (12). This technique relies on the use of a laser to vaporize solids without decomposing molecules. Therefore, it is particularly suitable for investigating the building blocks of biomolecules (13–17), especially when coupled with supersonic jet expansion that cools molecules to low rotational temperatures. This permits a great simplification of the resulting rotational spectrum. Furthermore, chirped-pulsed microwave spectrometers (18, 19), which allow for covering a broadband spectral region, combined with fast-mixing nozzles permit the investigation of noncovalent interactions in large systems (20–24). However, for flexible molecules, interpretation of the experimental data often requires accurate computational characterizations of the low-energy minima and their interconversion routes in order to unravel the actual conformers contributing to experimental outcomes. In the same vein, gas-phase infrared (IR) spectra of building blocks of biomolecules have become accessible owing to fast thermal heating of solid samples followed by fast-scan Fourier-transform (FT) IR spectroscopy prior to decomposition (25, 26). At the high temperatures of such experiments, no conformational freezing is possible, thus leading to complicated spectra that require accurate computational predictions for their correct interpretation. Furthermore, a large spectral range from near- to mid-IR is accessible, thus providing high-quality data for benchmarking vibrational calculations (27, 28). The combination of Raman spectroscopy with supersonic expansion provides complementary results (29, 30).

Experimental outcomes have traditionally been considered to be the unquestionable and definitive answers to any chemical problem. However, already for medium-sized systems (in the present context, up to two dozen atoms), such as most molecular bricks of life (see **Figure 1**), direct interpretation of the spectroscopic signals in structural and dynamic terms is not straightforward.

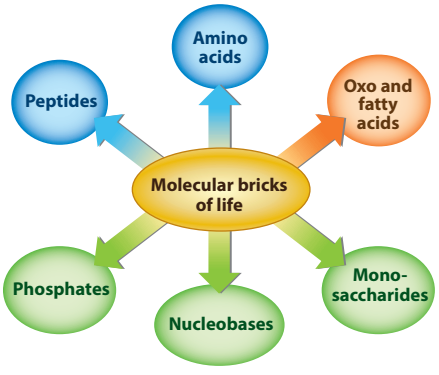

Figure 1

Main families of the molecular bricks of life.

In these cases, although conveying additional information, spectral congestion makes quantitative interpretation of experimental data even more difficult. In this respect, molecular simulations can play an invaluable role, provided that they are able to couple accuracy and feasibility (5, 31, 32).

For small semirigid molecules (in the present context, up to a dozen atoms), the accuracy of state-of-the-art quantum-chemical (QC) methodologies can rival that of experimental techniques (33–37). However, their extension to large (possibly flexible) systems faces a number of difficulties (even within the Born-Oppenheimer approximation), ranging from the very unfavorable scaling of QC methods with the number of active electrons to the correct description of the interactions/effects at work to (possibly) the proper characterization of flat potential energy surfaces (PESs) ruled by a large number of stationary points (5, 38). Focusing on the spectroscopic and molecular properties at the basis of vibrational and rotational spectroscopy, a viable route to obtain accurate results, even for relatively large molecular systems (a few dozens of atoms), is provided by hybrid QC/QC′ models, which combine accurate QC calculations of the primary properties (e.g., molecular structures and harmonic force field) with cheaper yet reliable QC approaches (QC′) for secondary properties (e.g., vibrational corrections and anharmonicity) (see, e.g., 5, 39–41). At the same time, accurate computations of spectroscopic parameters often require purposely tailored basis sets, whose setup and/or choice always requires extensive benchmarks (42–44).

In the field of rotational and vibrational spectroscopy, the rigid rotor (RR)–harmonic oscillator (HO) approximation, which lays the foundation of these spectroscopies and their separability, is not suitable for quantitative purposes, and models incorporating semirigidity and anharmonicity must be employed. Second-order vibrational perturbation theory (VPT2) (45–47) can be considered the gold standard for treatments beyond the RR-HO approximation at affordable costs. Indeed, it has been exploited with considerable success for semirigid molecules of increasing dimensions (5, 48–51). However, concerning vibrational spectroscopy, resonances largely limit VPT2 treatments, and their identification and accurate description still remain daunting tasks due to the arbitrariness of their definition and their indirect influence on the energy and intensities. Note that such interactions also affect rotational spectroscopy (46, 52). It is therefore of crucial importance to be able to systematically find the true resonances and then to correct them appropriately. The resonance conditions are strongly related to the quality of the electronic structure calculations, but they also depend on the coupling between the potentially resonant states. The most obvious solution is to combine perturbative and variational models, with the generalized (G)VPT2 model offering a very good accuracy-to-cost ratio not only for energies but also for transition moments (49). Indeed, GVPT2 incorporates a systematic unsupervised treatment of Fermi and Darling-Dennison resonances, as explained in Section 2.

The focus of this review is on rotational and vibrational spectroscopies for medium-sized molecular bricks of life in the gas phase, which still represent a great challenge for molecular spectroscopy. In view of the different techniques exploited in vibrational and rotational spectroscopy (see above), their synergistic combination guided and supported by computational chemistry in their interpretation paves the route toward the comprehensive study of flexible molecules in the gas phase. From the theoretical point of view, the first issue to address is the selection of suitable QC methods, which should be able to provide an optimal balance between accuracy and feasibility for the properties under consideration (5). Their use for deriving geometric structures and anharmonic force fields of all the stationary points of interest as well as their accuracy has a great impact on the subsequent spectroscopic characterization. Indeed, geometry and force constants are the input of the subsequent (G)VPT2 treatment. However, such a well-defined procedure cannot be applied to flexible systems. Indeed, the exploration of their rugged high-dimensional PESs is already a great challenge that cannot be performed in a standard fashion. Systematic scans of the so-called soft degrees of freedom ruling large-amplitude motions (LAMs) become

extremely expensive for systems involving more than a few LAMs. In these cases, one can resort to algorithms based on artificial intelligence for both the definition of the most suitable set of variables characterizing the PES and its exploration (53).

To address the challenges mentioned in the preceding paragraph, we try to provide a comprehensive picture of an integrated experimental/computational strategy aimed at the structural and spectroscopic characterization of prototypical molecular bricks of life. Starting from semirigid medium-sized molecules, we proceed toward systems with increasing flexibility. Each step is illustrated by representative test cases, which have been mostly selected from recent work performed in the authors' laboratories.

2. GENERAL METHODOLOGY

Even if not addressed in the discussion, the state of the art for small semirigid molecules showing a single well-defined energy minimum is presented in the description of the methodology (1, 5, 52, 54). However, difficulties arise with increasing molecular dimensions. The first difficulty is the increased number of degrees of freedom, which requires the use of effective approximate methods (e.g., hybrid QC/QC′) tested for small systems (5, 39–41, 55). The second difficulty is related to the presence of soft degrees of freedom, which can give rise to LAMs. Furthermore, flat PESs containing several low-lying energy minima separated by (possibly) small energy barriers require the location of all minima and the transition states ruling their interconversion (in order to unravel possible fast relaxation paths). This task, broadly referred to as exploration, can be performed according to different strategies (56–60). We consider systematic searches for molecules characterized by a small number of soft degrees of freedom (61, 62) and, for more flexible systems, the so-called metaheuristics (63), namely iterative generation procedures that guide subordinate heuristics (i.e., deterministic local optimizations) with the aim of escaping from the nearest local minimum (53). The latter topic is analyzed in Section 2.1.2. Subsequently, the so-called exploitation phase, namely the computation of structural, thermochemical, and spectroscopic parameters, is addressed. This phase is common to all types of molecular systems provided that, for flexible molecules, it is repeated for all of the low-lying structures having a nonnegligible population. The general framework of the spectroscopic approach is perturbation theory employing the Watson Hamiltonian (5, 27, 52, 64, 65), which allows a direct link between the computed parameters and those obtained by spectroscopists from the analysis of the recorded spectra (32). To summarize, in the next sections, the QC methodology and the foundations of rotational and vibrational spectroscopy are presented.

2.1. The Quantum-Chemical Model

In the field of rotational and vibrational spectroscopy, the computation of reliable spectroscopic parameters requires the derivation of accurate energies, structures, and anharmonic force fields for each low-lying minimum of the studied molecule. Unfortunately, for the molecular systems of interest in the present context, the state-of-the-art QC composite schemes exploited for the spectroscopic characterization of small semirigid molecules are at present not affordable. Thus, some approximations are unavoidable; however, methodologies that limit the reduction of accuracy have been developed (43, 44). The general approach presented here is to resort to the most accurate composite scheme compatible with the dimension of the system under investigation.

2.1.1. Equilibrium structure. Composite approaches were developed to account for both electron correlation and basis set effects by exploiting the additivity approximation. Among the possible additive schemes available for small semirigid species, in the following we present

the so-called gradient scheme introduced by Gauss and coworkers (66, 67). This is entirely based on coupled-cluster (CC) techniques and is a modular black-box approach that allows for incorporating the desired terms (68). An example is provided here:

$$\frac{dE_{best}}{dx} = \frac{dE^{\infty}(HF-SCF)}{dx} + \frac{d\Delta E^{\infty}(CCSD(T))}{dx} + \frac{d\Delta E(CV)}{dx} + \frac{d\Delta E(fT)}{dx} + \frac{d\Delta E(fQ)}{dx}. \quad 1.$$

While the reader is referred to References 66 and 67 for additional details, in Equation 1, the extrapolation to the complete basis set (CBS) is considered at the CCSD(T) level within the frozen-core (fc) approximation [CCSD(T) standing for the CC ansatz including single, double, and perturbative triple excitations (69)]. Equation 1 also incorporates the core-valence (CV) correlation effects and corrections due to the full treatment of triple (fT) and quadruple (fQ) excitations. To extend the applicability of composite approaches to larger molecules, an effective solution is provided by the so-called geometry scheme. This is based on the assumption that the additivity approximation can be directly applied to geometric parameters and only requires geometry optimizations at several levels of theory. The different contributions are thus evaluated separately and then combined together. To extend the size of systems to those of interest in this review, the computational cost is further reduced by resorting to the so-called cheap geometry scheme (hereinafter ChS) (55):

$$r(ChS) = r(CCSD(T)/TZ) + \Delta r(CBS) + \Delta r(CV), \quad 2.$$

with r denoting a generic structural parameter. In Equation 2, the first term of the right-hand side is the fc-CCSD(T) calculation with a triple-zeta (TZ) quality basis set. The extrapolation to the CBS limit and the CV contribution are instead evaluated using second-order Møller-Plesset (MP) perturbation theory (MP2) (70). Recently, the ChS approach has been extended to incorporate the effects of diffuse functions in the basis set, thus leading to the definition of the junChS model (43, 44). Explicitly correlated F12 methods (71) have also been introduced, thus defining the junChS-F12 model (44, 72). Interestingly, geometries obtained at the CCSD(T)-F12 level (73) with the jun-cc-pVTZ (74) or cc-pVDZ-F12 (75) basis sets are already remarkably accurate without any further extrapolation, provided that CV contributions are incorporated at least at the MP2-F12 level (76) combined with a CV basis set (44, 76, 77). Despite the reduction of the computational cost when moving from a composite scheme entirely based on CC techniques to the ChS model and its variants, for large systems, less expensive levels of theory are needed. In this respect, hybrid [e.g., B3LYP (78) and PW6B95 (79)] and, especially, double-hybrid [e.g., B2PLYP (80) or rev-DSD-PBEP86, hereinafter rDSD (81)] density functionals perform remarkably well when augmented by empirical dispersion corrections [usually using the D3BJ model (82, 83)]. Note that partially augmented double-zeta (DZ) and TZ basis sets [jul-cc-pVDZ and jun-cc-pVTZ (84)] offer the best cost-accuracy compromise for hybrid and double-hybrid density functionals, respectively (50, 85). When these combinations of functional/basis set are used, only the acronym of the functional is indicated explicitly.

