PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2025, **27**, 16383

Computational efficiency meets spectroscopic accuracy: an unsupervised workflow for equilibrium geometries and vibrational effects in gas-phase prebiotic molecules†

Marco Mendolicchio, *[a] Lina Uribe,[ab] Federico Lazzari, [b] Luigi Crisci, [b] Giovanni Scalmani, [c] Micheal J. Frisch[c] and Vincenzo Barone *[d]

Received 28th April 2025,
Accepted 11th July 2025

DOI: 10.1039/d5cp01611h

rsc.li/pccp

Equilibrium molecular geometries are essential for understanding molecular systems, particularly in the gas phase, where intrinsic stereoelectronic effects can be disentangled from environmental influences. High-resolution rotational spectroscopy offers direct structural information and is now applicable to molecules with up to 50 atoms, significantly expanding its scope and demanding more advanced computational support to account for vibrational averaging and spectral complexity. Herein, we present an automated workflow that integrates the Pisa composite schemes (PCS) with efficient vibrational correction models, interfacing seamlessly with Gaussian and MSR programs. The protocol is designed for medium-sized molecules where relativistic and static correlation effects can be neglected, and is demonstrated on a set of prebiotic and biologically relevant compounds. Reliable equilibrium geometries are obtained for both semi-rigid and flexible species, provided that a second-order vibrational perturbation theory (VPT2) treatment is adequate; proline is included as a representative flexible case. Additionally, the phenyl radical is considered as a prototypical open-shell system, supported by extensive isotopic experimental data. This strategy enables the accurate and cost-effective determination of equilibrium geometries for molecules beyond the small-molecule regime, outperforming conventional methods and offering broad applicability in astrochemistry, prebiotic chemistry, and molecular spectroscopy.

1 Introduction

Accurate structural parameters—such as bond lengths, valence angles, and dihedral angles—are essential for understanding molecular behavior across a wide range of chemical and physical contexts. Gas-phase studies, in particular, allow intrinsic stereoelectronic effects to be disentangled from environmental perturbations, offering direct insight into the factors that govern molecular structure and reactivity. This need has become increasingly evident as molecular sciences expand into more complex and multidisciplinary domains.

The growing importance of precise geometries is especially pronounced in astrochemistry and atmospheric chemistry, where accurate structures are required to model exotic environments, including interstellar space[1] and planetary atmospheres.[2] In astrochemistry, structural precision is critical for interpreting spectroscopic data and identifying novel species under extreme conditions. Similarly, atmospheric models depend on reliable geometrical inputs to simulate reaction pathways and assess the impact of reactive species. In both areas, gas-phase data are indispensable, as they provide access to intrinsic molecular properties unaffected by solvation or matrix effects.

Among experimental techniques, high-resolution rotational spectroscopy stands out for its ability to probe molecular structures with exceptional detail.[3–6] The resulting rotational constants are directly linked to the molecular moments of inertia and hence to the geometry. However, these constants correspond to vibrationally averaged structures, as measured in the ground vibrational state, and thus differ from the true equilibrium geometry. Consequently, recovering equilibrium structures from spectroscopic data requires the application of vibrational corrections, which can be derived experimentally only for small, rigid molecules. For larger or more flexible species, accurate theoretical predictions become indispensable—not only to determine

[a] *Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy. E-mail: marco.mendolicchio@sns.it*
[b] *Scuola Superiore Meridionale, Largo San Marcellino 10, 80138 Napoli, Italy*
[c] *Gaussian, Inc., Wallingford 06492, Connecticut, USA*
[d] *INSTM, Via Giuseppe Giusti 9, 50121 Firenze, Italy. E-mail: vincebarone52@gmail.com*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d5cp01611h

structural parameters but also to assist spectral assignment and interpretation, particularly in congested spectra or in the presence of multiple conformers.

Here, we present a computational protocol that delivers accurate equilibrium geometries and vibrational corrections at moderate computational cost. This integrated strategy extends the applicability of rotational spectroscopy to increasingly complex molecular systems.

Our study focuses on closed-shell, neutral molecules composed exclusively of light elements (up to the third period) and free of transition metals, ensuring that relativistic and multireference effects can be neglected. While the protocol is also applicable to open-shell[7] and charged species,[8] current high-resolution data for such systems are generally limited to small molecules. As an illustrative exception, we include the phenyl radical, which has been extensively characterized across several isotopologues. Second-order vibrational perturbation theory (VPT2) delivers reasonable vibrational corrections also for flexible molecules, provided that a well-defined equilibrium geometry exists and anharmonicity remains moderate. As a representative example, we analyze proline, a biologically relevant molecule that exemplifies internal flexibility.

For small species, spectral assignment and interpretation can often proceed with minimal computational support.[9,10] However, advances in instrumentation have extended the reach of rotational spectroscopy to molecules with up to 50 atoms,[11–13] significantly increasing the size and complexity of accessible systems. This expansion introduces challenges such as spectral congestion and the presence of several low-energy conformers, which make purely experimental analysis impractical without reliable computational guidance.

A key difficulty arises from the already mentioned mismatch between experimentally derived rotational constants and theoretical equilibrium geometries. Since the former are affected by vibrational averaging, accurate vibrational corrections are essential for meaningful structural interpretation. Yet, high-level computations of such corrections remain costly for large molecules, while experimental determinations are feasible only for a limited subset of rigid systems.

These considerations underscore the need for computational methods that balance accuracy and efficiency. Standard approaches based on density functional theory (DFT) or second-order Møller–Plesset perturbation theory (MP2) provide equilibrium geometries that are not sufficiently accurate and neglect vibrational effects,[14–16] thereby preventing unambiguous spectral assignment and interpretation. In practice, empirical adjustments or chemical intuition are frequently used to compensate for these shortcomings.

To address these challenges, we have developed the Pisa composite schemes (PCS),[17,18] a family of hierarchical protocols that combine high-level electronic structure corrections—either explicitly computed or efficiently estimated—with manageable computational cost. These schemes are tailored to the needs of rotational spectroscopy, providing accurate geometrical data for medium-sized molecules.

A key component of the protocol is the implementation of vibrational correction models within the VPT2 framework.[19,20] These models are fully automatable and computationally affordable, as the cost of computing vibrational corrections for any set of isotopologues is essentially the same as that for the parent molecule.

However, these methods risk remaining underutilized in many experimental workflows, largely due to their perceived complexity and the manual intervention they require. To bridge this gap, we introduce a fully automated, black-box workflow that integrates the Gaussian[21] and MSR (molecular structure refinement)[22] programs. This tool enables the routine prediction of spectroscopically accurate geometries and vibrational corrections with minimal user input, making advanced computational methods accessible to non-specialists.

We demonstrate the performance of this workflow through a diverse set of case studies spanning a wide range of molecular sizes and structural characteristics, from rigid frameworks to highly flexible backbones. These examples showcase the robustness and generality of our approach, and its ability to support both spectral assignment and structural refinement. Ultimately, our strategy enables the routine application of high-precision quantum chemical tools in rotational spectroscopy, at a computational cost comparable to that of conventional, less accurate methods.

2 Theory and methods

In this section, the theoretical framework underlying the calculation of vibrational averages of molecular properties will be briefly summarized. Then, the focus will shift to the development of effective computational strategies suitable for the inclusion of anharmonic effects in the calculation of vibrational averages, explicitly treating the case of a single and multiple isotopic species. Finally, the enhancement and implementation of novel procedures in the MSR code,[22] specifically devised for the calculation of accurate molecular structures through the semi-experimental (SE) approach,[9] will be the object of discussion.

2.1 Fast and accurate calculation of vibrational contributions to molecular properties

Over the last decades, different methodologies have been devised for the inclusion of anharmonic effects in the calculation of (ro-)vibrational properties, including perturbative and variational approaches, as well as their combination.[23–40] Among these, the vibrational second-order perturbation theory (VPT2) features an appealing accuracy/cost ratio, allowing for the tailoring of medium-to-large sized chemical systems.[41–48]

Let us consider a molecule characterized by N_a atoms and N normal modes of vibration, with $N = 3N_a - 6$ for non-linear molecules and $N = 3N_a - 5$ for linear ones. The starting point is the well-known Watson ro-vibrational Hamiltonian H^{vr}, which can be expanded in terms of the dimensionless normal coordinates $q = \{q_1, \ldots, q_N\}$ leading to the following expression:

$$H^{vr} = \sum_f \sum_g H_{fg} \quad (1)$$

where f and g respectively denote the order in the vibrational and rotational operators. The calculation of averages of the rotational constants involves the perturbative treatment of ro-vibrational contributions arising from eqn (1), namely H_{21}, H_{12}, H_{22} and H_{30}. Through a series of mathematical steps, a closed-form expression of the ground-state rotational constants (B_τ^0) can be derived:[49]

$$B_\tau^0 = B_\tau^{\mathrm{eq}} + \Delta B_\tau^{\mathrm{vib}} \tag{2}$$

where B_τ^{eq} and $\Delta B_\tau^{\mathrm{vib}}$, respectively, represent the equilibrium rotational constant and its vibrational correction along the same principal axis τ. The latter is composed of harmonic ($\Delta B_\tau^{\mathrm{harm}}$), Coriolis ($\Delta B_\tau^{\mathrm{Cor}}$) and anharmonic ($\Delta B_\tau^{\mathrm{anh}}$) terms,

$$\Delta B_\tau^{\mathrm{vib}} = \Delta B_\tau^{\mathrm{harm}} + \Delta B_\tau^{\mathrm{Cor}} + \Delta B_\tau^{\mathrm{anh}} \tag{3}$$

whose expressions are given below,

$$\Delta B_\tau^{\mathrm{harm}} = \left(B_\tau^{\mathrm{eq}}\right)^2 \sum_{\eta=x,y,z} \sum_{i=1}^{N} \frac{3\left(a_{i,\tau\eta}\right)^2}{4\omega_i I_\eta^{\mathrm{eq}}} \tag{4}$$

$$\Delta B_\tau^{\mathrm{Cor}} = -\left(B_\tau^{\mathrm{eq}}\right)^2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\left(\zeta_{ij,\tau}\right)^2 \left(\omega_i - \omega_j\right)^2}{\omega_i \omega_j \left(\omega_i + \omega_j\right)} \tag{5}$$

$$\Delta B_\tau^{\mathrm{anh}} = \left(B_\tau^{\mathrm{eq}}\right)^2 \left[\pi \sqrt{\frac{c}{h}} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{f_{iij} a_{j,\tau\tau}}{\omega_j^{3/2}}\right] \tag{6}$$

In the above equations $a_{i,\tau\eta}$ is the derivative of the $\tau\eta$ component of the moment of inertia tensor with respect to the q_i coordinate, I_η^{eq} is one of the principal moments of inertia, ω_i is the harmonic wavenumber (in cm^{-1}) associated with the ith normal mode, $\zeta_{ij,\tau}$ is the element of the Coriolis matrix coupling modes i and j along the τ axis, c is the speed of light in vacuum, h is the Planck constant, and the following notation has been introduced,

$$f_{ijk\ldots} = \left(\frac{\partial^n V}{\partial q_i \partial q_j \partial q_k \ldots}\right)_{\mathrm{eq}} \tag{7}$$

with V representing the potential energy.

Within the VPT2 framework, the calculation of anharmonic transition energies and intensities requires both the full set of anharmonic force constants up to semi-diagonal quartic terms (f_{ijkk}) and the derivatives of the relevant property (*e.g.*, dipole moment for infrared) up to the semi-diagonal third-order terms.[50,51] These data are typically obtained using central finite differences by displacing the molecular geometry along normal coordinates and performing a frequency calculation at each point, which also provides the property gradient.[45,52] This methodology, from now on referred to as full Hessian (FH), implies $2N + 1$ force constant computations (including the equilibrium geometry).

