RESEARCH

Harnessing the power of curvilinear internal coordinates: from molecular structure prediction to vibrational spectroscopy

Marco Mendolicchio[1]

Abstract

Different standard VPT2 codes employ Cartesian coordinates for the computation of rotational and vibrational spectroscopic parameters. However, curvilinear internal coordinates offer a number of advantages provided that a general non-redundant set of coordinates can be built and employed in an unsupervised workflow. In the present paper I summarize the main results and perspectives of a general engine employing curvilinear internal coordinates and perturbation theory for the computation of rotational and vibrational spectroscopic parameters of large molecules beyond the conventional rigid rotor/harmonic oscillator model. Some examples concerning biomolecule building blocks are discussed in some detail in order to better analyze the performance of the proposed strategy.

Keywords Internal coordinates · Anharmonicity · Semi-experimental · Computational spectroscopy

1 Introduction

The quest for a reliable yet practical modeling of large molecular systems has always played a central role in the field of theoretical and computational chemistry, and this balance between accuracy and feasibility is likely to persist into the future [1]. Combination of different complementary techniques is often the key to determining in an univocal way the molecular structure and bond topology for a wide range of systems, ranging from small molecules to large biomolecules. In particular, high-resolution spectroscopic studies both in the gas-phase (microwave, MW) and in inert matrices (infrared, IR) allow an unbiased disentanglement of intrinsic stereo-electronic effects without any strong perturbing effect from the environment. However, the tuning of experimental spectra by different low-lying structures and the interplay of different contributions can make the interpretation of experimental data troublesome or even impossible without the aid of trustworthy *in silico* simulations. In fact, ongoing improvements of hardware, software, and, above all, underlying physical–mathematical models are allowing the analysis of experimental data and their interpretation for molecular

systems of increasing complexity. Despite the undisputed effectiveness of static structure–property correlations and the fundamental rigid rotor/harmonic oscillator (RRHO) model, results that are directly comparable to experiment can only be achieved through more advanced models: (i) at the electronic level, employing highly-correlated methods; (ii) at the nuclear level, by incorporating anharmonic effects in the description of the nuclear motions. In fact, a strong limitation of the harmonic approximation is that the real vibrations are intrinsically anharmonic, and as a consequence vibrational energies are systematically overestimated. A common empirical scheme for improving the agreement with experimental results rests on the application of scaling factors depending on the employed quantum chemical (QC) method and, possibly, on the range of investigated frequencies. However, those scaling factors, generally obtained from statistical analyses of fundamental transitions, do not improve intensities and are often not suitable for higher-quanta transitions. The introduction of anharmonic effects requires a representation of potential energy surfaces (PESs) around stationary points beyond the quadratic approximation, introducing a series of higher-order terms involving one or more vibrational coordinates. In this context, a crucial point is represented by the inclusion of couplings involving different vibrations, which have to be properly considered in order to obtain accurate estimates of line shapes, frequency shifts, etc. Computational strategies aimed at the study of vibrational properties of molecular

✉ Marco Mendolicchio
marco.mendolicchio@sns.it

1   Scuola Normale Superiore, Piazza dei Cavalieri 7,
    56126 Pisa, Italy

systems are usually based on perturbative [2–9] variational [10–12] approaches or their combinations [13, 14]. Within variational methods, the energy is minimized starting from an expansion of the wave functions over a basis of known states. This class of methods includes the vibrational self-consistent-field (VSCF) [12, 15–20] and vibrational configuration interaction (VCI) [21–23]. While the use of accurate variational approaches in conjunction with high-order force fields allows to obtain estimates vibrational energies close to their experimental counterparts, their use is mainly limited to systems containing few atoms due to their unfavorable scaling with the size of the molecule [24]. With the aim of reducing the computational cost, several strategies have been devised, such as the iterative subspace expansion algorithms [25–28] and protocols for the reduction of the VCI basis [29, 30]. Over the last decades, several variational approaches have been proposed for the resolution of the vibrational problems, not only relying on the Watson's Hamiltonian and exploiting different sets of nuclear coordinates. In particular, different approaches and sophisticated variational, semi-variational nuclear motion and dynamics theoretical frameworks (and codes) built upon the use of internal coordinates have been discussed by Carrington [31, 32], Csaszar [33, 34], Lauvergnat [35–37], Tennyson [38, 39], and Yurchenko [40], among others. Variational approaches exploiting Monte Carlo diffusion methods in internal coordinates have been also investigated [41, 42], as well as the use of optimized and localized coordinates [43, 44], for example, in the vibrational coupled-cluster (VCC) framework [45]. Another option is the parametrization of the vibrational wavefunction in tensor formats such as canonical decomposition [46] and matrix product states. Following that, the wavefunction representations can be optimized through the recent extension of the density matrix renormalization group (DMRG) [47, 48] algorithm to the study of the vibrational problem, the resulting method being referred to as vibrational DMRG (vDMRG) [49, 50].

When comparing the various methods for including anharmonic effects, [2, 3, 16–19, 21, 22, 51–53] those based on perturbation theory applied to the Watson Hamiltonian offer a remarkable balance between accuracy and computational cost provided that resonance effects are taken into proper account. In particular, the vibrational second-order perturbation theory (VPT2), [2, 54] based on a fourth-order polynomial approximation of the potential energy in terms of normal coordinates, allows to study medium-to-large size molecular systems. Furthermore, the formulation of VPT2 based on the Van Vleck contact transformation method [55] enables the extension to a generalized model (GVPT2) [6]. Within this framework, the diagonalization of a limited set of reduced-dimensionality Hamiltonians involving strongly interacting states is carried out, while the other states are corrected by second-order perturbative contributions. At the VPT2 level, also the vibrational dependence of rotational constants on molecular vibrations can be expressed through a perturbative expansion leading to the determination of specific vibro-rotational interaction constants and paving the route for the calculation of accurate molecular structures through the semi-experimental (SE) approach [56].

While VPT2 can be successfully applied in many cases, it is also characterized by some intrinsic drawbacks that prevent its use in some scenarios. As mentioned above, the first problem is related to the presence of resonances (both Fermi and Darling-Dennison), even though several computational strategies as well as robust schemes have been developed over the years [6, 57–62] even for the treatment of vibrational intensities. Second, the theoretical framework largely depends on the symmetry of the molecule since the presence of degenerate vibrations implies a distinct derivation of the equations [63, 64]. As a consequence, different derivations and implementations of VPT2 are required for the treatment of systems belonging to different classes of symmetry. Recently, this issue was solved by the formulation of a unified model relying on the proper application of *a posteriori* transformations to the wavefunctions [65]. Last but not least, large amplitude motions (LAMs) are poorly described at the VPT2 level and more generally with a quartic force field. Unfortunately, variational approaches such as the VCI or more refined perturbative methods (such as VPT4 [66]) become rapidly prohibitive as the size of the molecular systems increases, due to the necessity of more demanding anharmonic force field calculations and the increasing complexity concerning the vibrational calculation. On the other hand, low-dimensionality methods tailored for describing one or a limited number of LAMs [67–72] imply their separation from the rest of vibrations, usually referred to as small amplitude motions (SAMs). As shown in a previous study, [73] the definition of a suitable set of internal coordinates able to decouple these two classes of vibration allows to treat each normal mode through the most appropriate method. Starting from the available literature, [74, 75] the VPT2 framework in terms of curvilinear coordinates was developed and successfully applied to both semi-rigid and flexible systems, showing in the latter case that not only LAMs can be effectively decoupled from SAMs, but also that each LAM can be treated independently from the others. As a matter of fact, SAMs can be treated through the internal-based VPT2, while variational approaches can be employed for the calculation of the energy levels of the floppy degrees of freedom. When a single LAM is present, it can be effectively treated through the discrete variable representation (DVR) [76–83]. On the other hand, a higher number of LAMs can be in principle handled through the reaction-surface Hamiltonian (RSH) [84], reaction-volume Hamiltonian (RVH) [85] or VCI-based methodologies.

In this paper, the versatility of internal coordinates and their application in both fields of rotational and vibrational

spectroscopy will be described. The paper is organized as follows. First, an overview of the Cartesian-based version of VPT2 will be provided, including the calculation of the vibrational contributions to the rotational constants required in the SE approach. Particular attention will be devoted to the use of different sets of internal coordinates in the non-linear least-squares fit needed for the calculation of the SE equilibrium molecular geometry. Then, the focus will shift to the extension of VPT2 to the use of internal coordinates, from the calculation of the anharmonic energy levels to the definition of a robust scheme for the automatic identification of Fermi resonances. In this context, the importance of internal coordinates in reducing inter-mode couplings will be underlined through the application to systems of biological and technological interest.

2 Theory

2.1 Overview of the Cartesian-based VPT2

In this section, the unified Cartesian-based, VPT2 framework for asymmetric, symmetric, linear, and spherical tops is described. In this theoretical model, all normal modes are treated independently from one another regardless the presence of vibrational degeneracies. In the latter instance, a series of linear transformations are employed *a posteriori* to determine the correct energies, wave functions and properties. The discussion is structured to give a broad overview of this methodology up to the derivation of the vibrational energy levels also accounting for the presence of Fermi resonances.