The accuracy of computed geometries can be further improved by means of the so-called linear regression approach (LRA) (86, 87). This makes use of the large database of accurate semi-experimental (SE) equilibrium structures (see Section 2.2.1 for definition), which allowed the derivation of reliable linear regressions for key geometric parameters to be used for correcting density functional theory (DFT) systematic errors. The LRA leads to equilibrium geometries rivaling those issuing from the most refined (and costly) QC schemes (86, 87). Another approach that allows for improving DFT structures is the template model approach (TMA) (87, 88), which implies the direct transfer of corrections to geometric parameters from suitable fragments to the molecular system under consideration. The combination of the TMA and the LRA has recently led to the definition of a fully black-box tool, referred to as nano-LEGO (86).

2.1.2. Potential energy surface exploration. The computational protocol for the exploration of flat PESs starts with an optional systematic search based on chemical knowledge (89, 90). This step generates a manageable number of guess structures (smaller than 3^n, where n is the number of degrees of freedom), which are submitted to constrained geometry optimizations using a fast semi-empirical method [typically PM7 (91) or GFN2-xTB (92)]. The application of energy and similarity filters generates the first set of candidates, whose energies are evaluated at the B3LYP or PW6B95 level. The most promising structures resulting from this step are saved for further processing. Then a genetic algorithm is employed to generate other candidates, avoiding any coalescence with those selected in the first step (53). Once again energy and structural filters are employed to select the most promising energy minima, whose geometries are reoptimized at the B3LYP or PW6B95 level. The combination of the candidates selected in these two steps (after further checking for any overlap) defines the panel of structures to be optimized at the B2PLYP or rDSD level. Harmonic frequencies are also computed to verify the nature of the located stationary points and to obtain the free energies (by standard statistical thermodynamic equations) needed for evaluating the relative populations at the temperature of interest. Possible relaxation paths are identified by locating transition states connecting low-energy minima. In some cases, the geometry of key structures is reoptimized at a higher computational level. Finally, spectroscopic parameters of the structures having nonnegligible populations (and not relaxing effectively to more stable energy minima) are computed, as described in Sections 2.2 and 2.3.

2.1.3. Energetics. In a spectroscopic characterization, energetics plays a central role, especially when several conformers are present in the gas-phase mixture. Indeed, as mentioned in Section 2.1.2, free energies allow the derivation of conformer populations, with electronic energy being the key ingredient for such an evaluation.

For small semirigid molecules, a number of highly accurate schemes have been introduced (see, e.g., 93–98). In particular, the HEAT (high accuracy extrapolated ab initio thermochemistry) protocol (96, 99, 100) provides subkilojoule per mole accuracy. Analogously to geometries, when increasing the size of the system, to reduce the computational cost while retaining high accuracy, the ChS approach can be employed. In particular, the junChS and junChS-F12 models have proven to provide kilojoule per mole accuracy (43, 44, 101). Interestingly, these two approaches use reference geometries obtained by double-hybrid density functionals (43, 44).

2.1.4. Harmonic and anharmonic force fields. Composite schemes can also be employed in the evaluation of harmonic force fields (39). Generally, state-of-the-art conventional or explicitly correlated approaches provide very accurate harmonic (35, 36, 102) and anharmonic (51, 103) force fields, but their cost becomes prohibitive for medium- to large-sized molecular systems. For the latter, harmonic frequencies can be obtained from analytical second-derivative techniques (48) applied to double-hybrid functionals with (partially) augmented TZ basis sets. Anharmonic corrections, when needed, are computed using the already mentioned GVPT2 model (28, 104). This requires the evaluation of third and semidiagonal fourth derivatives, which are obtained by the numerical differentiation of second derivatives. Either double-hybrid functionals or, for large systems, hybrid functionals are employed for this task.

2.2. Rotational Spectroscopy

The leading terms in rotational spectroscopy are the rotational constants, which mainly depend on the molecular structure (52, 105). Within the RR approximation, the rotational Hamiltonian is

$$\hat{H}_{\text{rot}} = \frac{1}{2}\left(\frac{\hat{J}_x^2}{I_x} + \frac{\hat{J}_y^2}{I_y} + \frac{\hat{J}_z^2}{I_z}\right), \quad 3.$$

where $\hat{J}_x$, $\hat{J}_y$, and $\hat{J}_z$ are the components of the angular momentum $\hat{J}$, and I_x, I_y, and I_z are the principal moments of inertia (i.e., the diagonal elements of the inertia tensor $\mathbf{I}$ in the principal inertia system). The multiplicative terms of $\hat{J}_x^2$, $\hat{J}_y^2$, and $\hat{J}_z^2$ are the rotational constants in atomic units,

$$B^\alpha = \frac{1}{2I_\alpha}, \quad 4.$$

where α stands for x, y, or z. The x, y, and z axes are those of the principal inertia systems, which are usually denoted as a, b, and c. However, the order of the correspondence is not unambiguous. The particular choice (which can be accomplished in six different ways) defines a representation (I^r, II^r, III^r, I^l, II^l, or III^l) (105). The a, b, c notation also leads to the usual definition of the rotational constants A ($=B^a$), B ($=B^b$), and C ($=B^c$), with the convention that $A \geq B \geq C$. Within the RR approximation, the rotational constants are thus given by the inverse of the principal moments of inertia I_α, which are obtained by diagonalizing the inertia tensor of the molecule, which is given by

$$\mathbf{I} = \sum_K M_K (R_K^2 \mathbf{1} - \mathbf{R}_K \mathbf{R}_K^T); \quad 5.$$

the sum runs over all K nuclei, with $\mathbf{R}_K$ denoting their Cartesian coordinates and M_K their atomic masses.

Going beyond the RR approximation, the rotational constants depend on the vibrational motion. Within the VPT2 treatment, the usual contact-transformation method gives the second-order result (45)

$$B_v^\alpha = B_{\text{e}}^\alpha - \sum_r \alpha_r^\alpha \left(v_r + \frac{1}{2}\right); \quad 6.$$

the sum is taken over all fundamental vibrational modes r, and the superscript α refers to the inertial axes (a, b, and c). B_v^α and B_{e}^α are the rotational constants of the vibrational state v and at equilibrium, respectively, while α_r^α denotes the vibration-rotation interaction constants. Based on the discussion above, the equilibrium rotational constants are completely defined once the equilibrium structure is known. Therefore, from a computational point of view, their determination only requires geometry optimizations. As is evident from Equation 6, the vibrational contribution does not vanish for the vibrational ground state:

$$B_0^\alpha = B_{\text{e}}^\alpha - \frac{1}{2}\sum_r \alpha_r^\alpha = B_{\text{e}}^\alpha + \Delta B_{\text{vib}}^\alpha. \quad 7.$$

From a computational point of view, vibrational corrections require anharmonic force field calculations (52). Since the vibrational corrections are small terms with respect to the equilibrium rotational constants (from less than 1% to 3% at most), in rotational spectroscopy, the computational effort should be put into accurate geometry optimizations, possibly employing the composite schemes introduced in Section 2.1.1.

Finally, going beyond the RR approximation also implies centrifugal distortion effects (64). The lowest-order perturbation treatments lead to the introduction of the quartic and sextic centrifugal distortion constants (47). Computationally, their evaluation needs harmonic and anharmonic force field calculations, respectively (47, 52). Therefore, they can be obtained as by-products of the computations carried out for obtaining vibrational corrections to rotational constants.

Rotational constants are at the basis of the so-called SE approach for the derivation of accurate structures. This is, however, hampered by the limited amount of data with respect to the number of geometric parameters and by the proper account of vibrational effects. The first limitation is overcome by considering, for a given molecule, a sufficient number of isotopic species in order to increase the amount of data (105). Since rotational constants obtained experimentally refer to the vibrational ground state, the structures directly derived from them are vibrationally averaged structures (105). To exclude vibrational effects in a rigorous manner, we can resort to Equation 7, which can be rewritten as

$$B_e^\alpha = B_0^\alpha + \frac{1}{2}\sum_r \alpha_r^\alpha = B_0^\alpha - \Delta B_{vib}^\alpha. \quad 8.$$

If in the equation above, the B_0^α value is obtained experimentally and the ΔB_{vib}^α correction is computed, the resulting B_e^α is the so-called SE equilibrium rotational constant. The availability of SE equilibrium rotational constants for a balanced set of isotopic species leads to the derivation of the SE equilibrium geometry (r_e^{SE}) by means of a least-squares fit. Since the sum in Equation 8 is, contrary to the individual terms, devoid of any possible resonance (106) and, owing to the fact that the vibrational corrections are very small terms with respect to the B_0^α constants, the SE approach provides equilibrium structures with great accuracy, also when the ΔB_{vib}^α contributions are obtained with hybrid density functionals. Additional electronic contributions can also be taken into account (3), but their role is negligible except in very peculiar cases. Using the SE approach, a data set of about 100 SE equilibrium structures for molecules containing H, B, C, N, O, F, P, S, and Cl atoms has been recently set up (86).