As can be deduced from eqn (4)–(6) and outlined in previous works,[9,19] the quantities allowing for the calculation of $\Delta B_\tau^{\mathrm{harm}}$, $\Delta B_\tau^{\mathrm{Cor}}$ can already be computed at the harmonic level, while the anharmonic term only requires the semi-diagonal cubic force constants, paving the way to computational strategies far less demanding. Vibrational corrections to the rotational constants of one or more isotopic species are required in different situations. In the former case, the computational protocol is similar to the FH method, but the evaluation of force constants is replaced by the much cheaper computation of analytical gradients. Firstly, a harmonic calculation is carried out, providing both $\Delta B_\tau^{\mathrm{harm}}$ and $\Delta B_\tau^{\mathrm{Cor}}$ contributions, together with the normal coordinates necessary to perform the displacements and finite differences. Then, at each geometry only the analytical gradient is evaluated, as this piece of information is sufficient to compute the semi-diagonal cubic force field,

$$f_{iij} = \frac{f_j(+\delta q_i) + f_j(-\delta q_i) - 2f_j(q^{\mathrm{eq}})}{\delta q_i^2} \tag{8}$$

where q^{eq} represents the equilibrium configuration, while δq_i is the displacement along the coordinate q_i. The inclusion of the terms obtained by eqn (8) within eqn (6) enables the calculation of the last contribution and then of the full vibrational corrections. Following the notation adopted in previous works, the method discussed above will be denoted as full gradient (FG).

2.2 Investigation of different isotopic species within the semi-experimental framework

As mentioned above, the $\Delta B_\tau^{\mathrm{vib}}$ contributions of several isotopic species are needed in the so-called semi-experimental (SE) approach.[9,53] Within this method, the equilibrium molecular geometry is determined through a non-linear least-squares fit of the experimental ground-state rotational constants ($B_\tau^{0,\mathrm{exp}}$) of a set of isotopologues refined by the corresponding vibrational corrections,

$$B_\tau^{\mathrm{SE}} = B_\tau^{0,\mathrm{exp}} + \Delta B_\tau^{\mathrm{vib}} \tag{9}$$

where B_τ^{SE} are known as SE rotational constants.

Once the experimental rotational constants for the isotopic species of interest have been determined, the corresponding vibrational corrections must be evaluated with efficient protocols. A widely used approach involves performing a number of FH calculations equal to the number of isotopologues. This is related to the availability of several QC packages able to compute analytical second derivatives of the energy with respect to Cartesian coordinates and then to carry out full anharmonic calculations, with the ro-vibrational analysis generated as a byproduct. However, from a computational perspective, this approach is highly inefficient when the primary focus is on ro-vibrational spectroscopic parameters. In this work, we present two computational strategies designed to provide the same information while significantly reducing computational cost. The first approach is based on the FG method and consists of two main steps. First, a harmonic calculation is performed for a single isotopic species, typically the parent molecule. Since all isotopologues share the same equilibrium geometry and Cartesian Hessian matrix, a single matrix can be systematically applied to the harmonic analysis of all isotopic species, incorporating their respective nuclear masses. As a result, a single Hessian matrix enables the calculation of the $\Delta B_\tau^{\mathrm{harm}}$ and $\Delta B_\tau^{\mathrm{Cor}}$ terms for all isotopologues.

At this point, the best course of action would be to calculate the semi-diagonal cubic force field of the parent species,

followed by its tensorial transformation to obtain the $\Delta B_\tau^{\mathrm{anh}}$ term for all other isotopic species. Unfortunately, this approach requires the full cubic force field of the original species. As demonstrated in previous work,[19,20] the removal of terms that involve three different normal modes (f_{ijk} with $i \neq j \neq k$) can cause serious problems in final vibrational corrections. Hence, the anharmonic contributions are estimated by applying the FG method to each isotopologue separately, leading in any case to a significant reduction in computational cost through the replacement of analytical force constants with gradients.

An alternative strategy relies on a single FH calculation, in which both second and third energy derivatives with respect to Cartesian coordinates are evaluated at the same time,

$$f_{rst}^{x} = \left(\frac{\partial^3 V}{\partial x_r \partial x_s \partial x_t}\right)_{\mathrm{eq}} \tag{10}$$

The purely harmonic and Coriolis contributions to $\Delta B_\tau^{\mathrm{vib}}$ can be determined following the same route discussed for the FG method. Conversely, the calculation of the semi-diagonal cubic force constants for all isotopic species is carried out through a tensorial transformation,

$$f_{iij}^{I} = \gamma_i \sqrt{\gamma_j} \sum_{r=1}^{3N_a} \sum_{s=1}^{3N_a} \sum_{t=1}^{3N_a} f_{rst}^{x} R_{ir}^{I} R_{is}^{I} R_{jt}^{I} \tag{11}$$

where $\mathbf{R}^I$ is the matrix linking the mass-weighted normal coordinates of isotopologue I in its Eckart orientation with the Cartesian coordinates of the parent species. The application of eqn (11) allows for the calculation of the anharmonic contributions for all isotopic species simultaneously. From now on, this procedure will be simply referred to as the tensorial Hessian (TH). Both the FG and TH methodologies lead to the calculation of the sought quantities with a significant reduction in computational cost with respect to the FH method.

The choice of the most effective strategy depends on the type of molecule, the number of isotopic species, the level of calculation, the computational resources, and, in particular, the possibility of parallelization. An indicative comparison of the computational costs of the different procedures is reported in Table 1.

In fact, after the evaluation of the Hessian at the equilibrium geometry, the FG method is more convenient than its TH counterpart whenever

$$\frac{C_{\mathrm{H}}}{C_{\mathrm{G}}} > N_{\mathrm{I}} \tag{12}$$

Table 1 Computational cost of the different methods employed in the present work. C_{H} and C_{G} respectively denote the cost of a single Hessian and gradient given the same computational resources, while N_{I} is the number of isotopologues and N the number of normal modes

Method	Computational cost
FH	$N_{\mathrm{I}} \cdot C_{\mathrm{H}} \cdot (2N + 1)$
FG	$C_{\mathrm{H}} + N_{\mathrm{I}} \cdot C_{\mathrm{G}} \cdot 2N$
TH	$C_{\mathrm{H}} \cdot (2N + 1)$

with C_{H} and C_{G} respectively being the cost of a single Hessian and gradient given the same computational resources, while N_{I} is the number of isotopologues.

3 Implementation and computational details

In this section, the technical aspects underlying the implementation of the procedures mentioned in Section 2.2 are discussed. In particular, a full automatization of the FG and TH approaches has been implemented in the MSR software. Finally, the QC packages and levels of calculation adopted in the present work will be briefly presented.

3.1 Overview of the MSR code

The MSR software[22] has been developed in our group and specifically designed for the calculation of accurate molecular structures through the SE approach. Over the last years, it has been employed in the characterization of different chemical systems,[54–56] including astrochemical and biological systems, as well as non-covalent complexes. In general, the minimal input required by the program includes the experimental rotational constants, the corresponding vibrational and electronic (if present) corrections, and the guessed geometry.

The optimization can be carried out through different algorithms, the default one being the Gauss–Newton method, over a set of nuclear coordinates defined by the user. In the present work, Z-matrix internal coordinates have been systematically adopted. After convergence of the optimization, in addition to reporting the final structure, a detailed error analysis is carried out. In particular, the standard deviations and confidence intervals for the different structural parameters are computed, together with the leverages and correlation coefficients. Regardless of the internal coordinates employed in the optimization process, a robust error propagation allows for the calculation of uncertainties in the desired coordinates, ranging from the canonical set of bond lengths, angles, and dihedrals to the distance between the centers of mass of different fragments. Furthermore, the Hessian matrix of the residual function is computed to confirm that the optimized geometry corresponds to a true minimum. While convergence to the absolute minimum cannot be guaranteed, different starting geometries and coordinate sets can be employed to mitigate this issue. In practice, the PCS starting geometries are sufficiently close to the SE structure that the final result typically corresponds to the absolute minimum.

More in detail, three main output files are provided:

- Main output file, containing all the information concerning the calculation;
- Summary file, containing the optimized geometry, the general trend of the optimization (which could in principle be plotted), part of the error analysis, and the rotational constants related to the different isotopologues with the corresponding uncertainties;

• A standard xyz file containing the Cartesian coordinates of the parent species in Eckart orientation, readable by any molecular editor.

3.2 Integration of the FG and TH methods within MSR

The MSR software has been enhanced with the possibility of extracting (or computing) the vibrational corrections to the rotational constants starting from QC calculations. In particular, the calculation of the semi-diagonal cubic force constants in terms of analytical gradients has been recently implemented in a development version of the Gaussian package[57] [new keyword freq = (numer,readharm) after a standard harmonic computation]. Whenever the FG method is applied, the harmonic data and the semi-diagonal cubic force constants are directly read from the resulting output files for all isotopic species and employed in the structural refinement. In this work, this protocol has been applied systematically for all the calculations based on the FG method. Furthermore, in order to make this framework fully accessible until the new commercial release of Gaussian becomes available, the framework has also been implemented in the MSR code. Since the calculation of vibrational corrections can be performed using any QC package featuring at least analytical gradients, the implementation has been designed to remain broadly applicable. For validation purposes, the MSR code has been fully interfaced with the Gaussian 16 QC package.[21] Starting from a formatted checkpoint file corresponding to the harmonic calculations, the normal coordinates are used to build the displaced geometries and define the corresponding single-point calculations. Then, MSR is run again to extract the data from the Gaussian output file and build the anharmonic force constants for isotopic species of interest.

The integration of the TH method is simpler, since it requires the calculation of the second and third derivatives of the energy with respect to Cartesian coordinates at the equilibrium geometry (freq = cubic keyword of Gaussian). Then, the vibrational corrections are computed by MSR through the tensorial transformation introduced in eqn (11).

Let us underline that the new developments have been fully integrated with the calculation of the SE structure, so that a single run of MSR allows for the calculation of the vibrational corrections, their use in the optimization process, and the determination of the structural parameters. For the sake of completeness, a diagram representing the whole engine is sketched in Fig. 1.

3.3 Reduced-dimensionality calculations

Although it is known that the SE method provides remarkably accurate structural parameters,[9,58] its success strongly depends on the availability of sufficient experimental data. Unfortunately, in several cases, the number of isotopic species investigated at the experimental level is not large enough to allow for a full structural characterization, making the development of strategies to handle this (not uncommon) scenario essential.

In general, this issue can be addressed by removing ill-defined parameters from the fit and fixing them at specified

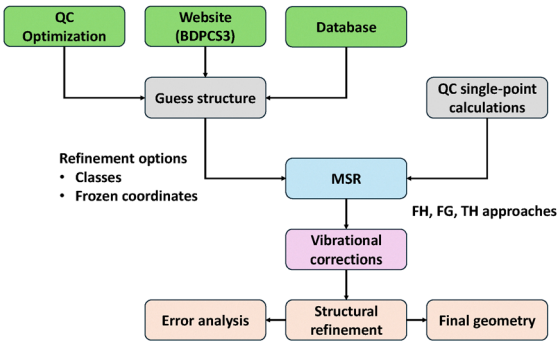


Fig. 1 Workflow describing the whole engine for the structural refinement developed in this work.

values, usually obtained through refined QC calculations. A viable alternative is the method of predicate observations (also known as mixed regression),[59] in which accurate estimates of structural parameters and their uncertainties are included in the fit as additional sources of information. While the MSR code has incorporated these procedures since its inception, this work explores a new method that provides valuable support, particularly in ill-conditioned fits.