Let us consider a molecule composed of N vibrational modes (N being equal to $3N_a - 6$ for nonlinear molecules and to $3N_a - 5$ for linear ones, where N_a is the number of atoms). The reference Hamiltonian is obtained by expanding the more general vibro-rotational Hamiltonian proposed by Watson [86] in terms of normal coordinates and then only considering the purely vibrational terms:

$$
\mathcal{H} = \frac{1}{2} \sum_{i=1}^{N} \omega_i (p_i^2 + q_i^2) + \frac{1}{6} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} f_{ijk} q_i q_j q_k
$$
$$
+ \frac{1}{24} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} f_{ijkl} q_i q_j q_k q_l
$$
$$
+ \sum_{\tau=x,y,z} B_\tau^{eq} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \zeta_{ij,\tau} \zeta_{kl,\tau} \sqrt{\frac{\omega_j \omega_l}{\omega_i \omega_k}} q_i p_j q_k p_l + \mathcal{U}
$$
(1)

where q_i is the i-th dimensionless normal coordinate, p_i is its conjugate momentum, ω_i is the corresponding harmonic wavenumber (in cm^{-1}), B_τ^{eq} is the equilibrium rotational

constant along the τ axis and ζ is the anti-symmetric matrix of Coriolis couplings. The terms f$_{ijk}$ and f$_{ijkl}$ are, respectively, referred to as cubic and quartic force constants, defined by the following notation,

$$
f_{ijk\ldots} = \left(\frac{\partial^n \mathcal{V}}{\partial q_i \partial q_j \partial q_k \ldots} \right)_{eq}
$$
(2)

where $\mathcal{V}$ represents the potential energy, while $\mathcal{U}$ is a mass-dependent term which does not contribute to the calculation of the transition energies, and for this reason it will be not considered from now on.

Within the VPT2 framework, the vibrational Hamiltonian is expanded through a perturbative series up to the second order, the resulting Schrödinger equation being typically solved by means of two different approaches, namely the Rayleigh–Schrödinger (RSPT) and Van Vleck (CVPT) [55] perturbation theory. As a result, the VPT2 vibrational energy of a state $|v_R\rangle$ is

$$
\varepsilon_R = \varepsilon_0 + \sum_{i=1}^{N} \omega_i v_{R,i} + \sum_{i=1}^{N} \sum_{j=i}^{N} \chi_{ij} \left[v_{R,i} v_{R,j} + \frac{1}{2} \left(v_{R,i} + v_{R,j} \right) \right]
$$
(3)

where $v_{R,i}$ is the number of quanta associated with the mode i in the state R, and ε_0 is the anharmonic resonance-free zero-point vibrational energy (ZPVE) [9, 87, 88].

The χ matrix contains the anharmonic corrections and is defined by the following expressions,

$$
16\chi_{ii} = f_{iiii} - \frac{5 f_{iii}^2}{3\omega_i} - \sum_{\substack{j=1 \\ (j \neq i)}}^{N} \frac{f_{iij}^2}{2}
$$
$$
\left[\frac{4}{\omega_j} + \frac{1}{2\omega_i + \omega_j} - \frac{1}{2\omega_i - \omega_j} \right]
$$
(4)

$$
4\chi_{ij} = f_{iijj} - \frac{f_{iij}^2}{2} \left[\frac{1}{2\omega_i + \omega_j} + \frac{1}{2\omega_i - \omega_j} \right]
$$
$$
- \frac{f_{ijj}^2}{2} \left[\frac{1}{2\omega_j + \omega_i} + \frac{1}{2\omega_j - \omega_i} \right] - \frac{f_{iii} f_{ijj}}{\omega_i} - \frac{f_{jjj} f_{iij}}{\omega_j}
$$
$$
+ \sum_{\substack{k=1 \\ (k \neq i,j)}}^{N} \left\{ \frac{f_{ijk}^2}{2} \left[\frac{1}{\omega_i + \omega_j + \omega_k} + \frac{1}{\omega_i - \omega_j + \omega_k} \right. \right.
$$
$$
\left. \left. - \frac{1}{\omega_i + \omega_j - \omega_k} - \frac{1}{\omega_i - \omega_j - \omega_k} \right] - \frac{f_{iik} f_{jjk}}{\omega_k} \right\}
$$
$$
+ 4 \left(\frac{\omega_i}{\omega_j} + \frac{\omega_j}{\omega_i} \right) \sum_{\tau=x,y,z} B_\tau^{eq} \{\zeta_{ij,\tau}\}^2
$$
(5)

Inspection of Eqs. 4 and 5 evidences that the presence of denominators approaching zero would lead to unphysical results. These conditions, collectively known as Fermi resonances (FRs), occur when $\omega_i \approx 2\omega_j$ or $\omega_i \approx \omega_j + \omega_k$. The first case is commonly referred to as type I, while the second as type II [89]. Several strategies can be adopted to handle this problem. In the deperturbed VPT2 (DVPT2), a sequential screening of all potentially resonant terms is performed, and an analysis based on one or more criteria allows to state which terms should be removed from the anharmonic calculation. Generally, this procedure is organized into two step. The first one is the evaluation of the energetic proximity of the interacting states at the harmonic level,

$$|\omega_i - \omega_j - \omega_k| \leq \Delta\omega^{1-2} \tag{6}$$

where i and j can be equal and $\Delta\omega^{1-2}$ is a threshold defined a priori. Second, the overall weight of the term is estimated. Several strategies have been proposed for the latter [7, 58, 90]. In this work, the test proposed by Martin and co-workers [90] has been employed, leading to the following condition, [91]

$$\frac{f_{ijk}^{\,4}}{64(1 + \delta_{jk})^2 |\omega_i - \omega_j - \omega_k|^3} \geq K^{1-2} \tag{7}$$

where K^{1-2} is a second threshold required in the procedure. In the DVPT2 scheme, each term fulfilling Eqs. 6 and 7 is labeled as resonant and removed from the anharmonic calculation. If on the one side this method avoids singularities in the calculation of the energy levels, on the other side it can lead to a truncate treatment, since the resonant terms are systematically neglected. In order to prevent this situation, the interaction terms related to FRs can be introduced back in a successive, variational step. For the purpose, a variational matrix $\tilde{\mathbf{H}}$ is built starting from the DVPT2 energies and Van Vleck Hamiltonian interaction terms. A new set of energies is obtained by diagonalizing $\tilde{\mathbf{H}}$, the full procedure being referred to as GVPT2^F (hereafter, simply GVPT2).

2.2 Vibrational corrections

The rotational constants of a target molecule are among the most important parameters obtained from the study of rotational and vibro-rotational spectra, and they are inversely proportional to the principal moments of inertia. In general, the availability of this kind of data for different isotopic species allows for the determination of structural parameters like bond lengths and angles [92]. Nevertheless, since molecules are not rigid rotors and are subject to vibrational motion, the notion of a

reference molecular geometry is far from being a simple concept. An important step forward has been done by Pulay and co-workers [56] through the introduction of the SE approach for the determination of the equilibrium structure of a molecule. While more difficult to measure at the experimental level, this kind of structure properly accounts for vibrational effects and it is independent from the isotopic species. Furthermore, it is directly comparable to quantum-mechanical data. Within the SE method, the structural parameters, are obtained through a fit of the principal moments of inertia (or the corresponding rotational constants). Hence, the SE rotational constant B_τ^{SE} is obtained as follows [93],

$$B_\tau^{SE} = B_\tau^{0,exp} - (\Delta B_\tau^0)^{QM} \tag{8}$$

where $\tau = a, b, c$ is one of the principal inertial axes, while $B^{0,exp}$ and $(\Delta B_\tau^0)^{QM}$ are, respectively, the experimental rotational constant and its correction along τ. In general terms, $(\Delta B_\tau^0)^{QM}$ has a double nature. It is composed of a vibrational contribution (ΔB_τ^{vib}) stemming from the VPT2 framework, and an electronic contribution (ΔB_τ^{ele}) [93–95] related to the effect of electron on the moments of inertia. In this context only the vibrational contribution will be analyzed, since the electronic term is often negligible and required only for highly accurate calculations.

The vibrational contribution can be derived from the contact-transformed Hamiltonian, leading to an explicit dependence of rotational constants on vibrational degrees of freedom [94],

$$B_\tau^{v_R} = B_\tau^{eq} - \sum_{i=1}^{N} \alpha_{\tau,i}\left(v_{R,i} + \frac{1}{2}\right) \tag{9}$$

where $B_\tau^{v_R}$ and B_τ^{eq} are the rotational constants associated with the vibrational state $|v_R\rangle$ and the corresponding equilibrium value, respectively, while $\alpha_{\tau,i}$ are the vibro-rotational interaction constants:

$$\alpha_{i,\tau} = -2(B_\tau^{eq})^2 \left[\sum_{\eta=x,y,z} \frac{3(a_{i,\tau\eta})^2}{4\omega_i I_\eta^{eq}} \right.$$
$$\left. + \sum_{j=1}^{N} \frac{(\zeta_{ij,\tau})^2(3\omega_i^2 + \omega_j^2)}{\omega_i(\omega_i^2 - \omega_j^2)} + \pi\sqrt{\frac{c}{h}} \sum_{j=1}^{N} \frac{f_{iij}a_{j,\tau\tau}}{\omega_j^{3/2}} \right] \tag{10}$$

where $a_{i,\tau\eta}$ are the derivatives of the effective moments of inertia tensor evaluated at the equilibrium geometry. When the vibrational ground state is considered, Eq. 9 can be rearranged to define the vibrational contribution $\Delta B_\tau^{vib} = B_\tau^0 - B_\tau^{eq}$ as,

$$\Delta B_\tau^{vib} = B_\tau^{eq} - B_\tau^0$$

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

$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(\zeta_{ij,\tau})^2(3\omega_i^2 + \omega_j^2)}{\omega_i(\omega_i^2 - \omega_j^2)}$$

$$\left. + \pi\sqrt{\frac{c}{h}} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{f_{iij} a_{j,\tau\tau}}{\omega_j^{3/2}} \right] \tag{11}$$