2.3. Vibrational Spectroscopy

The VPT2 model permits us to write the energies of vibrational bands in terms of harmonic wave numbers (ω_i) and elements of the effective anharmonicity matrix (χ_{ij}):

$$\nu_i = \omega_i + 2\chi_{ii} + \sum_{r\neq i=1}^{3N-6}\frac{\chi_{ir}}{2}, \quad 9.$$

$$[\nu_i\nu_j] = \omega_i + \omega_j + 2\chi_{ii} + 2\chi_{jj} + 2\chi_{ij} + \frac{1}{2}\sum_{r\neq\{i,j\}=1}^{3N-6}(\chi_{ir} + \chi_{jr}). \quad 10.$$

Both first overtones ($i = j$) and combination bands ($i \neq j$) are described by Equation 10. Diagonal and off-diagonal elements of the χ matrix are evaluated as follows:

$$16\chi_{ii} = \phi_{iiii} - \frac{5\phi_{iii}^2}{3\omega_i} - \sum_{j\neq i=1}^{3N-6}\frac{(8\omega_i^2 - 3\omega_j^2)\phi_{iij}^2}{\omega_j(4\omega_i^2 - \omega_j^2)}, \quad 11.$$

$$4\chi_{ij} = \phi_{iijj} - \frac{2\omega_i\phi_{iij}^2}{(4\omega_i^2 - \omega_j^2)} - \frac{2\omega_j\phi_{ijj}^2}{(4\omega_j^2 - \omega_i^2)} - \frac{\phi_{iii}\phi_{ijj}}{\omega_i} - \frac{\phi_{jjj}\phi_{iij}}{\omega_j}$$
$$+ \sum_{r\neq\{i,j\}=1}^{3N-6}\left[\frac{2\omega_r(\omega_i^2 + \omega_j^2 - \omega_r^2)\phi_{ijr}^2}{\Delta_{ijr}} - \frac{\phi_{iir}\phi_{jjr}}{\omega_r}\right] + \frac{4(\omega_i^2 + \omega_j^2)}{\omega_i\omega_j}\sum_{\alpha=a,b,c}B_e^\alpha\{\zeta_{ij}^\alpha\}. \quad 12.$$

In the equations above, ζ_{ij}^α and B_e^α are the Coriolis constant (coupling normal modes i and j) and the rotational constant associated with the principal rotation axis α, respectively; ϕ_{ijr} and ϕ_{iijj} denote the cubic and semidiagonal quartic force constants, respectively, and Δ_{ijr} is given by

$$\Delta_{ijr} = \omega_i^4 + \omega_j^4 + \omega_r^4 - 2(\omega_i^2\omega_j^2 + \omega_i^2\omega_r^2 + \omega_j^2\omega_r^2). \quad 13.$$

In Equations 11 and 12, nearly singular contributions (resonances) must be removed and taken into account in a successive variational treatment. This model [known as GVPT2 (28) or VPT2+K (107)] can be seen as a perturb then diagonalize approach, which is more effective in the vibrational context than its more traditional diagonalize then perturb counterpart that is employed in most multireference electronic structure models.

Analytical expressions for VPT2 transition moments have been derived for one-, two-, and three-quanta excitations from the vibrational ground state, thus giving access to IR, Raman, vibrational circular dichroism, and Raman optical activity intensities including both mechanical and electrical anharmonicity (49, 51). The corresponding variationally corrected VPT2 properties have also been implemented by means of proper deperturbation followed by projection onto the eigenvectors of the Hamiltonian matrix, including nearly resonant couplings (27, 108). Although the conventional VPT2 approach employs different equations for asymmetric, spherical, symmetric, and linear tops, it has been recently shown that the asymmetric-top formulation can be employed for the other cases, provided that all the degeneracy issues are handled properly and the customary spectroscopic signatures of non-Abelian groups (e.g., ℓ-type doubling) are derived by a posteriori transformations of the eigenvectors (65, 109). An additional advantage of this formulation is that it allows the computation of intensities for non-Abelian symmetry groups without resorting to complex algebra.

For flexible molecules, some degrees of freedom correspond to LAMs, whose perturbative treatment can lead to unphysically huge anharmonic corrections. The challenges related to this issue depend on the strength of the couplings between the LAMs and the other small-amplitude motions. Whenever these couplings are small enough (e.g., isolated torsions or ring deformations), the LAMs can be removed from the VPT2 treatment, and their contribution taken into account by different one-dimensional discrete variable representations (DVRs). Details and successful applications of this type of procedure can be found, for example, in References 110–113.

3. THE MOLECULAR BRICKS OF LIFE

This section is devoted to the spectroscopic characterization of representative examples of the molecular building blocks of life shown in **Figure 1**. We start from the analysis of pyrimidines and purines that are medium-sized semirigid molecules and, thus, suitable for hybrid QC/QC′ approaches. Uracil and thiouracil are chosen as examples for systems with only one minimum. Then, we tackle the problem of different tautomeric forms using guanine as a test case. Next, we move our attention to two species related to sugars, namely the smallest carboxylic acids containing also a ketonic [pyruvic acid (CH_3–CO–COOH)] or alcoholic [glycolic acid (CH_2OH–COOH)] moiety and representing paradigmatic models of small flexible systems with a reduced number of soft degrees of freedom. The situation is similar for dipeptide analogs of amino acids with small side chains, whose conformational landscape is ruled by two dihedral angles. Glycine and alanine dipeptide analogs are the chosen examples. When the number of soft degrees of freedom increases and their coupling is not negligible, preliminary exploration of the corresponding PES becomes mandatory. This aspect is illustrated by different amino acids involving up to six dihedral angles (using glycine, alanine, cysteine, and homocysteine as examples of increasing complexity).

3.1. Pyrimidines and Purines: Uracil, Thiouracil, and Guanine

The investigations discussed in References 114 and 115 demonstrated that composite approaches fully based on CC techniques are able to provide rotational constants with great accuracy (see also 5 and 52). As explained in Section 2, such schemes are not affordable for medium-sized molecules and approximated approaches must be employed. The performance of the latter is presented here

Table 1 Equilibrium rotational constants of some molecular bricks of life

	B_e	B3LYP/ SNSD	B2PLYP/ cc-pVTZ	MP2/ cc-pVTZ	CCSD(T)/ cc-pVTZ	ChS	SE
Uracil	A_e	3,875.9	3,904.5	3,906.8	3,884.1	3,913.9	3,912.4
	B_e	2,010.9	2,020.8	2,018.6	2,015.7	2,039.1	2,035.3
	C_e	1,324.0	1,331.6	1,330.9	1,327.0	1,340.7	1,338.9
Thiouracil	A_e	3,547.4	3,571.9	3,569.7	3,550.5	3,578.6	3,578.3
	B_e	1,299.5	1,309.6	1,313.6	1,306.6	1,322.4	1,321.9
	C_e	951.1	958.3	960.3	955.1	965.6	965.4
Pyruvic acid (Tc)	A_e	5,517.3	5,528.3	5,502.9	5,501.8	5,564.5	5,559.3
	B_e	3,564.5	3,603.0	3,625.6	3,602.4	3,611.4	3,621.5
	C_e	2,194.5	2,210.5	2,214.9	2,206.3	2,219.6	2,222.4
Glycolic acid (sSc)	A_e	10,653.0	10,718.3	10,699.3	10,701.9	10,801.2	10,797.3
	B_e	4,026.8	4,063.1	4,092.9	4,073.0	4,102.9	4,098.0
	C_e	2,976.9	3,000.9	3,015.6	3,005.3	3,029.2	3,025.5

All values are in megahertz. B3LYP results are from Reference 106 and this work; B2PLYP results are from References 88, 112, and this work; MP2, CCSD(T), and ChS results are from References 55, 112, 116, and 117; and SE equilibrium rotational constants are from References 5, 88, 112, and 117.
Abbreviations: ChS, cheap geometry scheme; MP2, second-order Møller-Plesset perturbation theory; SE, semi-experimental.

for uracil and thiouracil. These molecules have been chosen not only for their biological relevance but also because they have been well characterized experimentally (see 55 and 116 and references therein).

In **Table 1**, equilibrium rotational constants of both species evaluated at different levels of theory are compared to the corresponding SE equilibrium values (see Equation 8). The results point out the accuracy of the ChS model introduced above, which, at the price of a marginal increase of computational cost [additional MP2 computations much less expensive than the CCSD(T) one], provides a terrific improvement with respect to the fc-CCSD(T)/cc-pVTZ level of theory. Despite strongly reduced computational costs, the results delivered by the B2PLYP and even the B3LYP computational models are competitive with respect to fc-CCSD(T) calculations performed with a TZ basis set. In Reference 5, figure 9, the impact on the prediction of the rotational spectrum of uracil is evident: B2PLYP/cc-pVTZ agrees better with experiment than does fc-CCSD(T)/cc-pVTZ. Concerning the ChS, a nice graphical representation of its very good performance can be found in Reference 116, figure 1. In passing, we note that the ChS approach and its variants can also be applied to other spectroscopic quantities such as harmonic frequencies, quartic centrifugal distortion constants, and hyperfine parameters (for the definition of the latter, see 52).

Moving to vibrational spectra, in Reference 28, the performance and accuracy of fully anharmonic calculations for medium-sized molecular systems have been investigated. The comparison of different hybrid functionals with partially augmented basis sets pointed out the superiority of B3LYP with a mean absolute error with respect to experimental data of 9 cm^{-1}. Focusing on uracil and thiouracil, the remarkable agreement between experimental and computed (at the B3LYP level) fundamentals (mean absolute error of 10 cm^{-1} for uracil, 5.3 cm^{-1} for thiouracil) confirms the outcome above. Since both molecules show Fermi resonances (39, 118), this result points out that the GVPT2 approach is able to deal satisfactorily with strong resonance effects in semirigid molecules (28, 118). Improved results are obtained by means of hybrid QC/QC′ computations in which anharmonic B3LYP contributions are employed together with more accurate harmonic terms. If B2PLYP/cc-pVTZ or ChS is considered for the latter, absolute mean errors within 10 cm^{-1} are obtained for uracil (39, 119). At these levels, similarly good agreement is also

noted for overtones and combination bands, thus demonstrating that assignment and interpretation of spectroscopic features for these types of molecular systems can be performed confidently with effective approaches going beyond the standard HO model.

In the gas phase, uracil, thiouracil, and thymine are present only in their diketo forms (55, 116, 120), whereas several tautomeric forms have been detected for cytosine and guanine (121, 122) due to the close stability of keto and enol forms (hereinafter K and E, respectively). In guanine, two nonequivalent structures of the imidazole ring (N7H and N9H) have comparable energy and further increase the number of detectable species. In fact, theory and experiment agree in predicting four low-energy tautomers (K-N7H and K-N9H, together with the *cis* and *trans* forms of E-N9H) (see References 121 and 123–125 for data and a detailed explanation of the nomenclature). Accurate W1-F12 computations (95) point out that the four tautomers lie close in energy (enthalpy at 298 K): 0.0 (K-N7H), 2.2 (K-N9H), 2.5 (*cis*-E-N9H), and 4.1 (*trans*-E-N9H) kJ mol^{-1} (125). Therefore, they all should be considered in the prediction of rotational and vibrational spectra, with an accurate estimate of relative populations being a crucial point. Several higher energy tautomers have also been predicted, but their relative stability (at least 15 kJ mol^{-1} above the most stable species) hampers any experimental detection.

3.2. Sugar-Related Molecules: Pyruvic and Glycolic Acids

Both pyruvic and glycolic acids can be produced from glucose and supply energy to cells through the Krebs (in animals) or glyoxylate (in plants, bacteria, and fungi) cycle, respectively. Both acids have been recently characterized using the ChS model for geometries and harmonic force fields combined with anharmonic contributions at the B2PLYP level with an augmented TZ basis set (112, 117).