Previous studies have shown that different types of coordinates (*e.g.*, C–H bond lengths) exhibit systematic errors,[56,58,59] paving the way for methods that refine structural parameters through corrections derived from linear regressions. Therefore, a whole set of related coordinates can be optimized using a single parameter. More specifically, the parameters in question share the same correction at each step of the optimization while preserving different initial guesses and final values.

From a mathematical perspective, this is equivalent to imposing that the difference between one of these parameters and the others in the set remains equal to its initial guess value.

3.4 Quantum chemical methods

The family of Pisa composite schemes (PCS) was recently developed in order to optimize the price/precision ratio for molecules of different sizes.[17] For molecules containing a few dozen atoms, the cost increases roughly by factors of 1, 10, 10^3 for the computation of the geometry, harmonic frequencies, and anharmonic contributions with a given method. At the same time, the cost of hybrid density functionals with double-zeta basis sets, double hybrid density functionals with triple-zeta basis sets, and explicitly correlated (F12) coupled cluster methods[60] including single, double, and (perturbatively) triple excitations [CCSD(T)][61] in conjunction with double-zeta basis sets increases by 1, 10, 10^3. Therefore, the standard hierarchy of PCS variants employed for the different contributions aims to reach the best accuracy/cost ratio, progressively reducing the computational level for calculations providing decreasing contributions to the overall result, but requiring increasing computational resources.

The cheapest variant (referred to as HPCS2 and used in the computation of anharmonic contributions) corresponds to a hybrid functional (B3LYP) paired with the 6-31+G* basis set,

and complemented by empirical dispersion (D4).[62] A more accurate variant (referred to as DPCS3 and used for the evaluation of harmonic force fields) corresponds to the revDSD-PBEP86-D4 double hybrid functional[63] in conjunction with the 3F12$^-$ basis set,[18] which is obtained from its standard cc-pVTZ-F12 (3F12) counterpart[64] by removing *d* functions on first-row atoms and replacing the two *f* functions on second- and third-row atoms by a single *f* function taken from the cc-pVTZ basis set.[65] Note that the D3BJ model for empirical dispersions[66] employed in previous PCS versions has now been replaced by the more accurate D4 model,[62] which has been implemented in a development version of the Gaussian software.[57] Although this modification has a negligible impact on the molecular parameters of interest in the present work, it should allow a better reproduction of intermolecular interactions. At the same time, other quantum chemical models have been implemented in the Gaussian software (HF-3c,[67] r^{2}SCAN-3c,[68] ...) and tested as alternatives to the quite old B3LYP hybrid functional for the computation of anharmonic force fields and equilibrium geometries. Unfortunately, none of the tested methods improved the B3LYP results and some of them actually provided unphysical results for some molecules (*e.g.*, HF-3c for molecules containing oxygen atoms). As a consequence, we have retained the HPCS2 and DPCS3 variants used in previous works, except for the use of D4 empirical dispersions.

The most accurate PCS variant (referred to as PCS2 and used in geometry optimizations) is based on the CCSD(T)-F12b ansatz[60] (CC-F12) in conjunction with the cc-pVDZ-F12 (2F12) basis set[18,69] to evaluate valence contributions for molecules containing first- and second-row atoms. For third-row atoms, an additional *f* function (taken from the cc-pVTZ basis set) is needed to obtain comparable errors, with the resulting basis set being referred to as 2F12$^+$.[17] A benchmark study[70] showed that further extension of the basis set has a negligible effect provided that the contribution of the complementary auxiliary basis set (CABS) is taken into account in the Hartree–Fock (HF) component.[71,72] On the other hand, the conventional MP2[73] model in conjunction with the cc-pwCVTZ[74] (wC3) basis set is sufficient for evaluating the contribution of core–valence (CV) correlation.

$$\Delta r^{\mathrm{CV}} = r^{(\text{ae-MP2/wC3})} - r^{(\text{fc-MP2/wC3})} \tag{13}$$

Therefore, the PCS2 geometrical parameters are given by

$$r^{\mathrm{PCS2}} = r^{\text{CC-F12}} + \Delta r^{\mathrm{CV}} \tag{14}$$

Equilibrium rotational constants require an accuracy on the order of 1 mÅ for bond lengths and 0.1 degrees for valence angles,[18,75] which is attained by the PCS2 variant. However, the computational cost of this approach becomes prohibitive for molecules with more than 20 atoms. To address this limitation, a more efficient variant—denoted BDPCS3—has been developed. This replaces the coupled-cluster treatment with the DPCS3 scheme described above and estimates core–valence (CV) correlation effects on bond lengths using the following one-parameter expression:

$$\Delta r_{ij}^{\mathrm{BCV}} = -k[1 + 1.1\delta(i,C)\delta(j,H)]\sqrt{N_iN_j - 1}\left(r_i^{\mathrm{cov}} + r_j^{\mathrm{cov}}\right) \tag{15}$$

where $N = \min(n_i,3)$, r_i^{cov}, r_j^{cov} are covalent radii from ref. 76, and the optimized value of k is 0.0011. The second term in square brackets has been introduced in this work following a systematic analysis of an extensive set of CH bond lengths (see section Results and discussion). Minor valence corrections ($\Delta r_{ij}^{\mathrm{VB}}$) are also applied to mitigate the slight over-delocalization typical of semi-local density functionals:

$$\Delta r_{ij}^{\mathrm{BV}} = \Delta r_{ij}^{\mathrm{BCV}}\left[\sqrt{|P_{ij}-2|}-1\right]\left[\delta(i,S)\delta(j,C) + \sum_{X,Y\in C,N}\delta(i,X)\delta(j,Y)\right]$$

In the above expressions, Kronecker δ's are used to restrict the corrections to specific bond types, and, by convention, the atomic number of atom i is greater than or equal to that of atom j. P_{ij} denotes the Pauling bond order, which is used to identify bonded pairs when its value exceeds 0.3:

$$P_{ij} = \exp[(r_i^{\mathrm{cov}} + r_j^{\mathrm{cov}} - r_{ij})/0.3] \tag{16}$$

The final BDPCS3 bond lengths are given by:

$$r^{\mathrm{BDPCS3}} = r^{\mathrm{DPCS3}} + \Delta r_{ij}^{\mathrm{BCV}} + \Delta r_{ij}^{\mathrm{BV}} \tag{17}$$

and can be obtained from their DPCS3 counterparts by a dedicated website (**https://www.skies-village.it/proxima/pcsbonds/**).[77] The characteristics of the different PCS variants are summarized in Table 2.

All the quantum chemical computations have been performed with the Gaussian package,[21,57] except the CC-F12 ones, which have been performed with the Molpro program.[79] The standard convergence thresholds for geometry optimizations (maximum gradient $< 4.5 \times 10^{-3}$ a.u. and maximum step $< 1.8 \times 10^{-3}$ a.u., with RMS values of 3×10^{-3} a.u. (gradient) and 1.2×10^{-3} a.u. (step) are too large both for the sought spectroscopic accuracy and the computation of converged vibrational corrections, especially in the presence of soft modes. Therefore, tighter criteria (opt = tight keyword of Gaussian, *i.e.*, max gradient 1.5×10^{-3} a.u., gradient RMS 1.0×10^{-5} a.u.; max step 6×10^{-5} a.u., step RMS 4×10^{-5} a.u.) have been always

Table 2 The different PCS variants used in the present paper. See main text for further details

Label	Valence (fc)	Core–valence (ae–fc)	Δ valence (fc)
PCS2	CCSD(T)-F12b[60]/2F12	MP2/wC3 (see eqn (13))	None
DPCS3	rev-DSD-PBEP86[63]/3F12-	None	D4
BDPCS3	rev-DSD-PBEP86[63]/3F12-	BCV (see eqn (15))	D4 + BV (see eqn (16))
HPCS2	B3LYP/6-31+G* (ref. 78)	None	D4

used. In the same vein, a pruned (99 590) grid is used for numerical integrations (about 15 000 points per atom) and a pruned (75 302) grid (about 7000 points per atom) in the CPHF step, together with RMS convergence thresholds for the density matrix of 10^{-8} in the SCF step and 10^{-10} in the CPHF step.

4 Results and discussion

4.1 Analysis of XH bonds

Experimental rotational spectra have been reported for the isotopologues of several molecules, with the exception of deuterated species. This makes it particularly important to estimate accurate XH bond lengths from quantum chemical computations. The SE111 dataset[80] contains the SE equilibrium values of 286 different CH bond lengths, which allow for an unbiased evaluation of possible systematic errors affecting the different PCS variants. The histograms of the deviations of computed CH bond lengths from their SE counterparts are shown in Fig. 2 for the PCS2, DPCS3, and HPCS2 variants. The PCS2 variant displays a very narrow Gaussian distribution centered at 0.0003 Å and with a standard deviation of 0.0008 Å, which points out the accuracy of the model and the lack of significant systematic errors. The largest negative deviation (−0.0027 Å) is observed for the CH bonds of *cis*-1-chloro-2-fluoroethylene, while the largest positive deviation (0.0029 Å) is observed for the CH bonds of pyruvic acid.

The average difference between DPCS3 and SE values is 0.0027 Å and the differences follow a Gaussian distribution with a standard deviation of 0.0007 Å, demonstrating the systematic nature of the error source. The largest positive bond length deviation (0.0052 Å) is observed for pyruvic acid and the largest negative deviation (−0.0003 Å) is observed for 1-chloro-2-fluoroethene. Therefore, a constant shortening of DPCS3 bond lengths rounded to 0.0025 Å provides results with spectroscopic accuracy for molecules with a large variety of bond patterns.

Finally, the HPCS2 variant is characterized by a much larger systematic error (0.0064 Å), and the noise is also higher, with a standard deviation of 0.0010 Å. These positive values indicate the tendency of HPCS2 to overestimate bond lengths. The largest deviation from SE values (0.0092 Å) is observed for the CH bond of pyruvic acid, while the smallest deviation (0.0032 Å) is observed for the CH bonds of the imidazole system.

In conclusion, only the PCS2 variant provides CH bond lengths with spectroscopic accuracy, but a systematic correction of DPCS3 values (as done in the BDPCS3 model) leads to comparable accuracy at DFT cost. On the other hand, much larger error bars affect HPCS2 values irrespective of systematic corrections. As the systematic overestimation of CH bond lengths at the DPCS3 level is more than twice the current BDPCS3 correction (0.0012 Å), the second term in square bracket of eqn (15) has been added in the new release of the BDPCS3 model. Although the BDPCS3 correction is possibly too small also for other XH bond lengths, the number of available SE structures containing those bonds is too small to permit an unbiased revision of the corresponding BDPCS3 corrections.

4.2 Validation of the computational protocols

The next step is the validation of the proposed protocols with reference to molecules for which the availability of a large number of isotopologues permits obtaining accurate SE equilibrium structures optimizing all the geometrical parameters.

The first example is thiophene (see Fig. 3) for which very accurate equilibrium geometry, vibrational, and electronic corrections are available, together with the experimental ground state rotational constants of a large number of isotopologues.[81] The main results of our analysis are collected in Table 3. Of course, optimization of all the geometrical parameters reproduces exactly the results of ref. 81. Already, BDPCS3 and PCS2 equilibrium geometries are in remarkable agreement with their SE counterpart, and optimization of the positions of heavy atoms (freezing the bond lengths and valence angles of hydrogen atoms at their PCS2 or BDPCS3 values) provides a final structure where all the geometrical parameters are within the error bars of their full SE counterparts. At the same time, the relative errors of the rotational constants are reduced by about one order of magnitude when going from purely theoretical equilibrium structures (either PCS2 or BDPCS3) to their counterparts issued from the refinement of the positions of heavy atoms. While optimization of the positions of hydrogen atoms further reduces the errors on the rotational constants, the effect is quite small and, above all, well within the error bar of vibrational corrections.