As it can be deduced from Eq. 10, the coefficients $\alpha_{\tau,i}$ are affected by resonances, occurring when $\omega_i \approx \omega_j$. However, in the calculation of the vibrational correction this issue is solved by recasting the Coriolis contribution as follows,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(\zeta_{ij,\tau})^2(3\omega_i^2 + \omega_j^2)}{\omega_i(\omega_i^2 - \omega_j^2)} = -\sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{(\zeta_{ij,\tau})^2(\omega_i - \omega_j)^2}{\omega_i\omega_j(\omega_i + \omega_j)} \tag{12}$$

Once the set of SE rotational constants is assembled, a non-linear least-squares fit is carried out in order to determine the molecular geometry, which is described by a proper set of nuclear coordinates. For the purpose, the theoretical framework has been implemented in the MSR (Molecular Structures Refinement) software, [96–98] developed in our research group and designed for the determination of equilibrium structures through the SE approach. Over the last years, the MSR code has been employed for the characterization of numerous molecular systems including sulfur-containing [96, 99] and astrochemical systems, [100] biological building blocks [97] and non-covalent complexes [101]. The program has been developed with the target of being as general as possible, and equipped with a series of features at different levels. First, a wide range of optimization algorithms (such as the Gauss–Newton and Levenberg-Marquardt) has been included. Second, the so-called predicate observations [102–104] can be added to the set of reference data. Within this approach, the set of SE data is augmented with quantum-mechanical estimates of one or more parameters. This feature can be particularly advantageous when the number of isotopic species is not sufficiently large for the full structural characterization. Third, a detailed error analysis has been implemented, with the possibility to calculate, for example, the standard deviations on the singular parameters, the outliers and the condition number.

2.3 Curvilinear coordinates formalism

As previously anticipated, the choice of a proper set of curvilinear internal coordinates plays a central role in vibro-rotational analyses. For the present discussion, let us consider a set of non-redundant internal coordinates s. Since the latter are nonlinear functions of Cartesian coordinates, a strategy commonly adopted is represented by a Taylor-series expansion around the reference structure,

$$s_i = s_i^{eq} + \sum_{i=1}^{N} B_{ij}(x_j - x_j^{eq}) + \frac{1}{2} \sum_{j=1}^{3N_a} \sum_{k=1}^{3N_a} B_{ijk}'(x_j - x_j^{eq})(x_k - x_k^{eq}) + \dots \tag{13}$$

where $x = \{x_1, x_2, \dots, x_{3N_a}\}$ collects the nuclear Cartesian coordinates, N_a is the number of atoms, and the Wilson **B** matrix and its Cartesian derivative $\mathbf{B}'$ have been introduced,

$$B_{ij} = \left(\frac{\partial s_i}{\partial x_j}\right)_{eq} \qquad B_{ijk}' = \left(\frac{\partial^2 s_i}{\partial x_j \partial x_k}\right)_{eq} \tag{14}$$

It is worth pointing out that Eq. 13 can be truncated at different levels depending on the area of study. In the present work, a first-order expansion is sufficient to obtain a set of internal coordinates for the structural refinement, while the Wilson $\mathbf{B}'$ tensor is required in the field of anharmonic vibrational spectroscopy.

Concerning the calculation of accurate molecular structures, Z-matrix internal coordinates (ZICs) represent the most common choice. ZICs are mostly based on chemical intuition; however, they are not completely unambiguous. Because of this, selecting a different set of ZICs could provide a distinct outcome, indicating a substantial user-dependency. Moreover, poorly designed Z-matrices could prevent the optimization process from converging. In order to bypass this problem, a protocol based on molecular symmetry has been adopted. The latter is supposed to be maintained during the entire optimization procedure, and the formulation of suitable geometrical constraints applied to the internal coordinates is typically sufficient when ZICs are the reference set of coordinates. Although this approach has shown to be quite effective, the use several dummy atoms may be necessary when studying complex chemical topologies, which would involve the addition of numerous degrees of freedom. From this perspective, a remarkable improvement can be achieved by incorporating the symmetry in the definition of coordinates itself. More specifically, the totally-symmetric (A_1) coordinates selected from a non-redundant set of symmetry internal coordinates are used, while all the other ones are kept fixed at their guess values. In the MSR code, the delocalized internal coordinates (DICs) [105–109] are used as a reference set, since they are characterized by simple criteria for the identification of redundancies. In order to build DICs, the starting point is the definition of N_r redundant internal coordinates, which usually correspond to the so-called primitive internal coordinates (PICs), given by the full list of all bond lengths, valence and dihedral angles. DICs are built as linear combinations of PICs, the transformation matrix **U** ·being generated by the eigenvectors of $\mathbf{BB}^\dagger$ corresponding to non-null eigenvalues,

$$BB^\dagger(UR) = \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}(UR) \tag{15}$$

while $\mathbf{R}$ contains the redundancies. The selection of A_1 coordinates is carried out through a multi-step procedure which is shortly outlined here, while a full description is reported in Ref. [97]. First, the guess structure is displaced along each DIC, and the resulting geometry is converted to its Cartesian counterpart. The latter is then subjected to the application of all symmetry operations of the point group, and only if it remains unchanged, the corresponding displacement internal coordinate is marked as totally symmetric. Finally, the matrix product $\mathbf{BB}^\dagger$ expressed in terms of DICs can be cast in a block-diagonal form, each block being related to an irreducible representation of the point group. Once all A_1 coordinates have been detected, their presence in the same block is checked as a further confirmation. The use of A_1 coordinates in the optimization presents two basic advantages. In fact, it is characterized by a black-box definition of a non-redundant set of internal symmetry coordinates. Second, the generation of geometrical constraints is completely automatized in order to handle in a transparent way even large molecular systems. With the aim of highlighting the effectiveness of the mentioned procedure, a complete study at different levels will be presented later. In particular, the stability of the fit with respect to the set of coordinates has been object of interest. The accuracy of the proposed methodology will be shown through the analysis of systems of biological and astrochemical interest. Furthermore, the comparison of our results with advanced, highly accurate protocols will be object of discussion as well.

2.4 Internal-based VPT2 framework

In this section, the main aspects of the internal-based VPT2 framework are described, while interested readers are referred to Ref. [65] for a more detailed discussion.

In order to set up the expressions required for the calculation of VPT2 energies within the internal-based VPT2 framework, the first step is the definition of the vibrational Hamiltonian. At variance with the Cartesian-based formulation, the kinetic energy operator is not diagonal anymore, implying additional terms arising from the perturbative expansion.

Let us consider a set of M internal coordinates $s = \{s_1, s_2, \ldots, s_M\}$ ($M \geq N$). The expression of the kinetic energy operator (KEO) in terms of s is well known and can be stated through the introduction of the Wilson $\mathbf{G}$ matrix,

$$\mathbf{G} = \mathbf{B}\mathbf{M}^{-1}\mathbf{B}^{\mathrm{T}} \tag{16}$$

where $\mathbf{M}$ is the diagonal matrix of atomic masses and $\mathbf{B}$ is built over the coordinates s. By introducing $\tilde{\mathsf{G}} = \det(\mathbf{G})$, the expression of the operator is [35, 110, 111]

$$\mathcal{T} = -\frac{\hbar^2}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{\partial}{\partial s_i} \mathsf{G}_{ij} \frac{\partial}{\partial s_j} + \mathcal{V}_{\mathsf{g}} \tag{17}$$

where $\hbar$ is the reduced Planck constant and

$$\mathcal{V}_{\mathsf{g}} = \frac{\hbar^2}{32} \sum_{i=1}^{M} \sum_{j=1}^{M} \left[\frac{\mathsf{G}_{ij}}{\tilde{\mathsf{G}}^2} \frac{\partial \tilde{\mathsf{G}}}{\partial s_i} \frac{\partial \tilde{\mathsf{G}}}{\partial s_j} - 4\frac{\partial}{\partial s_i}\left(\frac{\mathsf{G}_{ij}}{\tilde{\mathsf{G}}} \frac{\partial \tilde{\mathsf{G}}}{\partial s_j} \right) \right] \tag{18}$$

is an inherently quantum-mechanical contribution, commonly known as extra-potential term. Similarly to the Cartesian-based treatment, the unperturbed eigenstates stem from the harmonic theory of vibrations. Within the present framework, the harmonic frequencies and normal modes are obtained through the Wilson GF method: [112]

$$GFL = L\mathbf{\Lambda} \tag{19}$$

The diagonal elements of $\mathbf{\Lambda}$ are the squared angular frequencies, the matrix L contains the eigenvectors linking the normal coordinates $Q = \{Q_1, Q_2, \ldots, Q_N\}$ to the vector s,

$$s = LQ \tag{20}$$

while the Wilson F matrix is the Hessian of the potential energy in terms of the internal coordinates, and can be directly calculated from its Cartesian counterpart H_x (see Appendix A). Since the extra-potential term can be generally approximated with its equilibrium value, it can be neglected in the calculation of transition energies. Following the introduction of the customary dimensionless normal coordinates q and their conjugate momenta p, the vibrational Hamiltonian $\mathcal{H}_v$ can be obtained,

$$\mathcal{H}_v = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} p_i \mathsf{g}_{ij} p_j + \mathcal{V} \tag{21}$$

where $\mathbf{g}$ is the G matrix expressed in wavenumbers and $\mathcal{V}$ is the potential energy operator. The anharmonic contributions due to the kinetic energies can be evaluated through the Taylor-series expansion of the g matrix,

$$\mathsf{g}_{ij} = \mathsf{g}_{ij}^{\mathrm{eq}} + \sum_{k=1}^{N} \mathsf{g}_{ij,k} q_k + \frac{1}{2} \sum_{k=1}^{N} \sum_{l=1}^{N} \mathsf{g}_{ij,kl} q_k q_l \tag{22}$$

where $\mathsf{g}_{ij}^{\mathrm{eq}} = \omega_i \delta_{ij}$ and $\mathsf{g}_{ij,kl\ldots}$ represent the derivatives of the $\mathbf{g}$ matrix expressed in wavenumbers:

$$\mathsf{g}_{ij,kl\ldots} = \left(\frac{\partial^n \mathsf{g}_{ij}}{\partial q_k \partial q_l \ldots} \right)_{\mathrm{eq}} \tag{23}$$