The conformational landscape of pyruvic acid is ruled by the O=C–C=O and H–O–C–C dihedral angles, whose *cis* and *trans* conformations are labeled as C, T and c, t, respectively. However, only three conformers (Tc, Tt, and Ct, in order of decreasing stability) have been located on the PES. For the most stable one (Tc), owing to the spectroscopic (rotational) characterization of different isotopic species, it has been possible to obtain the SE equilibrium structure (117). This accurate geometry could be employed together with the TMA to derive geometric parameters of the other two conformers with an accuracy approaching that of the r_e^{SE}, i.e., better than 0.001 Å and 0.1° for bond lengths and valence angles, respectively. This strategy allowed for greatly improving the prediction of rotational spectra of the Tt and Ct conformers. In passing, we note that the results of **Table 1** lead to conclusions similar to those pointed out for uracil and thiouracil.

Moving to IR spectroscopy, experimental data are available for the Tc and Tt conformers, thus permitting us to verify the accuracy reached by the QC/QC′ model, with QC and QC′ being the ChS and B2PLYP level, respectively. The computed vibrational frequencies were found to agree with their experimental counterparts well within 10 cm^{-1}, on average. The investigation reported in Reference 117 allows us to point out the importance of predicting not only accurate vibrational frequencies but also reliable IR intensities. Indeed, the computation of the latter led to simulated spectra that effectively reproduce all main features in the mid- and near-IR spectral ranges, thus allowing us to disentangle the contributions of low-intensity fundamentals from those of high-intensity overtones and combination bands (117). A noteworthy outcome of the work presented in Reference 117 is the remarkable accuracy of the B2PLYP functional, which can be further improved by using the TMA for geometric parameters and a single scaling factor for the harmonic part of the OH stretching. These findings pave the way toward the accurate study of larger acids and/or carbohydrate moieties at the limited cost of DFT computations.

The conformational landscape of glycolic acid is ruled by three dihedral angles: H–O–C–C (with *syn*, *anti*, and *gauche* minima), O–C–C=O (with *syn* and *anti* minima), and O=C–C–H

(with *cis* and *trans* minima). The resulting conformers are labeled with three letters denoting the conformation of each dihedral angle (e.g., sSc), with only 7 out of 12 being, however, true energy minima (see 112 and references therein). These are sSc, gAc, aAt, aSc, aAc, sSt, and aSt, in order of decreasing stability (see 112, figure 1). Characterization of the energetics is not sufficient to understand which conformers can be experimentally observed, but interconversion paths are also needed. According to Reference 112, the energy barriers leading from aSt to sSt, from aSc to sSc, and from aAc to gAc are extremely low (below 1 kJ mol^{-1}), thus preventing the detection of the aSt, aSc, and aAc conformers. The remaining energy barriers are instead much higher, ranging from 15 to more than 40 kJ mol^{-1}. As a consequence, only the sSc, gAc, aAt, and sSt conformers should be amenable to experimental investigations, and indeed, vibrational spectra have been recorded for these four species either in the gas phase or in inert matrices (126). At variance, only two conformers have been characterized by microwave spectroscopy, possibly due to the high energy of sSt and the very small dipole moment components (which imply very weak rotational transitions) of gAc (127, 128). In **Table 1**, the equilibrium rotational constants of the most stable conformer (sSc) evaluated at different levels of theory are compared with the SE equilibrium counterparts, derived as explained in Equation 8. It is quite apparent that glycolic acid, despite its increased flexibility with respect to pyruvic acid, confirms the effectiveness of the ChS approach, with its accuracy not being degraded compared to that of semirigid systems. The availability of rotational constants for several isotopologs of the sSc conformer (128) allowed the determination of an accurate SE equilibrium structure, which could allow the application of the TMA methodology to similar compounds (112). Note that all the other spectroscopic parameters are accurately reproduced by a GVPT2 treatment based on ChS harmonic and B2PLYP anharmonic force fields.

For the most stable conformer, GVPT2 computations combined with one-dimensional DVR treatment of the torsion around the C–C bond led to remarkable agreement between computed and experimental vibrational spectra recorded in the gas phase (112, 126). In particular, the mean absolute error for fundamental bands is as small as 4.4 cm^{-1}; furthermore, the small splittings between the OH stretchings are reproduced accurately. For the other conformers, the agreement is less satisfactory for some vibrations, possibly due to the role played by matrix effects. In conclusion, pyruvic and glycolic acids demonstrate that accurate predictions of the rotational and vibrational spectroscopic parameters can also be obtained for molecules characterized by some degree of flexibility, with the ChS approach performing well for equilibrium structures and harmonic force fields.

3.3. Peptides: Glycine and Alanine Dipeptide Analogs

Besides simple amides, the smallest realistic systems showing the peptide linkage (CO–NH) are the so-called dipeptide analogs, 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., a methyl group) at the N terminus have a negligible effect, whereas this is not the case for those at the C terminus (129, 130). Consequently, CH_3–CO–NH–CHR–CO–NH_2 is the smallest representative model (see **Figure 2*a***).

While the effect of side-chain flexibility is investigated in Section 3.4, here the attention is focused on the system backbone. For this reason, innocent side chains (H, CH_3) are considered. In such circumstances, the conformational flexibility of dipeptide analogs is smaller than that of the corresponding amino acids, with only the C_5 ($\phi \approx 180°$, $\psi \approx 180°$) and C_7 ($|\phi| \approx 90°$, $|\psi| \approx 60°$, with ϕ and ψ of opposite sign) conformers being populated in the gas phase. For chiral residues, two different situations, namely C_7^{eq} and C_7^{ax} (eq and ax stand for equatorial and axial, respectively), are possible; however, only the first one is experimentally accessible.

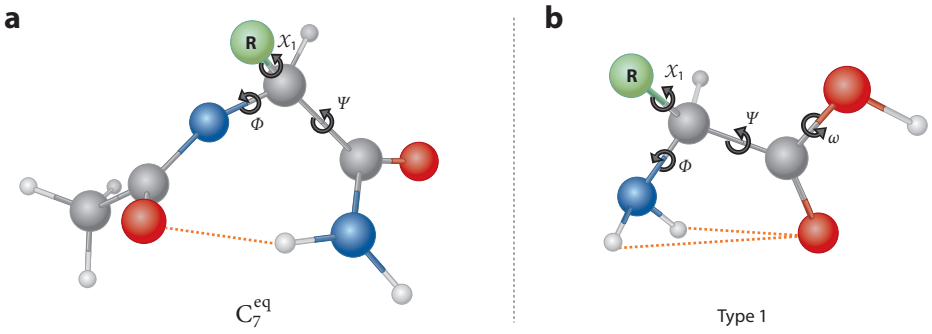

Figure 2

The structure and main dihedral angles of the most stable conformer of (*a*) a dipeptide analog and (*b*) an amino acid with so-called innocent side chains. Dihedral angles are ϕ = LP–N–C^{α}–C′ (LP is the nitrogen lone pair), ψ = N–C^{α}–C′–XH with X = N or O, ω = C^{α}–C′–O–H, and χ_1 is the side-chain torsional angle.

Both the C_5 and C_7 conformers of glycine (15) and alanine (C_7^{eq}) (131) dipeptide analogs have been accurately characterized by rotational spectroscopy guided by QC computations. While the flexibility appears to be reduced here with respect to the systems of Section 3.2, the molecular size increases, thus challenging accurate QC computations.

For the glycine dipeptide analog, rotational constants were evaluated by exploiting the ChS approach for equilibrium and B3LYP for incorporating vibrational corrections (15) (see **Table 2**). This approach led to agreement between experimental and computed values within a few megahertz (the only exception being the *A* constant of the C_5 conformer, which is, however, not well determined by experiment). On the contrary, for the alanine dipeptide analog, the mean and maximum absolute errors are 32 and 94 MHz, respectively (131). Such a worsening is ascribable to both the level of theory [MP2/6-311++G(d,p)] and the lack of vibrational corrections (131). This example allows us to draw attention to the fact that a fully a priori disentanglement of the conformational bath of more flexible systems (see Section 3.4) cannot rely on standard approaches

Table 2 Equilibrium rotational constants of glycine and its dipeptide analog

	B_e	B3LYP/ SNSD	B2PLYP/ cc-pVTZ	MP2/ cc-pVTZ	CCSD(T)/ cc-pVTZ	ChS	SE
Glycine dipeptide analog (C_7)	A_e	4,453.3	4,397.6	4,369.5	4,417.3	4,456.6	4,445.2
	B_e	1,199.2	1,225.5	1,236.1	1,220.2	1,228.9	1,232.5
	C_e	1,061.1	1,086.7	1,097.6	1,087.1	1,096.5	1,094.8
Glycine dipeptide analog (C_5)	A_e	5,206.8	5,243.8	5,229.2	5,221.5	5,281.2	5,311.2
	B_e	998.5	1,011.9	1,017.7	1,011.7	1,016.8	1,016.0
	C_e	846.7	857.1	860.9	856.5	861.7	862.3
Glycine (I)	A_e	10,283.1	10,364.8	10,328.0	10,328.2	10,396.6	10,418.2
	B_e	3,831.1	3,883.3	3,905.0	3,884.4	3,901.1	3,906.9
	C_e	2,882.9	2,917.1	2,926.2	2,915.0	2,930.4	2,934.4
Glycine (II)	A_e	10,135.0	10,174.9	10,178.7	10,112.3	10,205.3	10,144.5
	B_e	4,043.4	4,086.7	4,104.7	4,083.8	4,095.6	4,094.5
	C_e	2,993.1	3,026.3	3,041.0	3,025.7	3,030.6	3,024.7

All values are in megahertz. B3LYP results are from Reference 106 and this work; B2PLYP results are from Reference 88 and this work; MP2, CCSD(T), and ChS results are from References 15 and 41; and SE equilibrium rotational constants are derived from data from References 15 and 41. Abbreviations: ChS, cheap geometry scheme; MP2, second-order Møller-Plesset perturbation theory; SE, semi-empirical.

but requires sophisticated yet affordable strategies. These are based on composite QC schemes combined with the proper account of vibrational effects and, when needed, thermal contributions.

Dipeptide analogs containing more complex side chains have also been investigated; however, they do not add any new features to our general discussion, since the detected conformers reduce to only the C_7^{eq} structure because of their increased backbone rigidity [e.g., proline dipeptide analog (132)] or the formation of additional hydrogen bridges between the backbone and the side chain [e.g., serine dipeptide analog (133)]. In fact, the free energy difference between the C_7^{eq} and C_5 conformers increases [at the MP2/6-311++G(d,p) level] from 1 to 7 kJ mol^{-1} when going from the alanine to the serine dipeptide model (134). This explains why the less stable conformer has been experimentally observed only for the lighter systems (131, 133, 135).

3.4. Amino Acids: Glycine, Alanine, and (Homo)Cysteine

As already mentioned, systematic searches can be performed only for a limited number of LAMs, whereas more effective methodologies are needed for the exploration of more complex PESs (53). Prototypical amino acids have been selected to illustrate both situations.