The second example is that of cyclopentadiene (see Fig. 3), whose equilibrium SE structure has been recently obtained from the experimental rotational constants and fc-CCSD(T)/ANO0

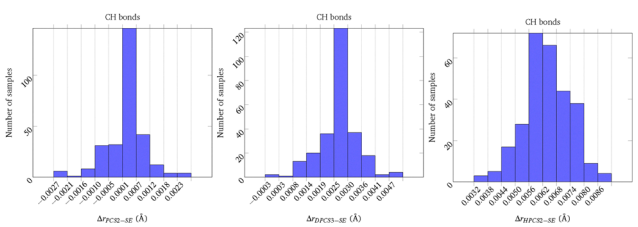
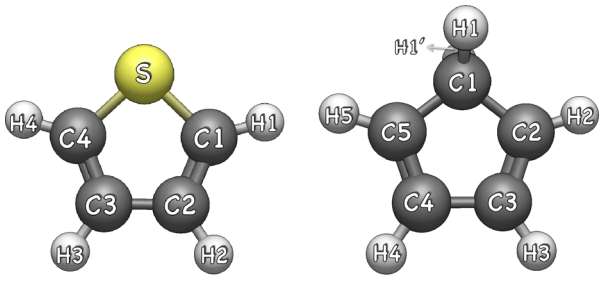


Fig. 2 Distributions of differences between computed (PCS2, DPCS3, or HPCS2) and semi-experimental equilibrium CH bond lengths for the molecules contained in the SE111 database.



Thiophene Cyclopentadiene (C_{2v})

Fig. 3 Structure and labeling of thiophene and cyclopentadiene.

Table 3 Comparison between SE and computed bond lengths (in Å) and angles (in degrees) of thiophene. The labeling of geometrical parameters follows the atom numbering of Fig. 3

	r_{eq}				$r_{eq}^{SE\,a,b}$		
Parameter	HPCS2	DPCS3	BDPCS3	PCS2[c]	BDPCS3[d]	PCS2[e]	Ref. 81
$r(S–C_1)$	1.7354	1.7137	1.7103	1.7102	1.7104(2)	1.7103(1)	1.71049(18)
$r(C_1–C_2)$	1.3705	1.3698	1.3680	1.3663	1.3654(3)	1.3655(2)	1.36564(31)
$r(C_1–H_1)$	1.0819	1.0795	1.0770	1.0765	1.0770✓	1.0765✓	1.07714(17)
$r(C_2–H_2)$	1.0847	1.0816	1.0791	1.0788	1.0791✓	1.0788✓	1.07856(14)
$\vartheta(C_4–S–C_1)$	91.50	92.10	92.10	92.06	92.02(1)	92.04(1)	92.047(15)
$\vartheta(S–C_1–C_2)$	111.60	111.54	111.54	111.61	111.630(7)	111.632(5)	111.608(16)
$\vartheta(H_1–C_1–S)$	120.31	120.33	120.33	120.30	120.33✓	120.30✓	120.065(28)
$\vartheta(H_2–C_2–C_1)$	123.35	123.37	123.38	123.44	123.38✓	123.44✓	123.414(23)
MAX%[f]	2.0302	0.4273	0.0878	0.0290	0.0104	0.0042	0.0023
MAE%[g]	1.8327	0.4161	0.0468	0.0368	0.0020	0.0014	0.0005

[a] The symbol ✓ denotes structural parameters kept fixed at their guess value. [b] Vibrational and electronic corrections at the CCSD(T)/cc-pCVTZ level of theory taken from ref. 81. [c] Data taken from the molecular database[70] (https://www.skies-village.it/databases/). [d] SE structure obtained starting from the BDPCS3 geometry. [e] SE structure obtained starting from the PCS2 geometry. [f] Maximum absolute percentage error in the rotational constants (for all isotopic species) of the refined structure relative to their experimental counterparts. [g] Mean absolute percentage error in the rotational constants (for all isotopic species) of the refined structure relative to their experimental counterparts.

vibrational corrections.[82] Also in this case, refinement of all the geometrical parameters by the MSR code reproduces exactly the results of the original work (see Table 4).

Furthermore, the results obtained freezing the relative positions of hydrogen atoms at their BDPCS3 or PCS2 values are even better than those issued from the complete optimization, possibly because of the well known numerical instabilities often affecting the optimization of hydrogen atom positions.[59]

As a further example we have considered *cis*-hexatriene (see Fig. 4), which was previously investigated by Demaison and coworkers.[83]

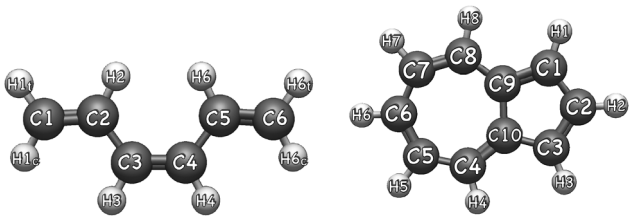
The results collected in Table 5 show that PCS2 and BDPCS3 provide more accurate reference values with respect to those employed in the mixed regression approach of ref. 83. Once again, the relative positions of hydrogen atoms can be safely frozen at their computed values while optimizing the positions of heavy atoms in a reduced-dimensionality SE approach.

The situation is slightly different in the case of azulene (see Fig. 4), as the optimization of all the geometrical parameters faces numerical instabilities, and only a stepwise procedure with some frozen parameters allowed obtaining a reasonable solution.[55] Under such circumstances, freezing the positions of hydrogen atoms to their BDPCS3 equilibrium values guarantees

Table 4 Comparison between SE and computed bond lengths (in Å) and angles (in degrees) of cyclopentadiene. The labeling of geometrical parameters follows the atom numbering of Fig. 3

	r_{eq}				$r_{eq}^{SE\,a,b}$				
Parameter	HPCS2	DPCS3	BDPCS3	PCS2[c]	BDPCS3[d]	RD(BDPCS3)[e]	PCS2[f]	RD(PCS2)[g]	Ref. 82
$r(C_1–C_2)$	1.5070	1.5016	1.4988	1.5005	1.4982(7)	1.49892(2)	1.4995(2)	1.499790(4)	1.49906(34)
$r(C_2–C_3)$	1.3521	1.3479	1.3466	1.3468	1.3465(6)	1.34672(2)	1.3461(2)	1.346090(4)	1.34635(35)
$r(C_1–H_1)$	1.1011	1.0969	1.0944	1.0940	1.0944✓	1.0944✓	1.0940✓	1.0940✓	1.09441(17)
$r(C_2–H_2)$	1.0841	1.0812	1.0787	1.0787	1.0787✓	1.0787✓	1.0787✓	1.0787✓	1.07865(24)
$r(C_3–H_3)$	1.0853	1.0820	1.0795	1.0795	1.0795✓	1.0795✓	1.0795✓	1.0795✓	1.07907(22)
$\vartheta(C_5–C_1–C_2)/2$	51.64	51.59	51.59	51.54	51.56(3)	51.5700(6)	51.528(9)	51.5258(5)	51.541(15)
$\vartheta(C_1–C_2–C_3)$	109.07	109.24	109.24	109.32	109.34(4)	109.2700(6)	109.32(1)	109.3048(5)	109.330(24)
$\vartheta(H_2–C_2–C_1)$	124.02	124.11	124.11	124.11	124.11✓	124.11✓	124.11✓	124.11✓	124.01(11)
$\vartheta(H_3–C_3–C_2)$	126.27	126.14	126.14	126.13	126.13✓	126.13✓	126.13✓	126.13✓	126.082(60)
$\vartheta(H_1–C_1–H_1')$	106.04	106.40	106.40	106.60	106.40✓	106.40✓	106.60✓	106.60✓	106.580(12)[h]
MAX%[i]	0.9744	0.3957	0.1970	0.1120	0.0100	0.0118	0.0033	0.0039	0.0772
MAE%[j]	0.9744	0.2775	0.1050	0.0608	0.0045	0.0050	0.0014	0.0016	0.0660

[a] The symbol ✓ denotes structural parameters kept fixed at their guess value. [b] Vibrational corrections at the fc-CCSD(T)/ANO0 level of theory taken from ref. 82. [c] Data taken from the molecular database[70] (https://www.skies-village.it/databases/). [d] SE structure obtained starting from the BDPCS3 geometry, taken from ref. 82. [e] SE structure obtained starting from the BDPCS3 geometry, taken from ref. 82, and refined for the C–C and C–C–C classes. [f] SE structure obtained starting from the PCS2 geometry, taken from ref. 82. [g] SE structure obtained starting from the PCS2 geometry, taken from ref. 82, and refined for the C–C and C–C–C classes. [h] Parameter derived from the SE structure reported in ref. 82. [i] Maximum absolute percentage error in the rotational constants (for all isotopic species) of the refined structure relative to their experimental counterparts. [j] Mean absolute percentage error in the rotational constants (for all isotopic species) of the refined structure relative to their experimental counterparts.



Cis-hexatriene Azulene

Fig. 4 Structure and labeling of *cis*-hexatriene and azulene.

the stability of the optimization process and provides very reliable results (see Table 6).

As previously noted in the Introduction, our computational protocol can be applied to open-shell species[7] and charged systems,[8] both of which are highly relevant in astrochemistry. However, currently available experimental data for such species are mostly limited to very small molecules, which fall outside the main scope of this study. As an illustrative exception, we include the phenyl radical, for which experimental data are available for a sufficient number of isotopologues to enable the determination of an accurate SE equilibrium structure (Table 7).[84]

The data clearly show that vibrational corrections computed at the HPCS2 level lead to SE structures that closely match those obtained using the more expensive DPCS3 and even CCSD(T)/ANO0 approaches. In all cases, the results are well converged, and the final errors are comparable to those observed for closed-shell systems. However, the accuracy of the QC equilibrium geometries is lower than usual, due to the known moderate multi-reference character of the phenyl radical.[84,85] In the case of PCS2, this effect primarily affects the C_i–C_o bond length and can be mitigated through a localized correction, as previously demonstrated for other multi-reference systems.[86,87] Conversely, the DPCS3 variant tends to underestimate all bond lengths involving atoms with significant spin density.[7] Although a general reparametrization of the method for open-shell species is currently hindered by the lack of accurate benchmark geometries—either SE equilibrium structures or computational results of at least CCSDT(Q) quality[10] for a sizeable set of radicals—reasonable results can be obtained for the phenyl radical by applying an empirical offset of 5 mÅ to the DPCS3 equilibrium values of all the CC bond lengths.