By inserting Eq. 23 into Eq. 21 and considering the expansion of potential energy, the following expression is obtained:

$$\mathcal{H}^{(0)} = \frac{1}{2} \sum_{i=1}^{N} \omega_i (q_i^2 + p_i^2) \tag{24}$$

$$16\chi_{ii}^{\mathcal{V}} = f_{iiii} - \frac{5f_{iii}^2}{3\omega_i} - \sum_{\substack{j=1 \\ (j \neq i)}}^{N} \frac{f_{iij}^2 (8\omega_i^2 - 3\omega_j^2)}{\omega_j (4\omega_i^2 - \omega_j^2)} \tag{28}$$

$$\mathcal{H}^{(1)} = \frac{1}{6} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \left(f_{ijk} q_i q_j q_k + 3 g_{ij,k} p_i q_k p_j \right) \tag{25}$$

$$4\chi_{ij}^{\mathcal{V}} = f_{iijj} - \frac{2f_{iij}^2 \omega_i}{4\omega_i^2 - \omega_j^2} - \frac{2f_{ijj}^2 \omega_j}{4\omega_j^2 - \omega_i^2} - \frac{f_{iii} f_{ijj}}{\omega_i} - \frac{f_{iij} f_{jjj}}{\omega_j}$$

$$- \sum_{\substack{k=1 \\ (k \neq i,j)}}^{N} \left[\frac{2\omega_k (\omega_k^2 - \omega_i^2 - \omega_j^2) f_{ijk}^2}{(\omega_i + \omega_j + \omega_k)(\omega_i - \omega_j - \omega_k)(\omega_i - \omega_j + \omega_k)(\omega_i + \omega_j - \omega_k)} + \frac{f_{iik} f_{jjk}}{\omega_k} \right] \tag{29}$$

$$\mathcal{H}^{(2)} = \frac{1}{24} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \left(f_{ijkl} q_i q_j q_k q_l + 6 g_{ij,kl} p_i q_k q_l p_j \right) \tag{26}$$

$$16\chi_{ii}^{\mathcal{T}} = 2g_{ii,ii} - \frac{3g_{ii,i}^2}{\omega_i}$$
$$- \sum_{\substack{j=1 \\ (j \neq i)}}^{N} \frac{g_{ii,j}^2 (8\omega_i^2 - 3\omega_j^2) - 4g_{ij,i}^2 \omega_j^2 + 8g_{ii,j} g_{ij,i} \omega_i \omega_j}{\omega_j (4\omega_i^2 - \omega_j^2)} \tag{30}$$

Analogously to the treatment in terms of Cartesian-based normal coordinates (CNCs), the anharmonic energies can be obtained by either CVPT or RSPT. Furthermore, thanks to the invariance of the harmonic Hamiltonian, the second-quantization formalism can be still applied. The main difference with respect to the Cartesian-based framework is that each vibrational property $\mathcal{Q}$ is composed of three contributions, namely a purely potential ($\mathcal{Q}^{\mathcal{V}}$), a purely kinetic ($\mathcal{Q}^{\mathcal{T}}$) and a cross term ($\mathcal{Q}^{\times}$). Within the internal-based VPT2, the energy is provided by the same expression as Eq. 3, where the anharmonic χ matrix is expressed as the sum of the three contributions mentioned above:

$$\chi = \chi^{\mathcal{V}} + \chi^{\mathcal{T}} + \chi^{\times} \tag{27}$$

As expected, the purely potential term does not show any difference compared with its Cartesian counterpart but the absence of the Coriolis term and the explicit expression of all contributions are reported in the following,

$$4\chi_{ij}^{\mathcal{T}} = g_{ii,jj} + g_{jj,ii}$$
$$- \frac{2(g_{ii,j}^2 \omega_i + 4g_{ij,i}^2 \omega_i - 2g_{ii,j} g_{ij,i} \omega_j)}{4\omega_i^2 - \omega_j^2}$$
$$- \frac{2(g_{jj,i}^2 \omega_j + 4g_{ij,j}^2 \omega_j - 2g_{jj,i} g_{ij,j} \omega_i)}{4\omega_j^2 - \omega_i^2}$$
$$- \frac{g_{ii,i} g_{jj,i}}{\omega_i} - \frac{g_{ii,j} g_{jj,j}}{\omega_j} \tag{31}$$

$$- \sum_{\substack{k=1 \\ (k \neq i,j)}}^{N} \left[\frac{2(g_{ij,k}^2 + g_{ik,j}^2 + g_{jk,i}^2)\omega_k(\omega_k^2 - \omega_i^2 - \omega_j^2) + 4g_{ij,k} g_{ik,j} \omega_j(\omega_j^2 - \omega_i^2 - \omega_k^2)}{(\omega_i + \omega_j + \omega_k)(\omega_i - \omega_j - \omega_k)(\omega_i - \omega_j + \omega_k)(\omega_i + \omega_j - \omega_k)} \right.$$
$$\left. + \frac{4g_{ij,k} g_{jk,i} \omega_i(\omega_i^2 - \omega_j^2 - \omega_k^2) + 8\omega_i \omega_j \omega_k g_{ik,j} g_{jk,i}}{(\omega_i + \omega_j + \omega_k)(\omega_i - \omega_j - \omega_k)(\omega_i - \omega_j + \omega_k)(\omega_i + \omega_j - \omega_k)} + \frac{g_{ii,k} g_{jj,k}}{\omega_k} \right]$$

$$16\chi_{ii}^{\times} = -\frac{2g_{ii,i} f_{iii}}{\omega_i} + \sum_{\substack{j=1 \\ (j \neq i)}}^{N} \frac{8g_{ij,i} f_{iij} \omega_i \omega_j - 2g_{ii,j} f_{iij}(8\omega_i^2 - \omega_j^2)}{\omega_j(4\omega_i^2 - \omega_j^2)} \tag{32}$$

$$4\chi_{ij}^{\times} = \frac{g_{ii,i}f_{ijj} + g_{jj,i}f_{iii}}{\omega_i} + \frac{f_{iij}g_{jj,j} + g_{ii,j}f_{jjj}}{\omega_j} - \frac{4(g_{ji,i}f_{iij}\omega_j - g_{ii,i}f_{iij}\omega_i)}{4\omega_i^2 - \omega_j^2}$$

$$- \frac{4(g_{ij,j}f_{ijj}\omega_i - g_{jj,i}f_{ijj}\omega_j)}{4\omega_j^2 - \omega_i^2} + \sum_{\substack{k=1 \\ (k \neq i,j)}}^{N} \left[\frac{4g_{ik,j}f_{ijk}\omega_i(\omega_i^2 - \omega_j^2 - \omega_k^2) + 4g_{jk,i}f_{ijk}\omega_j(\omega_j^2 - \omega_i^2 - \omega_k^2)}{(\omega_i + \omega_j + \omega_k)(\omega_i - \omega_j - \omega_k)(\omega_i - \omega_j + \omega_k)(\omega_i + \omega_j - \omega_k)} \right.$$

$$\left. + \frac{8\omega_i\omega_j\omega_k g_{ij,k}f_{ijk}}{(\omega_i + \omega_j + \omega_k)(\omega_i - \omega_j - \omega_k)(\omega_i - \omega_j + \omega_k)(\omega_i + \omega_j - \omega_k)} - \frac{g_{ii,k}f_{jjk} + f_{iik}g_{jj,k}}{\omega_k} \right] \tag{33}$$

where few misprints of Eqs. 35 and 37 of Ref. [73] have been corrected. By applying the partial fraction decomposition and introducing the tensors}

$$\eta_{ijkl} = f_{ijkl} + g_{ij,kl} + g_{kl,ij} \tag{34}$$

$$\sigma_{ijk} = f_{ijk} - (g_{ij,k} + g_{ik,j} + g_{jk,i}) \tag{35}$$

$$\rho_{ijk} = f_{ijk} - (g_{ij,k} - g_{ik,j} - g_{jk,i}) \tag{36}$$

the elements of the χ matrix can be rewritten in a more compact form, which is also required for the application of the DVPT2 scheme:

$$16\chi_{ii} = \eta_{iiii} - \frac{\sigma_{iii}^2/2 + 9\rho_{iii}^2/2}{3\omega_i}$$

$$- \frac{1}{2} \sum_{\substack{j=1 \\ (j \neq i)}}^{N} \left[\frac{4\rho_{jii}^2}{\omega_j} + \frac{\sigma_{iij}^2}{2\omega_i + \omega_j} - \frac{\rho_{iij}^2}{2\omega_i - \omega_j} \right] \tag{37}$$

$$4\chi_{ij} = \eta_{iijj} - \frac{1}{2} \left[\frac{\sigma_{iij}^2}{2\omega_i + \omega_j} + \frac{\rho_{iij}^2}{2\omega_i - \omega_j} \right]$$

$$- \frac{1}{2} \left[\frac{\sigma_{jji}^2}{2\omega_j + \omega_i} + \frac{\rho_{jji}^2}{2\omega_j - \omega_i} \right] - \frac{\rho_{iii}\rho_{ijj}}{\omega_i}$$

$$- \frac{\rho_{jjj}\rho_{jii}}{\omega_j} - \frac{1}{2} \sum_{\substack{k=1 \\ (k \neq i,j)}}^{N} \left[\frac{\sigma_{ijk}^2}{\omega_i + \omega_j + \omega_k} \right.$$

$$- \frac{\rho_{ijk}^2}{\omega_i + \omega_j - \omega_k} + \frac{\rho_{ikj}^2}{\omega_i - \omega_j + \omega_k}$$

$$\left. - \frac{\rho_{jki}^2}{\omega_i - \omega_j - \omega_k} + \frac{2\rho_{kii}\rho_{kjj}}{\omega_k} \right] \tag{38}$$