The structure of isolated amino acids is ruled by both backbone (ϕ, ψ, and ω) and side-chain (χ) torsional angles, as shown in **Figure 2*b***. The nonplanarity of the NH_2 moiety suggests replacing the customary ϕ dihedral angle (HNCC) with $\phi' = \text{LP–C–C–C} = \phi + 120°$, where LP is the nitrogen lone pair perpendicular to the plane defined by the two aminic hydrogens and the C^α atom. The most stable backbone structures involve the formation of hydrogen bonds, which can be classified as I (bifurcated $NH_2\cdots O{=}C$; $\phi' \approx 180°$, $\psi \approx 180°$, $\omega \approx 180°$), II [$N\cdots H(O)$; $\phi' \approx 0°$, $\psi \approx 0°$, $\omega \approx 0°$], or III (bifurcated $NH_2\cdots OH$; $\phi' \approx 180°$, $\psi \approx 0°$, $\omega \approx 180°$) (16). Higher energy minima can be classified as type I′ (single $NH\cdots O{=}C$ hydrogen bond; $\phi' \approx 90°$, $\psi \approx 180°$, $\omega \approx 180°$) or type III′ (single $NH\cdots OH$ hydrogen bond; $\phi' \approx 180°$, $\psi \approx 90°$, $\omega \approx 180°$). Furthermore, conformers of type I, I′, and III can present higher energy counterparts for $\omega \approx 0°$, these being labeled as I_c, I'_c, and III_c in the following. The conformations of the side chain are described by the conventional g^-, g, and t labels, which are used to indicate the *gauche* and *trans* conformations of each χ dihedral angle.

While several studies of amino acids have been performed (see, e.g., 13, 14, 17, 136–139), the computational characterization is very often restricted to systematic searches of energy minima employing levels of theory of limited accuracy. These approaches usually only permit a posteriori interpretations based on the agreement between experimental and computed spectroscopic parameters for a reduced number of conformers, which are not necessarily the most stable ones. However, they are not appropriate for performing any fully unbiased disentanglement of the conformational landscape. The examples discussed in the following are based instead on a more comprehensive computational strategy allowing fully a priori predictions.

First of all, we consider the two simplest α-amino acids, glycine and alanine, for which a systematic search of the conformational PES is possible. For glycine, all of the eight conformers mentioned above (I, II, III, I′, III′, I_c, III_c, and I'_c) have been characterized, with four of them (I, III, I_c, and III_c) belonging to the C_s point group (equivalent C_α hydrogens) and the other four lacking any symmetry (41, 140). Glycine has been extensively characterized from both experimental and computational points of view (see 140 and references therein). Its limited size also allowed the exploitation of composite schemes based entirely on CC techniques [CCSD(T)/CBS+CV (67), which corresponds to the first three terms of Equation 1]. Furthermore, the availability—for the two most stable conformers—of the experimental rotational constants for several isotopic species allowed the determination of SE equilibrium structures (3). While the reader is referred to References 3 and 41, we point out the good performance of the ChS approach (see also **Table 2**), which provides results close to those obtained at the CCSD(T)/CBS+CV level.

Table 3 Rotational (megahertz) and quartic centrifugal distortion (kilohertz) constants,[a] vibrational frequencies (wave numbers) and, in parentheses, depolarization ratios[b] of alanine

Parameter	I_{exp}	I_{calc}	IIA_{exp}	IIA_{calc}	IIB_{exp}	IIB_{calc}
A_0	5,066.1455 (7)	5,031.47	4,973.0546 (35)	4,950.180		
B_0	3,100.9507 (5)	3,067.44	3,228.3375 (56)	3,183.80		
C_0	2,264.0131 (4)	2,258.84	2,307.8090 (42)	2,316.25		
Δ_J	2.445 (7)	2.409	2.11 (6)	1.40		
Δ_{JK}	−6.38 (1)	−6.37	−4.8 (3)	−2.61		
Δ_K	5.37 (5)	5.42	4.6 (5)	2.77		
δ_J	0.574 (2)	0.570	0.41 (2)	0.26		
δ_K	10.37 (3)	9.66	7.2 (7)	4.69		
$\nu_1(\rho_1)$	287 (0.20)	298 (0.14)	298 (0.21)	324 (0.26)	312 (0.19)	342 (0.19)
$\nu_2(\rho_2)$	364 (0.41)	374 (0.46)	376 (>0.6)	384 (0.72)	393 (>0.6)	403 (0.56)

Abbreviations: calc, calculated; exp, experimental; MP2, second-order Møller-Plesset perturbation theory.
[a]Experimental and calculated [CCSD(T)/cc-pVTZ equilibrium rotational constants and MP2/6-31G(d) vibrational corrections] data from Reference 138.
[b]Experimental and calculated [B3LYP/aug-cc-pVTZ harmonic values combined with MP2/6-31G(d) anharmonic contributions] data from Reference 30.

Moving to alanine (13, 25, 26, 30, 138, 141, 142), the two sides of the backbone plane are nonequivalent, with two nearly iso-energetic minima (referred to, in the following, as A and B) being found for structures of types II, III, I′, III′, and I'_c. The number of conformers thus increases to 13: I, IIA, IIB, IIIA, IIIB, IA′, IB′, IIIA′, IIIB′, I_c, III_c, IA'_c, and IB'_c, in order of decreasing stability (141). The comparison with experiment requires the computation of free energies of the different conformers in order to evaluate their population and of transition states ruling their interconversion. Only the I, IIA, IIB, IIIA, and IIIB conformers are sufficiently stable for an unbiased spectroscopic characterization. However, the relaxation of the III conformers to their I counterparts is ruled by low-energy barriers, which are easily overcome in the typical experimental conditions of rotational spectroscopy; the same applies to the relaxation of IIB to IIA. As a consequence, only the I and IIA conformers could be detected in microwave studies, with the former collecting the populations of the I, IIIA, and IIIB conformers and the latter those of the IIA and IIB conformers. It is remarkable that the computed relative populations (0.79 and 0.21 for I and IIA, respectively) are in excellent agreement with the experimental findings (0.8 and 0.2) (13). **Table 3** reports the comparison between experiment and theory [CCSD(T)/cc-pVTZ] for the rotational parameters of these two conformers (138). Fair agreement is noted, which is consistent with what was already pointed out: The CCSD(T)/cc-pVTZ level of theory is not able to predict rotational constants with an accuracy of a few megahertz.

Jet-cooled Raman spectroscopy was instead able to observe four conformers. In fact, Raman vibrational spectra were interpreted in terms of nonnegligible contributions from I, IIA, IIB, and IIIA (30). Very recently, the IIIB conformer was also tentatively detected in gas-phase CD experiments (142). **Table 3** illustrates good agreement between experimental and computed [B3LYP/aug-cc-pVTZ harmonic terms with MP2/6-31G(d) anharmonic contributions] Raman frequencies and depolarization ratios. For frequencies, we note surprisingly good agreement between theory and experiment (on average, within 10 cm^{-1}), possibly due to the very small anharmonic contributions to the investigated vibrations.

We now turn to the case of cysteine, which has two more degrees of freedom in the side chain, with the SH moiety being possibly involved in additional hydrogen bonds with the polar groups

of the backbone. The increased number of soft degrees of freedom makes this system suitable for applying the PES exploration procedure introduced in Section 2.1.2. As discussed in Reference 137, a systematic scan of the conformational PES at the MP2/cc-pVTZ level led to the identification of 71 unique conformers, thus defining a reference data set for our strategy. This latter strategy started by exploiting the conformational landscape using the PM7 semi-empirical method (91), with constrained geometry optimizations followed by single-point energy evaluations at the B3LYP-D3/6-31+G(d) level. Subsequently, a filtering procedure based on structural and energetic criteria was applied to remove redundancies. At the end, all the structures of the reference data set (137) were retrieved by performing about 2,000 energy evaluations and 150 geometry optimizations (compared to the 11,664 calculations needed in the original study) (53). Refinement of the results by geometry optimization at the MP2/6-311++G(d,p) level, followed by MP4/6-31++G(d,p) [MP4 is fourth-order MP perturbation theory (70)] energy evaluation led to 11 conformers within 12 kJ mol^{-1} (14). These results were in semiquantitative agreement with those obtained at higher computational levels (137). Among these 11 conformers, the 2 less stable ones probably have populations that are too low to be experimentally detectable. The structures of type III not involving side-chain hydrogen bonds easily relax to the corresponding structures of type I. Finally, the type I′ conformer can relax through a concerted path involving rotation of both NH_2 and SH groups to I structures. As a consequence, the number of detectable conformers reduces to six: two for each type (I, II, and III). Spectroscopic parameters computed at the MP2 level with a TZ basis set were then found to be in fair agreement with their experimental counterparts (14). Note that the populations of the different conformers estimated experimentally are significantly different from those derived from the electronic energies, whereas there is good agreement with those obtained using the free energies evaluated at the carrier gas temperature. This further stresses that unbiased comparison between theory and experiment requires proper consideration of thermal contributions.

Moving to homocysteine, the flexibility of the system is further enhanced because of the introduction of a methylene group in the side chain. A filter of 15 kJ mol^{-1} after the exploration step produced 22 candidate structures, with 5 of them relaxing to more stable conformers after full geometry optimization. At the rDSD level, the remaining 17 conformers span an energy range of 8.5 kJ mol^{-1} (143). Although one structure of type II is the most stable conformer in terms of electronic energies, this is significantly destabilized once vibrational and thermal effects are incorporated. Consequently, a structure of type I was found to have the lowest free energy. Interestingly, a comprehensive analysis of possible relaxation paths further reduces the number of detectable structures to seven (six of type I and one of type II) when employing a free energy threshold of 6 kJ mol^{-1}. In view of the large number of conformers possibly contributing to the experimental rotational spectrum, it was crucial to accurately improve their molecular structure. An F12 variant of ChS, the ChS-F12 approach (144), was employed. Notably, this level of theory is still affordable for systems of the size of homocysteine and provides remarkable agreement between theory and experiment [e.g., an absolute average error of 7 Mhz on rotational constants (143)]. Despite the extremely complex conformational landscape, theory is able to predict the correct number and type of detectable conformers (143).

The different conformational landscapes of cysteine and homocysteine are reflected in different distinctive features of their IR and Raman spectra (145). For example, the NH_2 in-plane bend and the COH bend modes observed in cysteine at 1,583 and 650 cm^{-1}, respectively, shift to 1,633 and 470 cm^{-1}, respectively, in homocysteine, thus highlighting different intramolecular noncovalent interactions. Harmonic B3LYP computations, albeit not quantitatively accurate, are able to correctly reproduce all experimental features, provided that Boltzmann averaging for the

low-lying conformers is taken into account. In structural terms, the most significant outcome of rotational and IR/Raman spectroscopy together with QC computations is that only in cysteine is the SH moiety involved in hydrogen bonds with the backbone.