4.3 Partial SE equilibrium structures

In this section, we discuss the case of some molecules for which isotopic substitutions are available only for heavy atoms and were used in the original works to obtain partial substitution structures (r_s).[9]

Cyanides and isocyanides have attracted considerable attention in the field of astrochemistry as proxies of their hydrocarbon analogues, which are microwave silent due to vanishing dipole moments. After systematic investigations of polyaromatic cyanides and isocyanides,[88–93] branched carbon-chain analogues are now under extensive investigation.[94,95] Among isocyanides, experimental rotational constants are available for all the isotopologues of *tert*-butyl isocyanide (see Fig. 5) involving single substitutions of heavy atoms.[96]

Due to the high symmetry of the molecule (C_{3v}) the number of independent parameters is quite limited and, above all, only two values are needed for CH bond lengths and two others for HCC valence angles. The results obtained from different quantum chemical models are compared in Table 8 with those issued from partial SE optimizations in which the CH bond

Table 5 Comparison between SE and computed bond lengths (in Å) and angles (in degrees) of *cis*-hexatriene. The labeling of geometrical parameters follows the atom numbering of Fig. 4

	r_{eq}				$r_{eq}^{SE\,a,b}$		
Parameter	HPCS2	DPCS3	BDPCS3	PCS2[c]	Ref. 83	BDCPS3[d]	PCS2[e]
$r(C_1–C_2)$	1.3457	1.3408	1.3996	1.3401	1.3400(4)	1.3397(2)	1.3394(2)
$r(C_2–C_3)$	1.4524	1.4533	1.4508	1.4525	1.4512(4)	1.4511(2)	1.4511(1)
$r(C_3–C_4)$	1.3568	1.3506	1.3492	1.3493	1.3488(9)	1.3481(2)	1.3492(2)
$r(C_1–H_1c)$	1.0883	1.0849	1.0824	1.0827	1.0826(3)	1.0824✓	1.0827✓
$r(C_1–H_1t)$	1.0859	1.0826	1.0801	1.0800	1.0798(2)	1.0801✓	1.0800✓
$r(C_2–H_2)$	1.0883	1.0850	1.0825	1.0823	1.0825(8)	1.0825✓	1.0823✓
$r(C_3–H_3)$	1.0897	1.0869	1.0844	1.0845	1.0841(5)	1.0844✓	1.0845✓
$\vartheta(C_1–C_2–C_3)$	123.52	122.91	122.91	122.75	122.77(3)	122.92(2)	122.86(2)
$\vartheta(C_2–C_3–C_4)$	126.80	126.40	126.40	126.22	126.28(2)	126.29(1)	126.27(2)
$\vartheta(H_1c–C_1–C_2)$	121.45	121.12	121.12	120.98	121.02(2)	121.12✓	120.98✓
$\vartheta(H_1t–C_1–C_2)$	121.69	121.51	121.51	121.48	121.46(2)	121.51✓	121.48✓
$\vartheta(H_2–C_2–C_1)$	118.57	118.78	118.78	118.90	118.91(9)	118.78✓	118.90✓
$\vartheta(H_3–C_3–C_4)$	117.70	117.91	117.91	118.05	118.02(7)	117.91✓	118.05✓
MAX%[f]	1.6068	0.4469	0.3385	0.1550	0.1536	0.0086	0.0052
MAE%[g]	1.0007	0.2419	0.2585	0.0795	0.0817	0.0022	0.0020

[a] The symbol ✓ denotes parameters kept fixed at their guess value. [b] Vibrational corrections at the HPCS2 level of theory. [c] Data taken from the molecular database[70] (https://www.skies-village.it/databases/). [d] SE structure obtained starting from the BDPCS3 geometry. [e] SE structure obtained starting from the PCS2 geometry. [f] Maximum absolute percentage error in the rotational constants (for all isotopic species) of the refined structure relative to their experimental counterparts. [g] Mean absolute percentage error in the rotational constants (for all isotopic species) of the refined structure relative to their experimental counterparts.

Table 6 Comparison between SE and computed bond lengths (in Å) and angles (in degrees) of azulene. The labeling of geometrical parameters follows the atom numbering of Fig. 4

	r_{eq}				$r_{eq}^{SE\,a,b}$				
Parameter	HPCS2[c]	DPCS3[c]	BDPCS3	PCS2[c]	Ref. 55	RD1(BDCPS3)[d]	RD2(BDCPS3)[e]	RD1(PCS2)[f]	RD2(PCS2)[g]
$r(C_1–C_9)$	1.4077	1.4035	1.4014	1.4012	1.402(4)	1.396(2)	1.401151(9)	1.399(2)	1.401200(7)
$r(C_4–C_{10})$	1.3917	1.3883	1.3864	1.3872	1.379(7)	1.389(3)	1.386171(9)	1387(3)	1.387200(7)
$r(C_4–C_5)$	1.3985	1.3950	1.3930	1.3939	1.399(4)	1.394(2)	1.392811(9)	1.395(2)	1.393900(7)
$r(C_5–C_6)$	1.3987	1.3949	1.3929	1.3940	1.393(3)	1.3933(7)	1.392701(9)	1.3936(6)	1.394000(7)
$r(C_1–H_{11})$	1.0837	1.0811	1.0786	1.0790	1.079✓	1.0786✓	1.0786✓	1.0790✓	1.0790✓
$r(C_2–H_{12})$	1.0849	1.0824	1.0799	1.0800	1.080✓	1.0799✓	1.0799✓	1.0800✓	1.0800✓
$r(C_4–H_{14})$	1.0899	1.0877	1.0852	1.0854	1.087(2)	1.0852✓	1.0852✓	1.0854✓	1.0854✓
$r(C_5–H_{15})$	1.0877	1.0850	1.0825	1.0827	1.082(1)	1.0825✓	1.0825✓	1.0827✓	1.0827✓
$r(C_6–H_{16})$	1.0889	1.0862	1.0837	1.0840	1.081(2)	1.0837✓	1.0837✓	1.0840✓	1.0840✓
$\vartheta(C_1–C_9–C_{10})$	106.57	106.64	106.65	106.64	106.5(3)	106.67(2)	106.6653(2)	106.65(2)	106.6426(2)
$\vartheta(C_4–C_5–C_6)$	128.66	128.68	128.68	128.68	128.68✓	128.70(2)	128.6990(2)	128.68(2)	128.6746(2)
$\vartheta(C_5–C_4–C_{10})$	129.03	128.80	128.81	128.70	128.70✓	128.83(2)	128.8252(2)	128.69(2)	128.6896(2)
$\vartheta(C_5–C_6–C_7)$	129.83	129.90	129.90	129.98	129.9(1)	128.70(2)	128.6989(2)	128.68(2)	128.6746(2)
$\vartheta(C_1–C_2–H_{12})$	125.08	125.07	125.07	125.08	125.08✓	125.07✓	125.07✓	125.08✓	125.08✓
$\vartheta(C_6–C_5–H_{15})$	115.68	115.61	115.61	115.59	115.7(2)	115.61✓	115.61✓	115.59✓	115.59✓
$\vartheta(C_9–C_1–H_{11})$	125.17	125.14	125.15	125.09	125.09✓	125.15✓	125.15✓	125.09✓	125.09✓
$\vartheta(C_{10}–C_4–H_{14})$	115.27	115.32	115.31	115.32	115.32✓	115.31✓	115.31✓	115.32✓	115.32✓
MAX%[e]	0.9986	0.2996	0.1568	0.0701	0.0081	0.0066	0.0086	0.0062	0.0062
MAE%[f]	0.7941	0.2341	0.0786	0.0366	0.0024	0.0020	0.0030	0.0019	0.0021

[a] The symbol ✓ denotes parameters kept fixed at their guess value. [b] Vibrational corrections at the HPCS2 level of theory. [c] Data taken from ref. 55. [d] SE structure obtained starting from the BDPCS3 geometry and optimizing the C–C–C classes. [e] SE structure obtained starting from the BDPCS3 geometry and optimizing the C–C and C–C–C classes. [f] SE structure obtained starting from the PCS2 geometry and optimizing the C–C–C classes. [g] SE structure obtained starting from the PCS2 geometry and optimizing the C–C and C–C–C classes.

lengths and HCC valence angles are frozen at their PCS2 or BDPCS3 values.

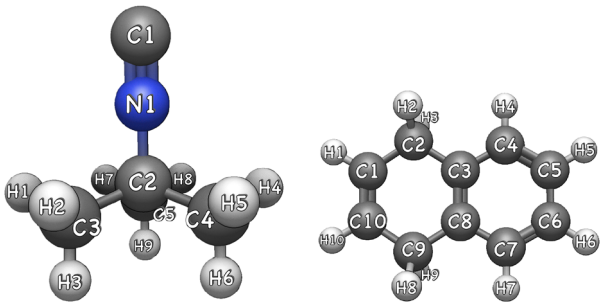
The remarkable accuracy of the BDPCS3 and PCS2 equilibrium geometries is well evidenced by the small deviations for all the available rotational constants, but in this case, the sought spectroscopic accuracy is reached only at the PCS2 level. Optimization of the relative positions of heavy atoms by a reduced-dimensionality SE approach leads to a slight increase in the N1C2 bond length and a corresponding shortening of the C2C3 bond length, while the C1N1 bond length and the N1C2C3 valence angle are only negligibly affected. Although this trend points out some residual problems in the balanced description of inductive and delocalization effects at the BDPCS3 level, the final SE equilibrium structures employing either PCS2 or BDPCS3 values for the hydrogen positions are nearly indistinguishable and reproduce all the available rotational constants with very small residual errors. Therefore, those geometrical parameters represent significant improvements with respect to those obtained for the so-called substitution structure (r_s) in ref. 96.

Another interesting case is that of 1,4-dehydronaphtalene (see Fig. 5), which has been recently investigated.[16] The results

Table 7 Comparison between SE and computed bond lengths (Å) and angles (degrees) of the phenyl radical

	r_{eq}				r_{eq}^{SE}		
Parameter[a]	HPCS2	DPCS3	BDPCS3	PCS2[b]	HPCS2 Δvib	DPCS3 Δvib	CCSD(T) Δvib[c]
$r(C_i–C_o)$	1.3784	1.3675	1.3725[d]	1.3745	1.3726(7)	1.3728(7)	1.3722(3)
$r(C_o–C_m)$	1.4060	1.3951	1.4001[d]	1.3999	1.3989(3)	1.3990(3)	1.3984(1)
$r(C_m–C_p)$	1.3988	1.3879	1.3929[d]	1.3932	1.3936(3)	1.3937(2)	1.3931(1)
$r(C_o–H)$	1.0866	1.0829	1.0804	1.0809	1.0799(2)	1.0799(2)	1.0808(1)
$r(C_m–H)$	1.0875	1.0844	1.0819	1.0820	1.0811(2)	1.0813(2)	1.0819(1)
$r(C_p–H)$	1.0865	1.0833	1.0806	1.0811	1.0802(2)	1.0804(2)	1.0809(1)
$\vartheta(C_o–C_i–C_o)$	125.86	126.18	126.18	125.73	125.87(3)	125.82(3)	125.85(1)
$\vartheta(C_i–C_o–C_m)$	116.57	116.38	116.38	116.61	116.59(3)	116.62(3)	116.60(1)
$\vartheta(C_o–C_m–C_p)$	120.18	120.28	120.28	120.16	120.17(2)	120.15(2)	120.16(1)
$\vartheta(C_m–C_p–C_m)$	120.63	120.57	120.57	120.66	120.62(2)	120.63(2)	120.63(1)
$\vartheta(C_i–C_o–H)$	122.44	122.40	122.40	122.36	122.97(3)	122.96(3)	122.90(1)
$\vartheta(C_o–C_m–H)$	119.63	120.13	120.13	120.19	120.41(3)	120.44(3)	120.39(1)
$\vartheta(C_m–C_p–H)$	119.69	119.59	119.59	119.65	119.69(1)	119.68(1)	119.68(1)

[a] The symbols C_i, C_o, C_m, and C_p refer to *ipso*, *ortho*, *meta*, and *para* carbon atoms, respectively. [b] Data taken from the PCS141 database[70] (https://www.skies-village.it/databases/). [c] Data taken from ref. 84. [d] DPCS3 value augmented by 0.005 Å (see main text for details).



t-butyl-isocyanate 1,4-dihydronaphtalene

Fig. 5 Structure and labeling of *t*-butyl-isocyanate and 1,4-dihydronaphtalene.

collected in Table 9 show that the deviations of BDPCS3 results are one order of magnitude smaller than their B3LYP counterparts reported in ref. 16. Actually, only a slight improvement is brought by the further optimization of heavy atom positions by the SE approach.