One of the most interesting aspects of Eqs. 37 and 38 is their analogy with the Cartesian-based treatment from the algebraic point of view. In the first place, the internal-based elements of χ show the same functional form as those introduced in Eqs. 4 and 5. Consequently, the internal-based χ matrix can be interpreted as a full-fledged generalization of the corresponding Cartesian counterpart, implying a remarkable simplification at the implementation level. In fact, the extension of any code implementing the Cartesian-based VPT2 becomes straightforward. Moreover, the diagnostic of Fermi resonances can be carried out through a direct extension of the Martin test. While the first step in the identification of FRs (see Eq. 6) does not require further changes, the second step becomes a straightforward generalization of Eq. 7:

$$\frac{\rho_{jki}^4}{64(1 + \delta_{jk})^2 |\omega_i - \omega_j - \omega_k|^3} \geq K^{1-2} \tag{39}$$

It is evident that Eq. 7 can be generated from Eq. 39 by setting the derivatives of the g matrix to zero, since in that case $\rho_{ijk} = f_{ijk}$.

Once the set of FRs has been identified, the corresponding interaction elements of the contact-transformed Hamiltonian can be calculated,

$$\langle v_R + 1_i | \tilde{\mathcal{H}} | v_R + 1_j + 1_k \rangle = \rho_{jki} \sqrt{\frac{(v_{R,i} + 1)(v_{R,j} + 1)(v_{R,k} + 1 + \delta_{jk})}{8(1 + 3\delta_{jk})}} \tag{40}$$

paving the route for the extension of the GVPT2 scheme to the use of internal coordinates. Let us underline that thanks to the analogy between the Cartesian- and internal-based frameworks of VPT2 in terms of the analytical expressions of wave functions, vibrational energies and diagnostic of FRs, the extension of the internal-based VPT2 for asymmetric tops to the treatment of linear, symmetric and spherical tops can be carried out through the procedure described in Ref. [65].

3 Implementation

In this section, the implementation of the algorithms employed in this work is briefly addressed, while a deeper discussion is reported elsewhere [65, 97].

3.1 Internal-based VPT2 framework

A full calculation within the internal-based VPT2 framework is carried out through different steps and exploits an interplay of implementations within a recently devised standalone code and a development version of the quantum-chemistry package GAUSSIAN. In particular, the standalone program performs the harmonic vibrational analysis, the generation of the single-point GAUSSIAN input files required for the calculation of both potential and kinetic energy derivatives with respect to the normal coordinates by finite differences, and the calculation of such derivatives. On the other side, the tasks carried out through the GAUSSIAN package are the initial optimization, the calculation of Cartesian force constants at both equilibrium and out-of-equilibrium geometries, and the application of the VPT2 framework for the calculation of the anharmonic frequencies. Summarizing, the new program is mainly aimed at generating the anharmonic force field and the Wilson g derivatives, while the internal-based VPT2 framework has been implemented in the GAUSSIAN package. Even though different quantum-chemistry programs can be employed for the generation of the derivatives required in the anharmonic treatment, in the present work the G16 [113] package has been always employed.

3.2 Calculation of semi-experimental structures

As previously discussed, the main ingredients for the calculation of SE structures are the experimental rotational constants, vibrational corrections, weights and the guess geometry. In general both the guess geometry and vibrational corrections are evaluated through a quantum-chemistry package. Similarly to the application of the VPT2 framework in internal coordinates, the G16 package has been used for this purpose. Once these preliminary operations have been carried out, the whole set of data is used to start the structural refinement by the MSR program, leading to the characterization of the molecular geometry.

Since the new code and MSR share a common set of libraries, they have been recently assembled in a single suit of programs for vibro-rotational analyses.

4 Computational details

Most of the available electronic structure methods implementing analytical second-order derivatives of potential energy have been employed. The hybrid density functional B3PW91 [114] has been used in conjunction with the jul-cc-pVDZ (hereafter julDZ) basis set [115]. Furthermore, tight d functions have been included (julDZd) for the treatment of atoms belonging to the third period. The double-hybrid functional revDSD-PBEP86 [116] and second-order Møller–Plesset perturbation theory (MP2) [117] have been employed in conjunction with the jun-cc-pVTZ (hereafter junTZ) basis set [115]. The latter has been augumented with tight d functions (junTZd) for calculations involving third-period atoms. The empirical dispersion contributions have been systematically accounted for in density functional theory (DFT) computations by means of Grimme's D3 model with Becke–Johnson damping [118, 119].

5 Results and discussion

Several systems have been analyzed in detail with the objective of validating the theoretical framework developed in the fields of both molecular structure prediction and vibrational spectroscopy. In this section, selected test cases will be discussed to highlight the main aspects characterizing the computational protocols developed so far.

5.1 Determination of SE equilibrium structures

As previously outlined, the choice of the set of coordinates represents one of the main steps needed to obtain reliable SE structures. While the effectiveness of ZICs for this type of calculations is now well recognized, in this context the attention will be focused on symmetry-based optimizations. With the aim of investigating different symmetries, the simplest Criegee intermediate (point group C_s) and thiophene (C_{2v}) have been selected as case studies. Furthermore, for both systems detailed analyses are available in the literature, which provided highly accurate SE structures. For this reason, they represent ideal test cases to validate the computational protocol discussed in the present work.

First, let us consider the Criegee intermediate (see Fig. 1), which is a planar system fully described by 7 degrees of freedom.

A detailed analysis has been performed by McCarthy and co-workers [120], where the experimental rotational

constants of nine isotopic species combined with vibrational corrections at the CCSD(T)/ANO1 [121] level of theory have been employed in the prediction of the SE structure. In the present work, the same set of experimental data has been used for the calculation of both the effective (r_0) structure and, in conjunction with vibrational corrections at the rDSD/junTZ level, its SE counterpart. However, due to the planarity of the system only two of the three moments of inertia are independent. As a matter of fact, 18 rotational constants have been actually included in the optimization. The fit has been performed employing both ZICs (see Section S1 of the Supplementary Information) and A_1-DICs as nuclear coordinates. Concerning the latter, a set of PICs has been generated from the molecular connectivity, followed by the calculation of DICs and the extraction of the totally-symmetric coordinates. As expected from a C_s planar systems, the A_1-DICs correspond to those DICs which do not alter dihedral angles. Since A_1-DICs are much less chemically intuitive if compared with ZICs, an error propagation has been applied to express both SE geometries in terms of the latter through a double-step procedure. In the first place, the variance–covariance matrix expressed in terms of A_1-DICs ($\mathbf{\Sigma}_{A_1}$) has been converted to the corresponding Cartesian-based counterpart ($\mathbf{\Sigma}_x$) through the following expression,

$$\mathbf{\Sigma}_x = \{\boldsymbol{B}_{A_1}^{\dagger}\}^{\mathrm{T}} \mathbf{\Sigma}_{A_1} \{\boldsymbol{B}_{A_1}^{\dagger}\} \tag{41}$$

where $\boldsymbol{B}_{A_1}^{\dagger}$ represents the pseudo-inverse of the Wilson $\boldsymbol{B}$ matrix expressed in terms of A_1-DICs evaluated at the optimized geometry. Secondly, the $\mathbf{\Sigma}_x$ matrix has been converted to ZICs,

$$\mathbf{\Sigma}_Z = \{\boldsymbol{B}_Z^{\dagger}\}^{\mathrm{T}} \mathbf{\Sigma}_x \{\boldsymbol{B}_Z^{\dagger}\} \tag{42}$$

where $\boldsymbol{B}_Z$ is the Wilson $\boldsymbol{B}$ matrix in terms of ZICs at the optimized geometry. The new standard deviations and confidence intervals have been obtained starting from the square roots of the diagonal elements of $\mathbf{\Sigma}_Z$. The different sets of
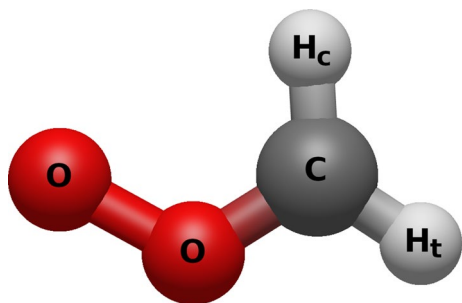
Fig. 1 Molecular structure and atom labeling of the simplest Criegee intermediate

structural parameters are reported and compared with those proposed by McCarthy and co-workers in Table 1.