4. CONCLUDING REMARKS

In this review, the main aspects of a general approach aimed at the accurate computational characterization of the rotational and vibrational signatures of the molecular bricks of life have been illustrated by selected examples. The main outcomes can be summarized as follows:

- Concerning rotational spectroscopy, accurate rotational constants (which are the leading terms of this technique) can be obtained from geometries (providing equilibrium values) computed at the ChS level (or one of its variants) and combining the equilibrium terms with vibrational corrections evaluated using suitable DFT models. This approach is affordable for systems containing up to about 10 nonhydrogen atoms and also showing a relevant flexibility. For larger molecules, one has to resort to double-hybrid functionals, which perform at least as well as the CCSD(T) method when used with a TZ basis set but with a much more favorable scaling with the number of basis functions. In addition, the TMA and LRA models can be employed to further improve these structures. Concerning other spectroscopic parameters, while composite schemes can be exploited for medium-sized species, they can be effectively computed at the DFT level for larger molecules.
- In terms of vibrational spectroscopy, for systems bearing up to about 10 nonhydrogen atoms, a reliable yet effective strategy for obtaining accurate vibrational parameters is based on harmonic frequencies evaluated by means of the ChS model (or one of its variants) coupled with anharmonic contributions computed by double-hybrid functionals with partially augmented TZ basis sets. For larger systems, one has to rely on harmonic contributions computed using double-hybrid functionals combined with anharmonic corrections at the same level. Anharmonic contributions can also be obtained using hybrid functionals with partially augmented DZ basis sets without any excessive degradation of the results.
- For flexible molecules containing a significant number of soft degrees of freedom, a crucial piece of information for both vibrational and rotational spectroscopy is the number and type of conformers contributing to the experimental spectra. This requires first of all an effective exploration of the PES that can be performed by genetic algorithms driving local geometry optimizations with last-generation semi-empirical methods. Further refinement of the surviving species by means of geometry optimizations using double-hybrid functionals, the definition of the interconversion paths between low-lying conformers, and the incorporation of vibrational and thermal effects on energetics complete the procedure. Subsequently, for each energy minimum, the characterization described in the first two bullet points has to be performed.

Even if further developments are needed, especially for large flexible systems, the implementation of the proposed computational strategy in a user-friendly context (1, 89) might pave the way toward a widespread application, which could also be used by nonspecialists, in the characterization of experimental spectra of the molecular bricks of life.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work has been supported by the Ministry of University and Research (MUR) (grants 2017A4XRCA and 202082CE3T), the Italian Space Agency (ASI) (Life in Space project, N.2019-3-U.0), and the University of Bologna (RFO funds).

LITERATURE CITED

1. Barone V. 2016. The virtual multifrequency spectrometer: a new paradigm for spectroscopy. *WIREs Comput. Mol. Sci.* 6:86–110
2. Sugiki T, Kobayashi N, Fujiwara T. 2017. Modern technologies of solution nuclear magnetic resonance spectroscopy for three-dimensional structure determination of proteins open avenues for life scientists. *Comput. Struct. Biotechnol. J.* 15:328–29
3. Puzzarini C, Barone V. 2018. Diving for accurate structures in the ocean of molecular systems with the help of spectroscopy and quantum chemistry. *Acc. Chem. Res.* 51:548–56
4. Lane JL, ed. 2018. *Frontiers and Advances in Molecular Spectroscopy*. Amsterdam: Elsevier
5. Puzzarini C, Bloino J, Tasinato N, Barone V. 2019. Accuracy and interpretability: the devil and the holy grail. New routes across old boundaries in computational spectroscopy. *Chem. Rev.* 119:8131–91
6. Melosso M, Bizzocchi L, Gazzeh H, Tonolo F, Guillemin JC, et al. 2022. Gas-phase identification of (*Z*)-1,2-ethenediol, a key prebiotic intermediate in the formose reaction. *Chem. Commun.* 58:2750–53
7. Rivilla VM, Colzi L, Jiménez-Serra I, Martín-Pintado J, Megías A, et al. 2022. Precursors of the RNA world in space: detection of (*Z*)-1,2-ethenediol in the interstellar medium, a key intermediate in sugar formation. *Astrophys. J. Lett.* 929:L11
8. Gardner JP, Mather JC, Clampin M, Doyon R, Greenhouse MA, et al. 2006. The James Webb Space Telescope. *Space Sci. Rev.* 123:485–606
9. He Y, Tang L, Wu X, Hou X, Lee Y. 2007. Spectroscopy: the best way toward green analytical chemistry? *Appl. Spectrosc. Rev.* 42:119–38
10. Lindon JC, Tranter GE, Koppenaal DW, eds. 2017. *Encyclopedia of Spectroscopy and Spectrometry*. Oxford, UK: Academic. 3rd ed.
11. Martin-Drumel MA, McCarthy MC, Patterson D, McGuire BA, Crabtree KN. 2016. Automated microwave double resonance spectroscopy: a tool to identify and characterize chemical compounds. *J. Chem. Phys.* 144:124202
12. Lesarri A, Mata S, López J, Alonso J. 2003. A laser-ablation molecular-beam Fourier-transform microwave spectrometer: the rotational spectrum of organic solids. *Rev. Sci. Instrum.* 74:4799–804
13. Blanco S, Lesarri A, López JC, Alonso JL. 2004. The gas-phase structure of alanine. *J. Am. Chem. Soc.* 126:11675–83
14. Sanz ME, Blanco S, López JC, Alonso JL. 2008. Rotational probes of six conformers of neutral cysteine. *Angew. Chem. Int. Ed.* 47:6216–20
15. Puzzarini C, Biczysko M, Barone V, Largo L, Peña I, et al. 2014. Accurate characterization of the peptide linkage in the gas phase: a joint quantum-chemical and rotational spectroscopy study of the glycine dipeptide analogue. *J. Phys. Chem. Lett.* 5:534–40
16. Alonso JL, López JC. 2015. Microwave spectroscopy of biomolecular building blocks. In *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, ed. A Rijs, J Oomens, pp. 335–401. Cham, Switz.: Springer
17. León I, Alonso ER, Mata S, Cabezas C, Alonso JL. 2019. Unveiling the neutral forms of glutamine. *Angew. Chem. Int. Ed.* 58:16002–7
18. Brown GG, Dian BC, Douglass KO, Geyer SM, Shipman ST, Pate BH. 2008. A broadband Fourier transform microwave spectrometer based on chirped pulse excitation. *Rev. Sci. Instrum.* 79:053103
19. Steber AL, Neill JL, Zaleski DP, Pate BH, Lesarri A, et al. 2011. Structural studies of biomolecules in the gas phase by chirped-pulse Fourier transform microwave spectroscopy. *Faraday Discuss.* 150:227–42
20. Pérez C, Neill JL, Muckle MT, Zaleski DP, Peña I, et al. 2015. Water–water and water–solute interactions in microsolvated organic complexes. *Angew. Chem. Int. Ed.* 54:979–82

21. Steber AL, Pérez C, Temelso B, Shields GC, Rijs AM, et al. 2017. Capturing the elusive water trimer from the stepwise growth of water on the surface of the polycyclic aromatic hydrocarbon acenaphthene. *J. Phys. Chem. Lett.* 8(23):5744–50
22. Puzzarini C, Spada L, Alessandrini S, Barone V. 2020. The challenge of non-covalent interactions: Theory meets experiment for reconciling accuracy and interpretation. *J. Phys. Condens. Matter* 32:343002
23. Xie F, Fusè M, Hazrah AS, Jaeger W, Barone V, Xu Y. 2020. Discovering the elusive global minimum in a ternary chiral cluster: rotational spectra of propylene oxide trimer. *Angew. Chem. Int. Ed.* 59:22427–30
24. Li X, Spada L, Alessandrini S, Zheng Y, Lengsfels KG, et al. 2022. Stacked but not stuck: unveiling the role of π–π^* interactions with the help of the benzofuran-formaldehyde complex. *Angew. Chem. Int. Ed.* 61:264–70
25. Linder R, Nispel M, Haber T, Kleinermanns K. 2005. Gas-phase FT-IR-spectra of natural amino acids. *Chem. Phys. Lett.* 409:260–64
26. Linder R, Seefeld K, Vavra A, Kleinermanns K. 2008. Gas phase infrared spectra of nonaromatic amino acids. *Chem. Phys. Lett.* 453:1–6
27. Bloino J, Biczysko M, Barone V. 2012. General perturbative approach for spectroscopy, thermodynamics and kinetics: methodological background and benchmark studies. *J. Chem. Theory Comput.* 8:1015–36
28. Barone V, Biczysko M, Bloino J. 2014. Fully anharmonic IR and Raman spectra of medium-size molecular systems: accuracy and interpretation. *Phys. Chem. Chem. Phys.* 16:1759–87
29. Balabin RM. 2010. Conformational equilibrium in glycine: experimental jet-cooled Raman spectrum. *J. Phys. Chem. Lett.* 1:20–23
30. Balabin RM. 2010. The identification of the two missing conformers of gas-phase alanine: a jet-cooled Raman spectroscopy study. *Phys. Chem. Chem. Phys.* 12:5980–82
31. Barone V. 2012. *Computational Strategies for Spectroscopy: From Small Molecules to Nano Systems*. Hoboken, NJ: Wiley
32. Barone V, Alessandrini S, Biczysko M, Cheeseman JR, Clary DC, et al. 2021. Computational molecular spectroscopy. *Nat. Rev. Methods Primers* 1:38
33. Cazzoli G, Puzzarini C, Harding ME, Gauss J. 2009. The hyperfine structure in the rotational spectrum of water: Lamb-dip technique and quantum-chemical calculations. *Chem. Phys. Lett.* 473(1):21–25
34. Puzzarini C, Cazzoli G, Harding ME, Vázquez J, Gauss J. 2009. A new experimental absolute nuclear magnetic shielding scale for oxygen based on the rotational hyperfine structure of $H_2^{17}O$. *J. Chem. Phys.* 131:234304
35. Martin JML, Kesharwani MK. 2014. Assessment of CCSD(T)-F12 approximations and basis sets for harmonic vibrational frequencies. *J. Chem. Theory Comput.* 10:2085–90
36. Agbaglo D, Fortenberry RC. 2019. The performance of explicitly correlated wavefunctions [CCSD(T)-F12b] in the computation of anharmonic vibrational frequencies. *Int. J. Quantum Chem.* 119:e25899
37. Heim ZN, Amberger BK, Esselman BJ, Stanton JF, Woods RC, McMahon RJ. 2020. Molecular structure determination: equilibrium structure of pyrimidine (m-$C_4H_4N_2$) from rotational spectroscopy (r_e^{SE}) and high-level *ab initio* calculation (r_e) agree within the uncertainty of experimental measurement. *J. Chem. Phys.* 152:104303
38. Barone V, Puzzarini C, Mancini G. 2021. Integration of theory, simulation, artificial intelligence and virtual reality: a four-pillar approach for reconciling accuracy and interpretability in computational spectroscopy. *Phys. Chem. Chem. Phys.* 23:17079–96
39. Puzzarini C, Biczysko M, Barone V. 2011. Accurate anharmonic vibrational frequencies for uracil: the performance of composite schemes and hybrid CC/DFT model. *J. Chem. Theory Comput.* 7:3702–10
40. Davisson JL, Brinckmann NR, Polik WF. 2012. Accurate and efficient calculation of excited vibrational states from quartic potential energy surfaces. *Mol. Phys.* 51:2587–98
41. Barone V, Biczysko M, Bloino J, Puzzarini C. 2013. Accurate structure, thermochemistry and spectroscopic parameters from CC and CC/DFT schemes: the challenge of the conformational equilibrium in glycine. *Phys. Chem. Chem. Phys.* 15:10094–111
42. Puzzarini C, Barone V. 2010. Toward spectroscopic accuracy for open-shell systems: molecular structure and hyperfine coupling constants of H_2CN, H_2CP, NH_2, and PH_2 as test cases. *J. Chem. Phys.* 133:184301