In this connection, it is remarkable that very accurate results are obtained by collecting all carbon atoms in a single class (*i.e.*, optimizing only two parameters, one for bond lengths and one for valence angles). In any case, the rotational constants of all the available isotopologues are reproduced with remarkable accuracy (see Table 10).

To further evaluate the performance of the methodology, we have considered a flexible molecule, namely the most stable conformer of proline (see Fig. 6).

The results collected in Table 11 show that, as expected, the errors are larger than for rigid molecules, possibly due to both limitations of the VPT2 model and of the underlying quantum chemical computations. In any case, the accuracy of the final BDPCS3 geometrical parameters is comparable to that of their PCS2 counterparts, while the computational cost is reduced by at least two orders of magnitude. Furthermore, the reduced-dimensionality SE equilibrium structures obtained by freezing the positions of hydrogen atoms at their BDPCS3 or PCS2 values do not show any numerical instability, while delivering an accuracy comparable (if not better) than that of the structure obtained in ref. 97 through a mixed regression approach based on a CCSD(T) geometry optimization.

Table 8 Comparison between SE and computed bond lengths (in Å) and angles (in degrees) of *t*-butyl isocyanide. The labeling of geometrical parameters follows the atom numbering of Fig. 4

	r_{eq}				$r_{eq}^{SE\,a,b}$	
Parameter	HPCS2	DPCS3	BDPCS3	PCS2	BDPCS3[c]	PCS2[d]
$r(C_1–N_1)$	1.1751	1.1730	1.1707	1.1712	1.1703(4)	1.1703(4)
$r(N_1–C_2)$	1.4429	1.4413	1.4386	1.4406	1.439(1)	1.439(1)
$r(C_2–C_3)$	1.5379	1.5289	1.5259	1.5255	1.5249(4)	1.5251(4)
$r(C_3–H_1)$	1.0945	1.0911	1.0885	1.0887	1.0885✓	1.0887✓
$r(C_3–H_3)$	1.0952	1.0919	1.0894	1.0895	1.0894✓	1.0895✓
$\vartheta(N_1–C_2–C_3)$	108.08	107.98	107.98	108.02	108.02(5)	108.02(5)
$\vartheta(C_2–C_3–H_1)$	110.67	110.58	110.58	110.52	110.58✓	110.52✓
$\vartheta(C_2–C_3–H_3)$	109.57	109.47	109.47	109.40	109.47✓	109.40✓
$\varphi(N_1–C_2–C_3–H_1)$	60.18	60.19	60.19	60.19	60.19✓	60.19✓
MAX%[e]	1.7026	0.6613	0.2224	0.0999	0.0004	0.0004
MAE%[f]	1.1329	0.3332	0.0997	0.0889	0.0002	0.0002

[a] The symbol ✓ denotes parameters kept fixed at their guess value. [b] Vibrational corrections at the HPCS2 level of theory. [c] SE structure obtained starting from the BDPCS3 geometry. [d] SE structure obtained starting from the PCS2 geometry.

Table 9 Comparison between SE and computed bond lengths (in Å) and angles (in degrees) of 1,4-dihydronaphthalene. The labeling of geometrical parameters follows the atom numbering of Fig. 4

	r_{eq}			$r_{eq}^{SE\,a,b}$	
Parameter	HPCS2	DPCS3	BDPCS3	RD1(BDCPS3)[c]	RD2(BDCPS3)[d]
$r(C_1–C_2)$	1.5019	1.4973	1.4945	1.495(14)	1.49445(3)
$r(C_2–C_3)$	1.5155	1.5100	1.5071	1.507(20)	1.50708(3)
$r(C_3–C_4)$	1.4035	1.4007	1.3986	1.398(22)	1.39852(3)
$r(C_4–C_5)$	1.3924	1.3877	1.3857	1.386(15)	1.38567(3)
$r(C_{10}–C_1)$	1.3357	1.3331	1.3320	1.332(26)	1.33192(3)
$r(C_1–H_1)$	1.0889	1.0859	1.0834	1.0834✓	1.0834✓
$r(C_2–H_2)$	1.1013	1.0973	1.0948	1.0948✓	1.0948✓
$r(C_4–H_4)$	1.0886	1.0860	1.0835	1.0835✓	1.0835✓
$r(C_5–H_5)$	1.0868	1.0838	1.0813	1.0813✓	1.0813✓
$\vartheta(C_1–C_2–C_3)$	114.32	114.12	114.12	114.1(3)	114.120(1)
$\vartheta(C_2–C_3–C_4)$	118.95	118.80	118.80	118.8(3)	118.796(1)
$\vartheta(C_3–C_4–C_5)$	121.41	121.39	121.39	121.4(3)	121.387(1)
$\vartheta(C_4–C_5–C_6)$	119.46	119.47	119.47	119.5(3)	119.475(1)
$\vartheta(H_1–C_1–C_2)$	116.54	116.83	116.83	116.83✓	116.83✓
$\vartheta(H_2–C_2–C_3)$	109.32	109.22	109.22	109.22✓	109.22✓
$\vartheta(H_4–C_4–C_5)$	119.69	119.72	119.72	119.72✓	119.72✓
$\vartheta(H_5–C_5–C_6)$	120.39	120.37	120.37	120.37✓	120.37✓
$\varphi(H_2–C_2–C_3–C_4)$	−57.09	−57.24	−57.24	−57.24✓	−57.24✓
MAX%[e]	1.0428	0.3329	0.0211	0.0187	0.0186
MAE%[f]	0.9437	0.3142	0.0146	0.0103	0.0102

[a] The symbol ✓ denotes parameters kept fixed at their guess value. [b] Vibrational corrections at the HPCS2 level of theory. [c] SE structure obtained starting from the BDPCS3 geometry and optimizing the C–C–C class and C–C bonds separately. [d] SE structure obtained starting from the BDPCS3 geometry and optimizing the C–C and C–C–C classes. [e] Maximum absolute percentage error in the rotational constants (for all isotopic species) of the refined structure relative to their experimental counterparts. [f] Mean absolute percentage error in the rotational constants (for all isotopic species) of the refined structure relative to their experimental counterparts.

4.4 Accurate rotational constants of large molecules

As already mentioned, experimental rotational constants have been recently reported for the parent species of molecules

Table 10 MAEs of the rotational constants of 1,4-dihydronaphthalene computed by refining the C–C class through the isotopic species with respect to their corresponding SE counterparts. The labeling of geometrical parameters follows the atom numbering of Fig. 4

Isotopic species	A	B	C
Parent	0.0095	0.0159	0.0074
^{13}C2/9	0.0091	0.0162	0.0069
^{13}C1/10	0.0094	0.0159	0.0074
^{13}C4/7	0.0101	0.0157	0.0077
^{13}C5/6	0.0093	0.0157	0.0075
^{13}C3/8	0.0096	0.0155	0.0074

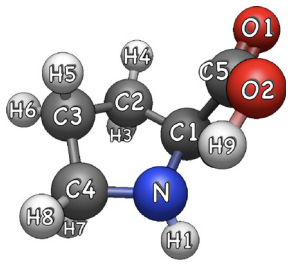


Fig. 6 Structure and labeling of the II_a conformer of proline.

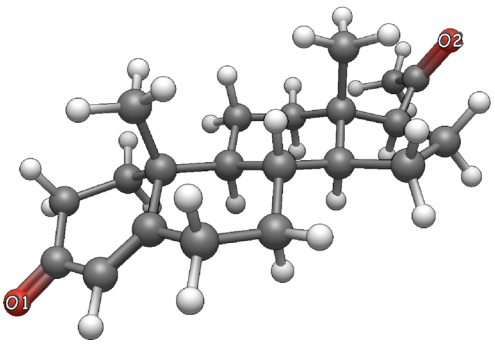


Fig. 7 Structure of progesterone.

Table 11 Comparison between SE and computed bond lengths (in Å) and angles (in degrees) of proline II_a. The labeling of geometrical parameters follows the atom numbering of Fig. 4

	r_{eq}				$r_{eq}^{SE\,ab}$		
Parameter	HPCS2	DPCS3	BDPCS3	PCS2	Ref. 97	RD(BDCPS3)[c]	RD (PCS2)[d]
$r(N–C_1)$	1.4845	1.4788	1.4759	1.4772	1.4785(18)	1.475(6)	1.478(2)
$r(C_1–C_2)$	1.5449	1.5355	1.5325	1.5324	1.5307(18)	1.532(6)	1.533(2)
$r(C_1–C_5)$	1.5425	1.5359	1.5329	1.5338	1.5321(18)	1.532(6)	1.534(2)
$r(C_2–C_3)$	1.5364	1.5291	1.5262	1.5271	1.5262(17)	1.525(6)	1.528(2)
$r(C_3–C_4)$	1.5321	1.5249	1.5220	1.5231	1.5215(15)	1.521(6)	1.524(2)
$r(C_5–O_1)$	1.2113	1.2043	1.2016	1.2019	1.2027(18)	1.2016✓	1.2019✓
$r(C_5–O_2)$	1.3405	1.3365	1.3338	1.3342	1.3343(19)	1.3338✓	1.3342✓
$r(N–H_1)$	1.0152	1.0100	1.0089	1.0085	1.0078(11)	1.0089✓	1.0085✓
$r(C_1–H_2)$	1.0951	1.0919	1.0894	1.0897	1.0891(23)	1.0894✓	1.0897✓
$r(C_2–H_3)$	1.0963	1.0928	1.0903	1.0907	1.0903(23)	1.0903✓	1.0907✓
$r(C_2–H_4)$	1.0923	1.0888	1.0863	1.0866	1.0865(23)	1.0863✓	1.0866✓
$r(C_3–H_5)$	1.0969	1.0935	1.0910	1.0915	1.0909(23)	1.0915✓	1.0915✓
$r(C_3–H_6)$	1.0939	1.0899	1.0874	1.0875	1.0872(23)	1.0874✓	1.0875✓
$r(C_4–H_7)$	1.1003	1.0961	1.0936	1.0937	1.0931(23)	1.0936✓	1.0937✓
$r(C_4–H_8)$	1.0946	1.0908	1.0883	1.0884	1.0880(23)	1.0883✓	1.0884✓
$r(O_2–H_2)$	0.9965	0.9844	0.9833	0.9821	0.9861(23)	0.9833✓	0.9821✓
$\vartheta(N–C_1–C_2)$	105.57	105.71	105.71	105.82	105.79(13)	105.5(5)	105.8(1)
$\vartheta(N–C_1–C_5)$	109.82	110.31	110.31	110.21	110.09(12)	110.31✓	110.21✓
$\vartheta(C_1–C_2–C_3)$	102.96	102.39	102.39	102.23	102.304(80)	102.2(5)	102.2(1)
$\vartheta(C_1–C_5–O_1)$	122.65	122.60	122.60	122.50	122.41(21)	122.60✓	122.50✓
$\vartheta(C_1–C_5–O_2)$	113.86	114.00	114.00	114.04	114.17(21)	114.00✓	114.04✓
$\vartheta(C_2–C_3–C_4)$	102.53	102.17	102.17	102.23	102.121(81)	102.0(5)	102.2(1)
$\vartheta(C_1–N–H_1)$	112.44	111.79	111.79	111.53	111.84(19)	111.79✓	111.53✓
$\vartheta(N–C_1–H_2)$	112.07	111.78	111.78	111.94	112.03(33)	111.78✓	111.94✓
$\vartheta(C_1–C_2–H_3)$	109.55	109.70	109.70	109.73	109.87(34)	109.70✓	109.73✓
$\vartheta(C_1–C_2–H_4)$	111.81	111.71	111.71	111.60	111.69(34)	111.71✓	111.60✓
$\vartheta(C_2–C_3–H_5)$	110.46	110.29	110.29	110.21	110.17(36)	110.29✓	110.21✓
$\vartheta(C_2–C_3–H_6)$	113.15	113.39	113.39	113.43	113.54(34)	113.39✓	113.43✓
$\vartheta(C_3–C_4–H_7)$	110.19	110.00	110.00	109.92	109.92(34)	110.00✓	109.92✓
$\vartheta(C_3–C_4–H_8)$	113.29	113.26	113.26	113.21	113.23(34)	113.26✓	113.21✓
$\vartheta(C_5–O_2–H_9)$	104.02	103.70	103.70	103.72	102.88(31)	103.70✓	103.72✓
$\varphi(N–C_1–C_2–C_3)$	23.94	26.11	26.11	26.82	26.47(18)	26.11✓	26.82✓
$\varphi(N–C_1–C_5–O_1)$	−179.98	−177.87	−177.87	−176.73	−177.36(35)	−177.87✓	−176.73✓
$\varphi(N–C_1–C_5–O_2)$	0.30	2.11	2.11	3.24	2.50(31)	2.11✓	3.24✓
$\varphi(C_1–C_2–C_3–C_4)$	−38.09	−39.91	−39.91	−40.25	−40.23(17)	−39.91✓	−40.25✓
$\varphi(N–C_1–C_2–H_3)$	−93.17	−90.62	−90.62	−89.81	−90.76(56)	−90.62✓	−89.81✓
$\varphi(N–C_1–C_2–H_4)$	147.00	149.05	149.05	149.64	148.91(56)	149.05✓	149.64✓
$\varphi(C_1–C_2–C_3–H_5)$	79.38	77.17	77.17	76.73	77.19(53)	77.17✓	76.73✓
$\varphi(C_1–C_2–C_3–H_6)$	−159.42	−161.26	−161.26	−161.59	−161.58(53)	−161.26✓	−161.59✓
$\varphi(C_1–C_5–O_2–H_9)$	0.21	−0.64	−0.64	−1.02	−0.41(56)	−0.64✓	−1.02✓
$\varphi(C_2–C_3–C_4–H_7)$	−80.97	−79.92	−79.92	−79.91	−79.57(55)	−79.92✓	−79.91✓
$\varphi(C_2–C_3–C_4–H_8)$	157.67	158.75	158.75	158.70	159.03(56)	158.75✓	158.70✓
$\varphi(C_5–C_1–N–H_1)$	−114.12	−117.67	−117.67	−119.35	−119.13(18)	−117.67✓	−119.35✓
$\varphi(H_1–N–C_1–C_2)$	125.65	122.38	122.38	120.94	121.09(17)	122.38✓	120.94✓
$\varphi(H_1–N–C_1–H_2)$	4.10	0.44	0.44	−1.25	0.01(56)	0.44✓	−1.25✓