Inspection of Table 1 shows that that the SE geometries obtained through ZICs and A_1-DICs exactly coincide both in terms of structural parameters and standard deviation and, what is even more important they present also structural parameters close to their reference counterparts. Consequently, the symmetry-based approach has been applied without the setup of a user-defined Z-matrix, but still reaching the same accuracy.

A further validation of the proposed methodology has been performed by the structure determination of thiophene (see Fig. 2), a C_{2v} heteroaromatic cycle.

A first characterization of this system performed by our research group employed the experimental rotational constants of eight isotopologues combined with vibrational corrections based on the B2PLYP and B3LYP functionals in conjunction with the cc-pVTZ and SNSD basis sets, respectively. [122] More recently, a detailed experimental and theoretical investigation of thiophene has been carried out by Orr and co-workers [123]. In particular, the set of experimental data has been extended to include 26 isotopic species, with the latter being used together with vibrational and electronic corrections at the CCSD(T)/cc-pCPVTZ level of theory for the structural refinement. In agreement with the calculations reported in Ref. [123], 24 isotopic species have been included in the present work. Similarly to the Criegee intermediate, only two rotational constants for each isotopologue have been considered, leading to a set of 48 experimental rotational constants usable in the nonlinear regression. At variance with Ref. [123], only vibrational corrections at the B3P/julDZd level have been used to correct the experimental data, as well as to compute the guess structure for the optimization. A full comparison of the effective and SE structure with those proposed in Ref. [123] is reported in Table 2.

In analogy with the Criegee intermediate, both the SE equilibrium structure and the standard deviation are invariant under change of coordinates. The search of totally symmetric coordinates provided the correct number of coordinates (8), and their use reaches the same structure obtained by using the Z-matrix indicated in Ref. [123] and reported in Section S1 of the Supplementary Information. Last but not least, a good agreement between the SE structural parameters retrieved in this work and those proposed by Orr and co-workers has been detected despite the different nature of the corrections to the experimental rotational constants.

5.2 Anharmonic calculations in internal coordinates

With the aim of highlighting the advantages of an internal-based formulation of VPT2, our recently developed engine has been be applied to both semi-rigid and flexible systems.

Table 1 Equilibrium molecular structure of the simplest Criegee intermediate (distances in Å, angles in degrees)

| Coordinate	$r_e^{[a]}$	ZICs[b]		A_1-DICs[b]		Ref. [120]
		r_0	$r_e^{SE\,[c]}$	r_0	$r_e^{SE\,[c]}$	$r_e^{SE\,[d]}$
$r(C-H_c)$	1.0830	1.087(1)	1.0794(2)	1.087(1)	1.0794(2)	1.0806(2)
$r(C-H_t)$	1.0797	1.079(2)	1.0761(4)	1.079(2)	1.0761(4)	1.0772(4)
$r(C-O)$	1.2696	1.2774(9)	1.2699(2)	1.2774(9)	1.2699(2)	1.2689(2)
$r(O-O)$	1.3243	1.3425(9)	1.3407(2)	1.3425(9)	1.3407(2)	1.3405(1)
$a(O-C-H_c)$	118.82	117.6(9)	118.70(3)	117.6(9)	118.70(3)	118.65(2)
$a(O-C-H_t)$	114.71	114.9(2)	114.96(7)	114.9(2)	114.96(7)	114.82(4)
$a(O-O-C)$	119.27	117.96(2)	117.834(4)	117.96(2)	117.834(4)	117.910(3)
$\sigma^{[e]}$		0.003	0.0005	0.003	0.0005	
$r^{[f]}$		11	11	11	11	

[a] Equilibrium geometry at the rDSD/junTZ level

[b] All fits have been performed on moments of inertia equally weighted

[c] Vibrational corrections at the rDSD/junTZ level of theory

[d] Vibrational corrections at the CCSD(T)/ANO1 level of theory

[e] Mean standard deviation (uÅ^2)

[f] $r = n - p$ is the number of degrees of freedom, n and p being the number of SE data and parameters, respectively
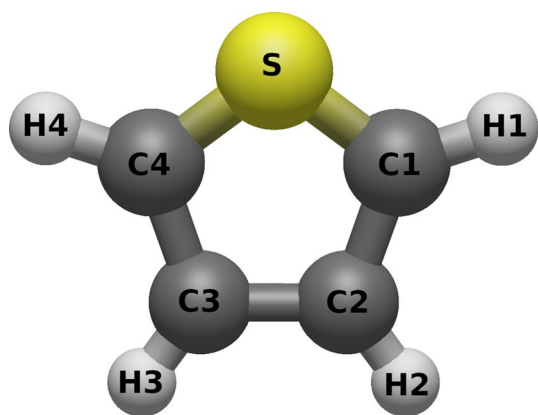
Fig. 2 Molecular structure and atom labeling of thiophene

In the former case, the choice of the coordinates set plays a more marginal role even though it leads to a remarkable reduction of inter-mode couplings. This finding paves the way to the systematic use of cost-effective protocols aimed at calculating vibrational properties. In fact, the reduction of couplings between different vibrations results in the calculation of low-dimensionality anharmonic force fields, which represent the bottleneck of an anharmonic calculation. This effect is even stronger when flexible systems are taken into account, especially concerning the couplings involving LAMs.

In the present work, 1,1-difluoroethylene (see Fig. 3) has been selected as test-case for semi-rigid systems.

The calculation of the fundamental vibrational frequencies at the anharmonic level has been performed by employing all VPT2 schemes discussed in the previous sections within both internal- and Cartesian-based frameworks. A full comparison of the results with the experimental data is reported in Table 3.

As expected, the Cartesian- and internal-based calculations do not show any difference in the resulting transition frequencies. It is also worth specifying that Coriolis contributions have been properly considered in the calculations based on Cartesian coordinates. As a matter of fact, specific second-order terms in the expansion of the KEO yield contributions formally equivalent to the Coriolis terms. Despite the GVPT2 results obtained on the basis of Cartesian and internal coordinates are numerically indistinguishable, the magnitude of the couplings is remarkably affected by the choice of the nuclear coordinates. This pattern has been also observed for terms which do not contribute directly to the calculation of anharmonic transition energies in this context, such as three-mode quartic force constants.

Inspection of Fig. 4 shows that the magnitude of inter-mode terms is definitively reduced when switching from the Cartesian-based representation of normal coordinates to that employing curvilinear coordinates. Furthermore, vibrational couplings are not transferred from potential to kinetic energy, since there exist only six first-order derivatives of the Wilson **g** matrix exceeding 100 cm^{-1}, while all second-order derivatives are well below this value.

Table 2 Equilibrium molecular structure of thiophene (distances in Å, angles in degrees)

Coordinate	$r_e^{[a]}$	ZICs[b]		A_1-DICs[b]		Ref. [123]
		r_0	$r_e^{SE\,[c]}$	r_0	$r_e^{SE\,[c]}$	r_e^{SE}
r(S−C1)	1.7168	1.7167(6)	1.7106(2)	1.7167(6)	1.7106(2)	1.71049(18)
r(C1−C2)	1.3733	1.3699(8)	1.3666(2)	1.3699(8)	1.3666(2)	1.36564(31)
r(C1−H1)	1.0870	1.0766(5)	1.0756(1)	1.0766(5)	1.0756(1)	1.07714(17)
r(C2−H2)	1.0892	1.0777(5)	1.0779(1)	1.0777(5)	1.0779(1)	1.07856(14)
r(C2−C3)	1.4255	1.430(1)	1.4229(3)	1.430(1)	1.4229(3)	1.4224(10)
a(C1−S−C4)/2	46.07	46.07(2)	46.035(5)	46.07(2)	46.035(5)	46.024(7)
a(S−C1−C2)	111.51	111.57(5)	111.61(1)	111.57(5)	111.61(1)	111.608(16)
a(S−C1−H1)	120.14	119.67(7)	120.26(2)	119.67(7)	120.26(2)	120.065(28)
a(C1−C2−H2)	123.35	123.45(7)	123.43(2)	123.45(7)	123.43(2)	123.414(23)
a(C1−S−C4)	92.14	92.13(4)	92.07(1)	92.13(4)	92.07(1)	92.047(15)
$\sigma^{[e]}$		0.002	0.0006	0.002	0.0006	
$r^{[f]}$		40	40	40	40	

[a] Equilibrium geometry at the B3P/julDZd level

[b] All fits have been performed on moments of inertia equally weighted

[c] Vibrational corrections at the B3P/julDZd level of theory

[d] Vibrational and electronic corrections at the CCSD(T)/cc-pCPVTZ level of theory

[e] Mean standard deviation (uÅ^2)

[f] $r = n − p$ is the number of degrees of freedom, n and p being the number of SE data and parameters, respectively
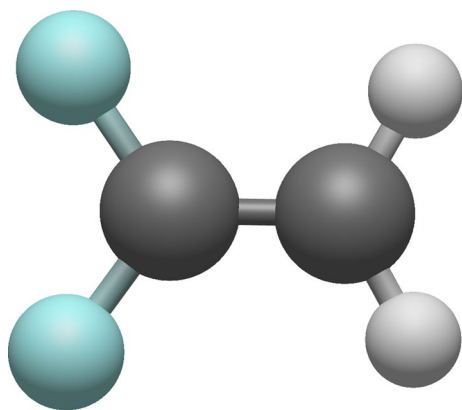
Fig. 3 Molecular structure of 1,1-difluoroethylene

The AAT conformer of glycolic acid (see Fig. 5) has been selected as a flexible system presenting LAMs.