43. Alessandrini S, Barone V, Puzzarini C. 2020. Extension of the "cheap" composite approach to noncovalent interactions: the jun-ChS scheme. *J. Chem. Theory Comput.* 16:988–1006
44. Lupi J, Alessandrini S, Puzzarini C, Barone V. 2021. junChS and junChS-F12 models: parameter-free efficient yet accurate composite schemes for energies and structures of noncovalent complexes. *J. Chem. Theory Comput.* 17(11):6974–92
45. Mills IM. 1972. Vibration-rotation structure in asymmetric- and symmetric-top molecules. In *Molecular Spectroscopy: Modern Research*, Vol. 1, ed. KN Rao, CW Mathews, pp. 115–40. New York: Academic
46. Papoušek D, Aliev MR. 1982. *Molecular Vibrational/Rotational Spectra*. Amsterdam: Elsevier
47. Aliev MR, Watson JKG. 1985. Higher-order effects in the vibration-rotation spectra of semirigid molecules. In *Molecular Spectroscopy: Modern Research*, Vol. 3, ed. KN Rao, pp. 2–67. New York: Academic
48. Biczysko M, Panek P, Scalmani G, Bloino J, Barone V. 2010. Harmonic and anharmonic vibrational frequency calculations with the double-hybrid B2PLYP method: analytic second derivatives and benchmark studies. *J. Chem. Theory Comput.* 6:2115–25
49. Bloino J, Biczysko M, Barone V. 2015. Anharmonic effects on vibrational spectra intensities: infrared, Raman, vibrational circular dichroism and Raman optical activity. *J. Phys. Chem. A* 119:11862–74
50. Barone V, Ceselin G, Fusè M, Tasinato N. 2020. Accuracy meets interpretability for computational spectroscopy by means of hybrid and double-hybrid functionals. *Front. Chem.* 8:584203
51. Franke PR, Stanton JF, Douberly GE. 2021. How to VPT2: accurate and intuitive simulations of CH stretching infrared spectra using VPT2+K with large effective Hamiltonian resonance treatments. *J. Phys. Chem. A* 125:1301–24
52. Puzzarini C, Stanton JF, Gauss J. 2010. Quantum-chemical calculation of spectroscopic parameters for rotational spectroscopy. *Int. Rev. Phys. Chem.* 29:273–367
53. Mancini G, Fusè M, Lazzari F, Chandramouli B, Barone V. 2020. Unsupervised search of low-lying conformers with spectroscopic accuracy: a two-step algorithm rooted into the island model evolutionary algorithm. *J. Chem. Phys.* 153:124110
54. Puzzarini C, Biczysko M, Barone V. 2010. Accurate harmonic/anharmonic vibrational frequencies for open-shell systems: performances of the B3LYP/N07D model for semirigid free radicals benchmarked by CCSD(T) computations. *J. Chem. Theory Comput.* 6:828–38
55. Puzzarini C, Barone V. 2011. Extending the molecular size in accurate quantum-chemical calculations: the equilibrium structure and spectroscopic properties of uracil. *Phys. Chem. Chem. Phys.* 13:7189–97
56. Vainio MJ, Johnson MS. 2007. Generating conformer ensembles using a multiobjective genetic algorithm. *J. Chem. Inf. Model.* 47:2462–74
57. O'Boyle NM, Vandermeersch T, Flynn CJ, Maguire AR, Hutchison GR. 2011. Confab – systematic generation of diverse low-energy conformers. *J. Cheminf.* 3:8
58. Goto H, Takahashi T, Takata Y, Ohta K, Nagashima U. 2003. CONFLEX: conformational behaviors of polypeptides as predicted by a conformational space search. *Nanotech* 1:32–35
59. Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT. 2010. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* 50:572–84
60. Pracht P, Bohle F, Grimme S. 2020. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* 22:7169–92
61. Beusen DD, Berkley Shands E, Karasek S, Marshall GR, Dammkoehler RA. 1996. Systematic search in conformational analysis. *J. Mol. Struct. THEOCHEM* 370:157–71
62. Rienstra CM, Tucker-Kellogg L, Jaroniec CP, Hohwy M, Reif B, et al. 2002. *De novo* determination of peptide structure with solid-state magic-angle spinning NMR spectroscopy. *PNAS* 99:10260–65
63. Brownlee J. 2012. *Clever Algorithms: Nature-Inspired Programming Recipes*, rev. 2. **http://www.CleverAlgorithms.com**
64. Watson JKG. 1967. Determination of centrifugal distortion coefficients of asymmetric-top molecules. *J. Chem. Phys.* 46:1935–49
65. Piccardo M, Bloino J, Barone V. 2015. Generalized vibrational perturbation theory for rotovibrational energies of linear, symmetric, and asymmetric tops: theory, approximations, and automated approaches to deal with medium-to-large molecular systems. *Int. J. Quantum Chem.* 115:948–82

66. Heckert M, Kállay M, Gauss J. 2005. Molecular equilibrium geometries based on coupled-cluster calculations including quadruple excitations. *Mol. Phys.* 103:2109–115
67. Heckert M, Kállay M, Tew DP, Klopper W, Gauss J. 2006. Basis-set extrapolation techniques for the accurate calculation of molecular equilibrium geometries using coupled-cluster theory. *J. Chem. Phys.* 125:044108
68. Matthews DA, Cheng L, Harding ME, Lipparini F, Stopkowicz S, et al. 2020. Coupled-cluster techniques for computational chemistry: the CFOUR program package. *J. Chem. Phys.* 152:214108
69. Raghavachari K, Trucks GW, Pople JA, Head-Gordon M. 1989. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* 157:479–83
70. Møller C, Plesset MS. 1934. Note on an approximation treatment for many-electron systems. *Phys. Rev.* 46(7):618–22
71. Kong L, Bischoff FA, Valeev EF. 2012. Explicitly correlated R12/F12 methods for electronic structure. *Chem. Rev.* 112:75–107
72. Barone V, Lupi J, Salta Z, Tasinato N. 2021. Development and validation of a parameter-free model chemistry for the computation of reliable reaction rates. *J. Chem. Theory Comput.* 17:4913–28
73. Knizia G, Adler TB, Werner HJ. 2009. Simplified CCSD(T)-F12 methods: theory and benchmarks. *J. Chem. Phys.* 130:054104
74. Papajak E, Zheng J, Xu X, Leverentz HR, Truhlar DG. 2011. Perspectives on basis sets beautiful: seasonal plantings of diffuse basis functions. *J. Chem. Theory Comput.* 7:3027–34
75. Peterson KA, Adler TB, Werner HJ. 2008. Systematically convergent basis sets for explicitly correlated wavefunctions: the atoms H, He, B–Ne, and Al–Ar. *J. Chem. Phys.* 128:084102
76. Werner HJ, Adler TB, Manby FR. 2007. General orbital invariant MP2-F12 theory. *J. Chem. Phys.* 126:164102
77. Spackman PR, Jayatilaka D, Karton A. 2016. Basis set convergence of CCSD(T) equilibrium geometries using a large and diverse set of molecular structures. *J. Chem. Phys.* 145:104101
78. Becke AD. 1988. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* 38:3098–100
79. Zhao Y, Truhlar DG. 2005. Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions. *J. Phys. Chem. A* 109:5656–67
80. Grimme S. 2006. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* 124:034108
81. Santra G, Sylvetsky N, Martin JM. 2019. Minimally empirical double-hybrid functionals trained against the GMTKN55 database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4. *J. Phys. Chem. A* 123:5129–43
82. Grimme S, Antony J, Ehrlich S, Krieg H. 2010. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* 132:154104
83. Grimme S, Ehrlich S, Goerigk L. 2011. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32:1456–65
84. Papajak E, Truhlar DG. 2011. Convergent partially augmented basis sets for post-Hartree-Fock calculations of molecular properties and reaction barrier heights. *J. Chem. Theory Comput.* 7:10–18
85. Yang Q, Mendolicchio M, Barone V, Bloino J. 2021. Accuracy and reliability in the simulation of vibrational spectra: a comprehensive benchmark of energies and intensities issuing from generalized vibrational perturbation theory to second order (GVPT2). *Front. Astron. Space Sci.* 8:665232
86. Ceselin G, Barone V, Tasinato N. 2021. Accurate biomolecular structures by the nano-LEGO approach: pick the bricks and build your geometry. *J. Chem. Theory Comput.* 17:7290–311
87. Melli A, Tonolo F, Barone V, Puzzarini C. 2021. Extending the applicability of the semi-experimental approach by means of "template molecule" and "linear regression" models on top of DFT computations. *J. Phys. Chem. A* 125:9904–16
88. Penocchio E, Piccardo M, Barone V. 2015. Semiexperimental equilibrium structures for building blocks of organic and biological molecules: the B2PLYP route. *J. Chem. Theory Comput.* 11:4689–707
89. Lazzari F, Salvadori A, Mancini G, Barone V. 2020. Molecular perception for visualization and computation: the Proxima library. *J. Chem. Inf. Model.* 60:2668–72