[a] The symbol ✓ denotes parameters kept fixed at their guess value. [b] Vibrational corrections at the HPCS2 level of theory. [c] SE structure obtained starting from the BDPCS3 geometry. [d] SE structure obtained starting from the PCS2 structure.

containing up to 50 atoms, including several hormones.[18] In these cases, PCS2 geometry optimizations become too expensive, but the BDPCS3 variant can still be exploited. At the same time, the implementation of the FG and HG methods for normal modes expressed in terms of either Cartesian[19] or internal coordinates[20] allows for the systematic computation of vibrational corrections. Of course, the lack of isotopic substitutions does not allow an unbiased estimation of geometrical parameters, but it has been shown that relative errors of around 0.1% on rotational constants correspond to errors of the order of 1 mÅ on bond lengths and 0.2 degrees on valence angles.[75] At the same time, vibrational corrections typically range between 0.3 to 0.7% of the corresponding rotational constants and the error of HPCS2 values is within 10%.[18] Therefore, the best realistic accuracy of computed rotational constants for molecules of such size is 0.1%, which is usually referred to as spectroscopic accuracy.[75]

The performance of the proposed computational strategy is illustrated by the rotational spectrum of progesterone (see Fig. 7), which has been recently reported and interpreted in terms of DFT computations neglecting vibrational corrections.[15]

The results collected in Table 12 show that the BDPCS3 equilibrium geometry in conjunction with HPCS2 vibrational corrections provides results about one order of magnitude more accurate than the computations reported in ref. 15, reaching spectroscopic accuracy with reasonable computational efforts. Optimization of the average value of the bond lengths between heavy atoms (by means of a single offset) with respect to the experimental rotational constants of the parent species further reduces the error, but already the starting values can be considered fully satisfactory.

A similar accuracy has been obtained with analogous methods for several other hormones and biomolecule building blocks[18] with a cost comparable to that of standard DFT approaches. Therefore, the implementation of the whole computational protocol in a black box and user-friendly tool paves the way for its widespread use also by experiment-oriented researchers.

Table 12 Comparison between the equilibrium rotational constants of progesterone computed through different computational protocols and their SE counterparts obtained from experimental ground-state values and HPCS2 vibrational corrections. All the data except relative errors are in MHz

	HPCS2	DPCS3	BDPCS3	Fit[a]	SE[b]
B_a^{eq}	677.71(−3.14[c])	676.72	679.45	679.68	679.52
B_b^{eq}	130.61(−1.37[c])	132.12	132.55	132.60	132.62
B_c^{eq}	119.85(−1.34[c])	121.37	121.77	121.81	121.79
MAE%	0.884	0.378	0.028	0.018	—

[a] Rotational constants of the equilibrium geometry, obtained by refining the C–C and C–O bond lengths as a single class using the parent species. [b] SE rotational constants of the parent species, obtained by combining the experimental ground-state data from ref. 15 with vibrational corrections at the HPCS2 level of theory. [c] Vibrational corrections to the rotational constants of the parent species at the HPCS2 level of theory.

5 Conclusions

In this study, we have developed and validated an advanced computational workflow for the accurate determination of equilibrium geometries of isolated molecular systems containing between 10 and 50 atoms. The protocol is tailored for closed-shell, neutral species composed of light elements, where relativistic and static correlation effects can be safely neglected. By integrating the Pisa composite schemes (PCS) with efficient vibrational correction models into a fully automated pipeline, we provide a powerful and user-friendly tool for high-resolution rotational spectroscopists.

The workflow has been tested on diverse molecular systems, demonstrating robustness and accuracy for both semi-rigid and flexible species. The phenyl radical illustrates the applicability to open-shell systems, while proline exemplifies the treatment of flexibility within the VPT2 framework, provided that vibrational anharmonicity remains moderate.

A key strength of the approach is its ability to bridge the gap between theoretical predictions and experimental observations, enabling the extraction of equilibrium structures from vibrationally averaged spectroscopic data. This is particularly valuable in gas-phase studies, where intrinsic stereoelectronic properties can be analyzed without interference from environmental effects.

Overall, the proposed methodology enhances both the accuracy and accessibility of structural predictions, outperforming standard computational approaches at comparable cost. It offers a practical and general solution for the routine interpretation of high-resolution spectra, with applications in astrochemistry, atmospheric chemistry, and molecular pharmacology, and lays the groundwork for extending spectroscopic-quality structural analysis to increasingly complex molecular systems.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the ESI.† HPCS2 vibrational corrections for the considered isotopologues are given in Tables S1.1–S1.5 (ESI†), while the new PCS2 equilibrium structures of *t*-butyl isocyanide and proline are given in the attached zip file. Additional data and the new version of the MSR program are available on request from the corresponding author(s).

References

1 V. M. Rivilla, L. Colzi, I. Jiménez-Serra, J. Martín-Pintado, A. Megías, M. Melosso, L. Bizzocchi, A. López-Gallifa, A. Martínez-Henares and S. Massalkhi, *Astrophys. J., Lett.*, 2022, **929**, L11.
2 J. P. Gardner, J. C. Mather, M. Clampin, R. Doyon, M. A. Greenhouse, H. B. Hammel, J. B. Hutchings, P. Jakobsen,