A recent detailed analysis of the conformers of glycolic acid [126] shows that most of them are characterized by the presence of this kind of vibrations, poorly described at the VPT2 level. In this work, the AAT conformer has been considered, since it is characterized by two distinct LAMs, representing then an ideal system for studying also inter-mode couplings between different LAMs. The simulations have been performed employing a dual-level method based on the substitution approach [73], where the anharmonic calculation at the B3P/julDZ level of theory has been refined by the inclusion of harmonic frequencies evaluated at the

rDSD/junTZ level (hereafter rDSD/junTZ//B3P/julDZ). Furthermore, the reduced-dimensionality scheme has been used, so that anharmonic corrections have been applied to all vibrations but LAMs. The set of fundamental wavenumbers obtained within both the Cartesian- and internal-based GVPT2 framework are compared with their experimental counterparts in Table 4.

The data reported in Table 4 show that both sets of GVPT2 transition frequencies are in remarkable agreement with their experimental counterparts, with a mean absolute error (MAE) always lower than 10 cm^{-1}. Despite the similarity of the results for Cartesian and internal coordinates, the behavior of the inter-mode couplings is very different, especially when interaction terms involving LAMs are examined. In order to highlight this aspect, the contribution of LAMs has been investigated by comparing the quartic force constants involving modes 20 and 21, which are reported in Fig. 6.

It is apparent that both diagonal and off-diagonal terms are much larger in Cartesian coordinates (exceeding 80,000 cm^{-1} for $f_{21,21,21,21}$), while the values are definitely more reasonable for the internal-based counterparts. Interestingly, this behavior also involves couplings between LAMs and stretching modes. As an example, the values the semi-diagonal quartic force constants $f_{20,20,1,1}$ and $f_{21,21,1,1}$ expressed in CNCs are, respectively, −3733 and −9347 cm^{-1}, while the corresponding internal-based counterparts are −27 and −62 cm^{-1}. Hence, the internal-based representation of vibrations allows not only for an improved

Table 3 Comparison of the Cartesian and curvilinear VPT2, DVPT2, and GVPT2 wavenumbers (in cm^{-1}) of 1,1-difluoroethylene at the MP2/junTZ level with the experimental data

| State	Assign.[a]	Symm	ω	Cartesian[b]			Curvilinear[b]			Exp.[b]
				ν_{VPT2}	ν_{DVPT2}	ν_{GVPT2}	ν_{VPT2}	ν_{DVPT2}	ν_{GVPT2}	
$|1_1\rangle$	CH$_2$ (S) str	A$_1$	3246	3139	3117	3083	3139	3119	3085	3058
$|1_2\rangle$	C-C str		1778	1744	1738	1743	1745	1738	1743	1743
$|1_3\rangle$	CH$_2$ sciss		1423	1354	1377	1360	1355	1380	1361	1359
$|1_4\rangle$	CF$_2$ (S) str		944	929	929	929	929	929	929	925
$|1_5\rangle$	CF$_2$ sciss		555	550	550	550	550	550	550	550
$|1_6\rangle$	C-C twist	A$_2$	730	714	714	714	713	713	713	708
$|1_7\rangle$	CH$_2$ op wag	B$_1$	813	789	789	789	792	792	792	802
$|1_8\rangle$	CF$_2$ op wag		641	631	631	631	631	631	631	610
$|1_9\rangle$	CH$_2$ (A) str	B$_2$	3360	3222	3222	3222	3222	3222	3222	3176
$|1_{10}\rangle$	CF$_2$ (A) str		1331	1296	1296	1296	1296	1296	1296	1301
$|1_{11}\rangle$	CH$_2$ rock		970	951	951	951	951	951	951	954
$|1_{12}\rangle$	CF$_2$ rock		438	437	437	437	437	437	437	437
MAE[c]				15	15	10	15	15	10	

[a] Ref. [124]

[b] Ref. [125]

[c] Mean absolute error with respect to the experimental data
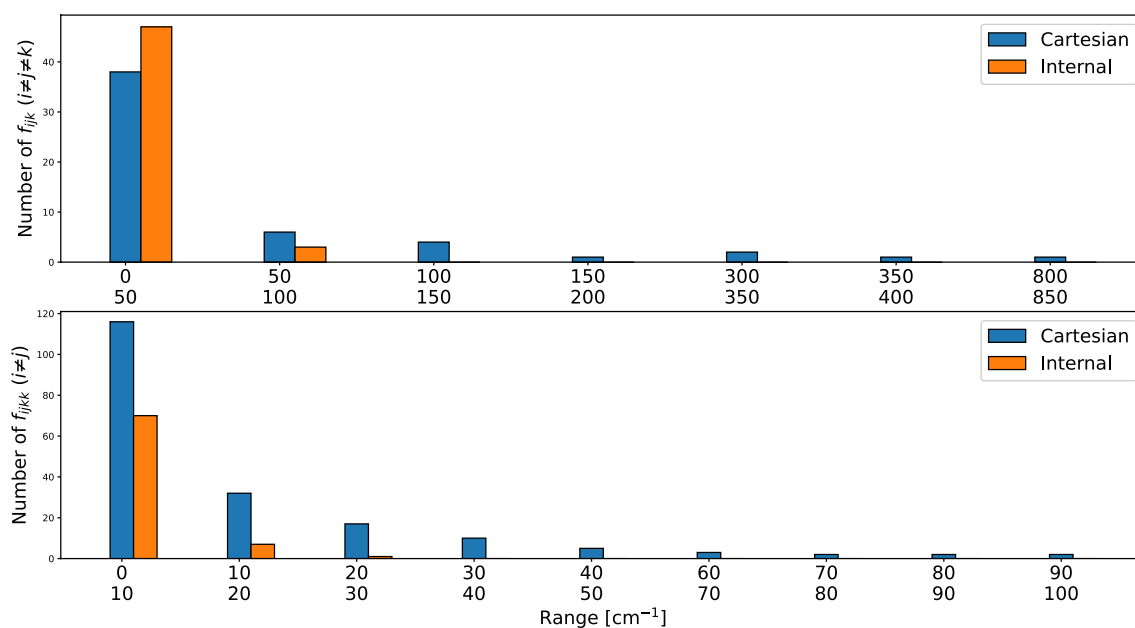
Fig. 4 Comparison of the number of cubic (f$_{ijk}$ ($i \neq j \neq k$)) and quartic (f$_{ijkk}$ ($i \neq j$)) force constants of 1,1-difluoroethylene above a given threshold (in cm^{-1}) computed at the MP2/junTZ level with Cartesian or curvilinear coordinates
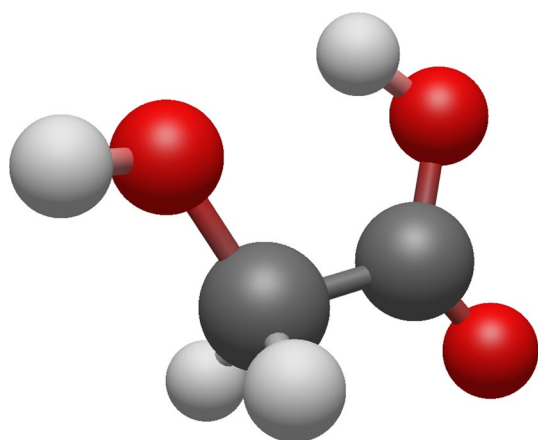
Fig. 5 Molecular structure of the AAT conformer of glycolic acid

separation of LAMs from SAMs, but it also lays the foundations for a multi-mode, one-dimensional treatment of LAMs.

6 Conclusion

In this work, I have presented a general engine for vibrational and rotational spectroscopy based on curvilinear internal coordinates.

Concerning the determination of semi-experimental molecular structures, the new methodology provides a transparent way for the *a priori* definition of the nuclear coordinates required in the optimization process and a unambiguous definition of geometrical constraints without loss of accuracy. The algorithm implemented into the MSR software enables SE nonlinear regression starting from a minimal set of input data without any intermediate, user-dependent step. The applications presented in this context confirmed our initial hypotheses and previous studies, paving the route for the study of larger systems.

The outcomes of a number of test cases demonstrate the reliability of the new GVPT2 engine based on curvilinear internal coordinates and allowing to effectively handle medium-to-large-size molecules. It is worth mentioning that this methodology is completely general for all those electronic structure methods implementing analytical Cartesian force constants. Comparison with the standard Cartesian-based framework shows two main benefits. The first one is the ease of implementation from an existing code based on the Cartesian formulation of VPT2. The second one is a significant improvement in the separation of large amplitude motions and

Table 4 Comparison of the Cartesian and curvilinear GVPT2 fundamental wavenumbers (in cm^{-1}) of the AAT conformer of glycolic acid with the experimental data

State	ω[a]	Cartesian[b]	Curvilinear[b]	Exp
$\lvert 1_1 \rangle$	3873	3681	3678	3672[c]
$\lvert 1_2 \rangle$	3705	3470	3475	3474[c]
$\lvert 1_3 \rangle$	3112	2976	2975	2976[d]
$\lvert 1_4 \rangle$	3062	2965	2960	2952[d]
$\lvert 1_5 \rangle$	1849	1817	1814	1806[c]
$\lvert 1_6 \rangle$	1503	1455	1450	1448[e]
$\lvert 1_7 \rangle$	1433	1390	1384	1388[e]
$\lvert 1_8 \rangle$	1390	1352	1346	1360[e]
$\lvert 1_9 \rangle$	1282	1258	1249	–
$\lvert 1_{10} \rangle$	1252	1224	1217	1198[e]
$\lvert 1_{11} \rangle$	1173	1132	1127	1136[e]
$\lvert 1_{12} \rangle$	1086	1048	1046	1059[e]
$\lvert 1_{13} \rangle$	1032	992	987	–
$\lvert 1_{14} \rangle$	863	848	844	846[e]
$\lvert 1_{15} \rangle$	683	677	652	653[e]
$\lvert 1_{16} \rangle$	627	615	608	–
$\lvert 1_{17} \rangle$	570	553	545	559[f]
$\lvert 1_{18} \rangle$	512	505	499	503[f]
$\lvert 1_{19} \rangle$	313	303	292	309[f]
$\lvert 1_{20} \rangle$	130	–	–	–
$\lvert 1_{21} \rangle$	87	–	–	–
MAE[g]		8	7	