90. Ferro-Costas D, Mosquera-Lois I, Fernandez-Ramos A. 2021. Torsiflex: an automatic generator of torsional conformers. Application to the twenty proteinogenic amino acids. *J. Cheminform.* 13:100
91. Stewart JJP. 2007. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* 13:1173–213
92. Bannwarth C, Ehlert S, Grimme S. 2019. GFN2-xTB, an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* 15:1652–71
93. Boese AD, Oren M, Atasoylu O, Martin JML, Kállay M, Gauss J. 2004. W3 theory: robust computational thermochemistry in the kJ/mol accuracy range. *J. Chem. Phys.* 120:4129–41
94. Karton A, Rabinovich E, Martin JML, Ruscic B. 2006. W4 theory for computational thermochemistry: in pursuit of confident sub-kJ/mol predictions. *J. Chem. Phys.* 125:144108
95. Karton A, Martin JML. 2012. Explicitly correlated W*n* theory: W1-F12 and W2-F12. *J. Chem. Phys.* 136:124114
96. Tajti A, Szalay PG, Császár AG, Kállay M, Gauss J, et al. 2004. HEAT: high accuracy extrapolated *ab initio* thermochemistry. *J. Chem. Phys.* 121:11599–613
97. Császár AG, Allen WD, Schaefer HF III. 1998. In pursuit of the *ab initio* limit for conformational energy prototypes. *J. Chem. Phys.* 108:9751–64
98. Peterson KA, Feller D, Dixon D. 2012. Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges. *Theor. Chem. Acc.* 113:1079
99. Bomble YJ, Vázquez J, Kállay M, Michauk C, Szalay PG, et al. 2006. High-accuracy extrapolated *ab initio* thermochemistry. II. Minor improvements to the protocol and a vital simplification. *J. Chem. Phys.* 125:064108
100. Harding ME, Vázquez J, Ruscic B, Wilson AK, Gauss J, Stanton JF. 2008. High-accuracy extrapolated ab initio thermochemistry. III. Additional improvements and overview. *J. Chem. Phys.* 128:114111
101. Alessandrini S, Tonolo F, Puzzarini C. 2021. In search of phosphorus in astronomical environments: the reaction between the CP radical ($X^2\Sigma^+$) and methanimine. *J. Chem. Phys.* 154:054306
102. Tew DP, Klopper W, Heckert M, Gauss J. 2007. Basis set limit CCSD(T) harmonic vibrational frequencies. *J. Phys. Chem. A* 111:11242–48
103. Watrous A, Westbrook BR, Fortenberry RC. 2021. F12-TZ-cCR: a methodology for faster and still highly accurate quartic force fields. *J. Phys. Chem. A* 125:10532–40
104. Barone V. 2005. Anharmonic vibrational properties by a fully automated second order perturbative approach. *J. Chem. Phys.* 122:014108
105. Gordy W, Cook RL. 1984. *Microwave Molecular Spectra*. New York: Wiley. 3rd ed.
106. Piccardo M, Penocchio E, Puzzarini C, Biczysko M, Barone V. 2015. Semi-experimental equilibrium structure determinations by employing B3LYP/SNSD anharmonic force fields: validation and application to semirigid organic molecules. *J. Phys. Chem. A* 119:2058–82
107. Rosnik AM, Polik WF. 2014. VPT2+K spectroscopic constants and matrix elements of the transformed vibrational Hamiltonian of a polyatomic molecule with resonances using Van Vleck perturbation theory. *Mol. Phys.* 112:261–300
108. Vázquez J, Stanton JF. 2007. Treatment of Fermi resonance effects on transition moments in vibrational perturbation theory. *Mol. Phys.* 105:101–9
109. Mendolicchio M, Bloino J, Barone V. 2021. A general perturb-then-diagonalize model for the vibrational frequencies and intensities of molecules belonging to Abelian and non-Abelian symmetry groups. *J. Chem. Theory Comput.* 19:1759–87
110. Puzzarini C, Tasinato N, Bloino J, Spada L, Barone V. 2019. State-of-the-art computation of the rotational and IR spectra of the methyl-cyclopropyl cation: hints on its detection in space. *Phys. Chem. Chem. Phys.* 21:3615–25
111. Baiardi A, Bloino J, Barone V. 2017. Simulation of vibronic spectra of flexible systems: hybrid DVR-harmonic approaches. *J. Chem. Theory Comput.* 13:2804–22
112. Ceselin G, Salta Z, Bloino J, Tasinato N, Barone V. 2022. Accurate quantum chemical spectroscopic characterization of glycolic acid: a route toward its astrophysical detection. *J. Phys. Chem. A* 126:2373–87
113. Barone V, Fusè M, Vieira Pinto SM, Tasinato N. 2021. A computational journey across nitroxide radicals: from structure to spectroscopic properties and beyond. *Molecules* 26:7404

114. Puzzarini C, Heckert M, Gauss J. 2008. The accuracy of rotational constants predicted by high-level quantum-chemical calculations. I. Molecules containing first-row atoms. *J. Chem. Phys.* 128:194108
115. Alessandrini S, Gauss J, Puzzarini C. 2018. Accuracy of rotational parameters predicted by high-level quantum-chemical calculations: case study of sulfur-containing molecules of astrochemical interest. *J. Chem. Theory Comput.* 14:5360–71
116. Puzzarini C, Biczysko M, Barone V, Peña I, Cabezas C, Alonso JL. 2013. Accurate molecular structure and spectroscopic properties of nucleobases: a combined computational-microwave investigation of 2-thiouracil as a case study. *Phys. Chem. Chem. Phys.* 15:16965–75
117. Barone V, Biczysko M, Bloino J, Cimino P, Penocchio E, Puzzarini C. 2015. CC/DFT route toward accurate structures and spectroscopic features for observed and elusive conformers of flexible molecules: pyruvic acid as a case study. *J. Chem. Theory Comput.* 11:4342–63
118. Barone V, Festa G, Grandi A, Rega N, Sanna N. 2004. Vibrational spectra of large molecules by density functional computations beyond the harmonic approximation. The case of uracil and 2-thiouracil. *Chem. Phys. Lett.* 388:279–83
119. Fornaro T, Biczysko M, Monti S, Barone V. 2014. Dispersion corrected DFT approaches for anharmonic vibrational frequency calculations: nucleobases and their dimers. *Phys. Chem. Chem. Phys.* 16:10112–18
120. López JC, Peña I, Sanz ME, Alonso JL. 2007. Probing thymine with laser ablation molecular beam Fourier transform microwave spectroscopy. *J. Chem. Phys.* 126:191103
121. Alonso JL, Peña I, López JC, Vaquero V. 2009. Rotational spectra signatures of four tautomers of guanine. *Angew. Chem. Int. Ed.* 48:6141–43
122. Alonso JL, Vaquero V, Peña I, López JC, Mata S, Caminati W. 2013. All five forms of cytosine revealed in the gas phase. *Angew. Chem. Int. Ed.* 52:2331–34
123. Choi MY, Miller RE. 2006. Four tautomers of isolated guanine from infrared laser spectroscopy in helium nanodroplets. *J. Am. Chem. Soc.* 128:7320–28
124. Chin W, Mons M, Piuzzi F, Tardivel B, Gorb IL, Lesczynski J. 2004. Gas phase rotamers of the nucleobase 9-methylguanine enol and its monohydrate: optical spectroscopy and quantum mechanical calculations. *J. Phys. Chem. A* 108:8237–43
125. Karton A. 2019. Thermochemistry of guanine tautomers re-examined by means of high-level CCSD(T) composite *ab initio* methods. *Austr. J. Chem.* 72:607–9
126. Nejad A, Meyer E, Suhm MA. 2020. Glycolic acid as a vibrational anharmonicity benchmark. *J. Phys. Chem. Lett.* 11:5228–33
127. Godfrey PD, Rodgers FM, Brown RD. 1997. Theory versus experiment in jet spectroscopy: glycolic acid. *J. Am. Chem. Soc.* 119:2232–39
128. Kisiel Z, Pszczołkowski L, Białkowska-Jaworska E, Charnley SB. 2016. Millimetre wave rotational spectrum of glycolic acid. *J. Mol. Spectrosc.* 321:13–22
129. Yu W, Xu X, Li H, Pang R, Fang K, Lin Z. 2009. Extensive conformational searches of 13 representative dipeptides and an efficient method for dipeptide structure determination based on amino acid conformers. *J. Comput. Chem.* 30:2105–21
130. Bhattacharya A, Bernstein ER. 2011. Influence of turn (or fold) and local charge in fragmentation of the peptide analogue molecule CH_3CO-Gly-NH_2 following single-photon VUV (118.2 nm) ionization. *J. Phys. Chem. A* 115:10679–88
131. Cabezas C, Varela M, Cortijo V, Jiménez AL, Peña L, et al. 2013. The alanine model dipeptide Ac-Ala-NH_2 exists as a mixture of $C7_{eq}$ and C_5 conformers. *Phys. Chem. Chem. Phys.* 15:2580–85
132. Cabezas C, Varela M, Alonso JL. 2013. Probing the γ-turn in a short proline dipeptide chain. *Chem. Phys. Chem.* 14:2539–43
133. Cabezas C, Robben MAT, Rijs AM, Peña I, Alonso JL. 2015. Fourier transform microwave spectroscopy of Ac-Ser-NH_2: the role of side-chain interactions in peptide folding. *Phys. Chem. Chem. Phys.* 17:20274–80
134. Chakraborty D, Banerjee A, Wales DJ. 2021. Side-chain polarity modulates the intrinsic conformational landscape of model dipeptides. *J. Phys. Chem. B* 125:5809–22
135. Grenie Y, Avignon M, Garrigou-Lagrange C. 1975. Molecular structure study of dipeptides isolated in an argon matrix by infrared spectroscopy. *J. Mol. Struct.* 24:293–307

136. Blanco S, Sanz ME, López JC, Alonso JL. 2007. Revealing the multiple structures of serine. *PNAS* 104:20183–88
137. Wilke JJ, Lind MC, Schaefer HF III, Császár AG, Allen WD. 2009. Conformers of gaseous cysteine. *J. Chem. Theory Comput.* 5:1511–23
138. Jaeger HM, Scheafer HF III, Demaison J, Császár A, Allen WD. 2010. Lowest-lying conformers of alanine: pushing theory to ascertain precise energetics and semiexperimental R_e structures. *J. Chem. Theory Comput.* 6:3066–78
139. Nguyen HVL, Kleiner I. 2022. Understanding (coupled) large amplitude motions: the interplay of microwave spectroscopy, spectral modeling, and quantum chemistry. *Phys. Sci. Rev.* 7:679–726
140. Barone V, Biczysko M, Bloino J, Puzzarini C. 2013. Glycine conformers: a never-ending story? *Phys. Chem. Chem. Phys.* 15:1358–63
141. Császár A. 1996. Conformers of gaseous alanine. *J. Phys. Chem.* 100:3541–51
142. Meinert C, Garcia AD, Topin J, Jones NC, Dieckmann M, et al. 2022. Amino acid gas phase circular dichroism and implications for the origin of biomolecular asymmetry. *Nat. Commun.* 13:502
143. León I, Fusè M, Alonso ER, Mata S, Mancini G, et al. 2022. Unbiased disentanglement of conformational baths with the help of microwave spectroscopy, quantum chemistry, and artificial intelligence: the puzzling case of homocysteine. *J. Chem. Phys.* 157:074107
144. Lupi J, Puzzarini C, Cavallotti C, Barone V. 2020. State-of-the-art quantum chemistry meets variable reaction coordinate transition state theory to solve the puzzling case of the $H_2S + Cl$ system. *J. Chem. Theory Comput.* 16(8):5090–104
145. Gunasekaran S, Bright A, Renuga Devi TS, Arunbalaji R, Anand G, et al. 2010. Experimental and semi-empirical computations of the vibrational spectra of methionine, homocysteine and cysteine. *Arch. Phys. Res.* 1:12–26

Annual Review of
Physical Chemistry
Volume 74, 2023

Contents

Errata

An online log of corrections to *Annual Review of Physical Chemistry* articles may be found at http://www.annualreviews.org/errata/physchem