S. J. Lilly, K. S. Long, J. I. Lunine, M. J. McCaughrean, M. Mountain, J. Nella, G. H. Rieke, M. J. Rieke, H.-W. Rix, E. P. Smith, G. Sonneborn, M. Stiavelli, H. S. Stockman, R. A. Windhorst and G. S. Wright, *Space Sci. Rev.*, 2006, **123**, 485–606.
3 T. Sugiki, N. Kobayashi and T. Fujiwara, *Comput. Struct. Biotechnol. J.*, 2017, **15**, 328–329.
4 C. Puzzarini and V. Barone, *Acc. Chem. Res.*, 2018, **51**, 548–556.
5 J. L. Lane, *Frontiers and Advances in Molecular Spectroscopy*, Elsevier, Amsterdam, 2018.
6 C. Puzzarini, J. Bloino, N. Tasinato and V. Barone, *Chem. Rev.*, 2019, **119**, 8131–8191.
7 S. Alessandrini, M. Melosso, L. Bizzocchi, V. Barone and C. Puzzarini, *J. Phys. Chem. A*, 2024, **128**, 5833–5855.
8 T. E. Field-Theodore, S. Alessandrini, M. Melosso and C. Puzzarini, *Chem. Phys. Lett.*, 2025, **868**, 141978.
9 J. Demaison, *Mol. Phys.*, 2007, **105**, 3109–3138.
10 N. P. Sahoo, P. R. Franke and J. F. Stanton, *J. Comput. Chem.*, 2024, **45**, 1419–1427.
11 S. V. M. Caliebe, P. Pinacho and M. Schnell, *J. Phys. Chem. Lett.*, 2022, **127**, 11913–11917.
12 I. León, E. R. Alonso, S. Mata and J. L. Alonso, *J. Phys. Chem. Lett.*, 2021, **12**, 6983–6987.
13 I. Peña, C. Cabezas and J. L. Alonso, *Angew. Chem., Int. Ed.*, 2015, **54**, 2991–2994.
14 S. Zinn and M. Schnell, *ChemPhysChem*, 2018, **19**, 2915–2920.
15 A. Insausti, E. R. Alonso, S. Municio, I. Leon, L. Kolesnikova and S. Mata, *J. Phys. Chem. Lett.*, 2025, **16**, 2425–2432.
16 Y. Li, F. Shen, Y. Feng, W. Lv, H. Huang, J. Huang and G. Feng, *J. Chem. Phys.*, 2025, **162**, 134301.
17 S. Di Grande and V. Barone, *J. Phys. Chem. A*, 2024, **128**, 4886–4900.
18 V. Barone, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2025, **15**, e70000.
19 M. Mendolicchio and V. Barone, *J. Chem. Theory Comput.*, 2024, **20**, 2842–2857.
20 M. Mendolicchio and V. Barone, *J. Chem. Theory Comput.*, 2024, **20**, 8378–8395.
21 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji *et al.*, *Gaussian 16 Revision C.01*, Gaussian Inc., Wallingford CT, 2016.
22 M. Mendolicchio, E. Penocchio, D. Licari, N. Tasinato and V. Barone, *J. Chem. Theory Comput.*, 2017, **13**, 3060–3075.
23 S. Carter, A. R. Sharma, J. M. Bowman, P. Rosmus and R. Tarroni, *J. Chem. Phys.*, 2009, **131**, 224106.
24 G. Rauhut and T. Hrenar, *Chem. Phys.*, 2008, **346**, 160–166.
25 O. Christiansen, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6672–6687.
26 D. Papp, T. Szidarovszky and A. G. Császár, *J. Chem. Phys.*, 2017, **147**, 094106.
27 T. Carrington, *J. Chem. Phys.*, 2017, **146**, 120902.
28 A. Erba, J. Maul, M. Ferrabone, P. Carbonnière, M. Rérat and R. Dovesi, *J. Chem. Theory Comput.*, 2019, **15**, 3755–3765.
29 A. Erba, J. Maul, M. Ferrabone, R. Dovesi, M. Rérat and P. Carbonnière, *J. Chem. Theory Comput.*, 2019, **15**, 3766–3777.
30 E. Mátyus, A. M. Santa Dara and G. Avila, *Chem. Commun.*, 2023, **59**, 366–381.
31 T. Carrington Jr, *Spectrochim. Acta, Part A*, 2021, **248**, 119158.
32 S. Manzhos, X. Wang and T. Carrington Jr, *Chem. Phys.*, 2018, **509**, 139–144.
33 D. Lauvergnat, E. Balotcha, G. Dive and M. Desouter-Lecomte, *Chem. Phys.*, 2006, **326**, 500–508.
34 S. N. Yurchenko, W. Thiel and P. Jensen, *J. Mol. Spectrosc.*, 2007, **245**, 126–140.
35 B. T. Sutcliffe and J. Tennyson, *Int. J. Quantum Chem.*, 1991, **39**, 183–196.
36 A. S. Petit and A. B. McCoy, *J. Phys. Chem. A*, 2013, **117**, 7009–7018.
37 I. W. Bulik, M. J. Frisch and P. H. Vaccaro, *J. Chem. Theory Comput.*, 2018, **14**, 1554–1563.
38 K. Yagi, M. Keçeli and S. Hirata, *J. Chem. Phys.*, 2012, **137**, 204118.
39 D. O. Harris, G. G. Engerholm and W. D. Gwinn, *J. Chem. Phys.*, 1965, **43**, 1515–1517.
40 A. Dickinson and P. Certain, *J. Chem. Phys.*, 1968, **49**, 4209–4211.
41 H. H. Nielsen, *Rev. Mod. Phys.*, 1951, **23**, 90–136.
42 I. M. Mills, *Molecular Spectroscopy: Modern Research*, Academic Press, New York, 1972, ch. 3.2, pp. 115–140.
43 D. A. Clabo Jr., W. D. Allen, R. B. Remington, Y. Yamaguchi and H. F. Schaefer III, *Chem. Phys.*, 1988, **123**, 187–239.
44 W. D. Allen, Y. Yamaguchi, A. G. Császár, D. A. Clabo Jr., R. B. Remington and H. F. Schaefer III, *Chem. Phys.*, 1990, **145**, 427–466.
45 V. Barone, *J. Chem. Phys.*, 2005, **122**, 014108.
46 S. V. Krasnoshchekov, E. V. Isayeva and N. F. Stepanov, *J. Phys. Chem. A*, 2012, **116**, 3691–3709.
47 A. M. Rosnik and W. F. Polik, *Mol. Phys.*, 2014, **112**, 261–300.
48 P. R. Franke, J. F. Stanton and G. E. Douberly, *J. Phys. Chem. A*, 2021, **125**, 1301–1324.
49 C. Puzzarini, J. F. Stanton and J. Gauss, *Int. Rev. Phys. Chem.*, 2010, **29**, 273–367.
50 M. Mendolicchio, J. Bloino and V. Barone, *J. Chem. Theory Comput.*, 2021, **19**, 1759–1787.
51 M. Mendolicchio, J. Bloino and V. Barone, *J. Chem. Theory Comput.*, 2022, **18**, 7603–7619.
52 W. Schneider and W. Thiel, *Chem. Phys. Lett.*, 1989, **157**, 367–373.
53 P. Pulay, W. Meyer and J. E. Boggs, *J. Chem. Phys.*, 1978, **68**, 5077–5085.
54 E. Penocchio, M. Mendolicchio, N. Tasinato and V. Barone, *Can. J. Chem.*, 2016, **94**, 1065–1076.
55 L. Uribe, S. Di Grande, M. Mendolicchio, N. Tasinato and V. Barone, *J. Phys. Chem. A*, 2024, **128**, 10474–10488.
56 L. Uribe, M. Mendolicchio, S. Municio, S. Mato, E. R. Alonso, J. L. Alonso, L. Iker and V. Barone, *J. Phys. Chem. Lett.*, 2025, **16**, 6523–6532.
57 M. J. Frisch, G. W. Trucks, G. Scalmani, J. R. Cheeseman, X. Li, J. Bloino, B. G. Janesko, A. V. Marenich, J. Zheng, F. Lipparini *et al.*, *Gaussian Development Version, Revision J-28*, Gaussian Inc., Wallingford CT, 2023.

58 V. Barone, G. Ceselin and N. Tasinato, *J. Phys. Chem. A*, 2023, **127**, 5183–5192.
59 N. Vogt, J. Demaison, J. Vogt and H. D. Rudolph, *J. Comput. Chem.*, 2014, **35**, 2333–2342.
60 G. Knizia, T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2009, **130**, 054104.
61 K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chem. Phys. Lett.*, 1989, **157**, 479–483.
62 E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, S. Bannwarth and C. Grimme, *J. Chem. Phys.*, 2019, **150**, 154122.
63 G. Santra, N. Sylvetsky and J. M. Martin, *J. Phys. Chem. A*, 2019, **123**, 5129–5143.
64 K. A. Peterson, T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2008, **128**, 084102.
65 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.
66 S. Grimme, *Wiley Interdiscip. Rev.:Comput. Mol. Sci.*, 2011, **1**, 211–228.
67 R. Sure and S. Grimme, *J. Comput. Chem.*, 2013, **34**, 1672–1685.
68 S. Grimme, A. Hansen, S. Ehlert and J.-M. Mewes, *J. Chem. Phys.*, 2021, **154**, 064103.
69 P. R. Spackman, D. Jayatilaka and A. Karton, *J. Chem. Phys.*, 2016, **145**, 104101.
70 S. Di Grande, F. Lazzari and V. Barone, *J. Chem. Theory Comput.*, 2024, **20**, 9243–9258.
71 K. E. Yousaf and K. A. Peterson, *J. Chem. Phys.*, 2008, **129**, 184108.
72 K. E. Yousaf and K. A. Peterson, *Chem. Phys. Lett.*, 2009, **476**, 303–307.
73 C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–622.
74 K. A. Peterson and T. H. Dunning Jr, *J. Chem. Phys.*, 2002, **117**, 10548–10560.
75 C. Puzzarini and J. F. Stanton, *Phys. Chem. Chem. Phys.*, 2023, **25**, 1421–1429.
76 B. Cordero, V. Gomez, A. E. Platero-Prats, M. Revés, J. Echevarria, E. Cremades, F. Barragan and S. Alvarez, *Dalton Trans.*, 2008, 2832–2838.
77 V. Barone and F. Lazzari, *J. Phys. Chem. A*, 2023, **127**, 10517–10527.
78 T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. V. R. Schleyer, *J. Comput. Chem.*, 1983, **4**, 294–301.
79 H.-J. Werner, P. J. Knowles, F. R. Manby, J. A. Black, K. Doll, A. Heßelmann, D. Kats, A. Köhn, T. Korona and D. A. Kreplin, *et al.*, *J. Chem. Phys.*, 2020, **152**, 144107.
80 F. Lazzari, S. Di Grande, L. Crisci, M. Mendolicchio and V. Barone, *J. Chem. Phys.*, 2025, **162**, 114310.
81 V. L. Orr, Y. Ichikawa, A. R. Patel, S. M. Kougias, K. Kobayashi, J. F. Stanton, B. J. Esselman, R. C. Woods and R. J. McMahon, *J. Chem. Phys.*, 2021, **154**, 244310.
82 L. Bonah, B. Helmstaedter, J.-C. Guillemin, S. Schlemmer and S. Thorwirth, *J. Mol. Spectrosc.*, 2025, **468**, 111967.
83 N. C. Craig, Y. Chen, H. A. Fuson, H. Tian, H. van Besien, A. R. Conrad, M. J. Tubergen, H. D. Rudolph and J. Demaison, *J. Phys. Chem. A*, 2013, **117**, 9391–9400.
84 O. J. Martinez, K. N. Crabtree, C. A. Gottlieb, J. F. Stanton and M. C. McCarthy, *Angew. Chem., Int. Ed.*, 2014, **54**, 1808–1811.
85 V. Barone, M. Biczysko, J. Bloino, F. Egidi and C. Puzzarini, *J. Chem. Phys.*, 2013, **138**, 234303.
86 V. Barone, L. Uribe, S. Srivastav and A. Pathak, *ACS Earth Space Chem.*, 2024, **8**, 2334–2344.
87 L. Crisci, F. Lazzari and V. Barone, *J. Chem. Theory Comput.*, 2025, **14**, 7188–7197.
88 C. Cabezas, J. Janeiro, D. Perez, W. Li, M. Agundez, A. L. Steber, E. Guitian, J. Demaison, C. Perez and J. Cernicharo, *et al.*, *J. Phys. Chem. Lett.*, 2024, **15**, 7411–7418.
89 C. Cabezas, J. Janeiro, A. L. Steber, D. Perez, C. Bermudez, E. Guitian, A. Lesarri and J. Cernicharo, *Phys. Chem. Chem. Phys.*, 2024, **26**, 23703–23709.
90 D. McNaughton, M. K. Jahn, M. J. Travers, D. Wachsmuth, P. D. Godfrey and J.-U. Grabow, *Mon. Not. R. Astron. Soc.*, 2018, **476**, 5268–5273.
91 G. Wenzel, I. R. Cooke, T. H. Speak, P. B. Changala, E. A. Bergin, S. Zhang, A. M. Burkhardt, A. N. Byrnel, S. B. Charnley, M. A. Cordiner, M. Duffy, Z. T. P. Fried, H. Gupta, M. S. Holdren, A. Lipnicky, R. A. Loomis, H. Toru Shai, C. N. Schlingledecker, M. A. Siebert, D. A. Stewart, R. H. Willis, C. Xue, A. J. Remijan, A. E. Wendlandt, M. C. McCarthy and B. McGuire, *Science*, 2024, **386**, 810–813.
92 G. Wenzel, T. H. Speak, P. B. Changala, R. H. J. Willis, A. M. Burkhardt, S. Zhang, E. A. Bergin, A. N. Byrnel, S. B. Charnley, Z. T. P. Fried, H. Gupta, E. Herbst, M. S. Holdren, A. Lipnicky, R. A. Loomis, C. N. Schlingledecker, C. Xue, A. J. Remijan, A. E. Wendlandt, M. C. McCarthy, I. R. Cooke and B. McGuire, *Nat. Astron.*, 2025, **9**, 262–270.
93 M. A. Zdanovskaia, B. J. Esselman, R. C. Woods and R. J. McMahon, *J. Chem. Phys.*, 2019, **151**, 024301.
94 A. Belloche, R. T. Garrod, S. P. Muller and K. M. Menten, *Science*, 2014, **345**, 1584–1587.
95 L. Pagani, C. Favre, P. F. Goldsmith, A. Bergin, R. Snell and G. Melnick, *Astron. Astrophys.*, 2017, **604**, A32.
96 N. W. Howard, A. C. Legon, C. A. Rego and A. L. Wallwork, *J. Mol. Struct.*, 1989, **197**, 181–191.
97 N. Vogt, J. Demaison, S. V. Krasnoshchekov, N. F. Stepanov and H. D. Rudolph, *Mol. Phys.*, 2017, **115**, 942–951.