[a] Harmonic frequencies at the rDSD/junTZ level of theory

[b] Hybrid model based on the substitution approach combining harmonic frequencies at the rDSD/junTZ level of theory and anharmonic corrections at the B3P/julDZ level

[c] Average of measurements of Refs. [127] (IR, Ar matrix), [128] (IR N_2 matrix) and [129] (Raman, Ar matrix)

[d] From Raman Ar matrix measurement (Ref. [129])

[e] From IR Ar matrix measurement (Ref. [127])

[f] Average of measurements of Refs. [127] (IR, Ar matrix) and [129] (Raman, Ar matrix)

[g] Mean absolute error (in cm^{-1}) with respect to the experimental data. States $\lvert 1_9 \rangle, \lvert 1_{13} \rangle, \lvert 1_{16} \rangle, \lvert 1_{20} \rangle$ and, $\lvert 1_{21} \rangle$ excluded

their further treatment by more appropriate methodologies. In fact, a significant reduction in the inherent problems of VPT2 applied to this kind of systems has been verified at different levels. More sophisticated strategies for treating LAMs are currently under study in our research group, with the aim to study molecular systems featured by a growing number of this type of vibrational motions.
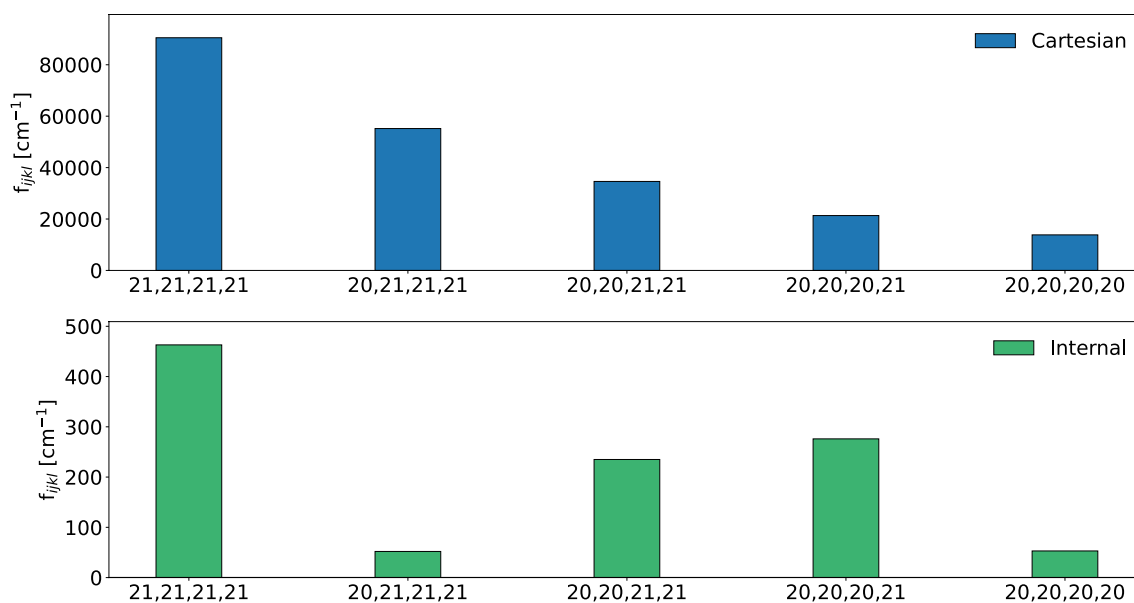
Fig. 6 Comparison of the Cartesian (top panel) and curvilinear (bottom panel) quartic force constants of the AAT conformer of glycolic acid involving modes 20 and 21 at the rDSD/junTZ//B3P/julDZ level of theory

Appendix A Transformation of force constants

One of the main steps in the internal-based anharmonic treatment is the conversion of the harmonic force constants, usually computed in terms of Cartesian coordinates, to internal coordinates. Let us consider a generic molecular geometry and the corresponding Cartesian gradient and force constants, referred to as g_x and H_x, respectively. In order to convert them to a basis of M internal coordinates represented by the vector s, the first step is the calculation of the internal-based gradient vector,

$$g_s = (B^\dagger)^T P g_x \tag{A1}$$

where B is the Wilson B matrix built over s, $B^\dagger$ is its pseudo-inverse, while P is the projection matrix employed to remove the contribution of translational and rotational motions.

Then, the calculation of the first-order Cartesian derivative of the B, namely B' is carried out in order to derive a compact expression for the Hessian matrix in internal coordinates, namely F,

$$F = (B^\dagger)^T P (H_x - g_s B') P (B^\dagger) \tag{A2}$$

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00214-023-03069-7.

Declarations

Conflict of interest The author declares no conflict of interest.

References

1. 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(13):8131–8191. https://doi.org/10.1021/acs.chemrev.9b00007

2. Nielsen HH (1951) The vibration–rotation energies of molecules. Rev Mod Phys 23(2):90–136. https://doi.org/10.1103/RevModPhys.23.90

3. Mills IM (1972) Vibration-rotation structure in asymmetric- and symmetric-top molecules. In: Rao KN, Mathews CW (eds) Molecular spectroscopy: modern research, Chap 3.2. Academic Press, New York, pp 115–140. https://doi.org/10.1016/B978-0-12-580640-4.50013-3

4. Clabo DA Jr, Allen WD, Remington RB, Yamaguchi Y, Schaefer HF III (1988) A systematic study of molecular vibrational anharmonicity and vibration-rotation interaction by self-consistent-field higher-derivative methods. Asymmetric top molecules. Chem. Phys. 123(2):187–239. https://doi.org/10.1016/0301-0104(88)87271-9

5. Allen WD, Yamaguchi Y, Császár AG, Clabo DA Jr, Remington RB, Schaefer HF III (1990) A systematic study of molecular vibrational anharmonicity and vibration-rotation interaction by self-consistent-field higher-derivative methods. Linear polyatomic molecules. Chem Phys 145(3):427–466. https://doi.org/10.1016/0301-0104(90)87051-C

6. Barone V (2005) Anharmonic vibrational properties by a fully automated second-order perturbative approach. J Chem Phys 122(1):014108. https://doi.org/10.1063/1.1824881

7. Barone V (2004) Vibrational zero-point energies and thermodynamic functions beyond the harmonic approximation. J Chem Phys 120(7):3059–3065. https://doi.org/10.1063/1.1637580

8. Krasnoshchekov SV, Isayeva EV, Stepanov NF (2012) Numerical-analytic implementation of the higher-order canonical van Vleck perturbation theory for the interpretation of medium-sized molecule vibrational spectra. J Phys Chem A 116(14):3691–3709. https://doi.org/10.1021/jp211400w

9. 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(2):261–300. https://doi.org/10.1080/00268976.2013.808386

10. Carter S, Sharma AR, Bowman JM, Rosmus P, Tarroni R (2009) Calculations of rovibrational energies and dipole transition intensities for polyatomic molecules using multimode. J Chem Phys 131(22):224106. https://doi.org/10.1063/1.3266577

11. Rauhut G, Hrenar T (2008) A combined variational and perturbational study on the vibrational spectrum of P_2F_4. Chem Phys 346(1–3):160–166. https://doi.org/10.1016/j.chemphys.2008.01.039

12. Christiansen O (2012) Selected new developments in vibrational structure theory: potential construction and vibrational wave function calculations. Phys Chem Chem Phys 14(19):6672–6687. https://doi.org/10.1039/C2CP40090A

13. Carbonnière P, Dargelos A, Pouchan C (2010) The VCI-P code: an iterative variation-perturbation scheme for efficient computations of anharmonic vibrational levels and ir intensities of polyatomic molecules. Theoret Chem Acc 125(3–6):543–554. https://doi.org/10.1007/s00214-009-0689-7

14. Pouchan C, Zaki K (1997) Ab initio configuration interaction determination of the overtone vibrations of methyleneimine in the region 2800–3200 cm^{-1}. J Chem Phys 107(2):342–345. https://doi.org/10.1063/1.474395

15. Christiansen O (2007) Vibrational structure theory: new vibrational wave function methods for calculation of anharmonic vibrational energies and vibrational contributions to molecular properties. Phys Chem Chem Phys 9(23):2942–2953. https://doi.org/10.1039/B618764A

16. Bowman JM (1978) Self-consistent field energies and wavefunctions for coupled oscillators. J Chem Phys 68(2):608–610. https://doi.org/10.1063/1.435782

17. Bowman JM (1986) The self-consistent-field approach to polyatomic vibrations. Acc Chem Res 19(7):202–208. https://doi.org/10.1021/ar00127a002

18. Jung JO, Gerber RB (1996) Vibrational wave functions and spectroscopy of $(H_2O)_n$, $n = 2, 3, 4, 5$: Vibrational self-consistent field with correlation corrections. J Chem Phys 105(23):10332–10348. https://doi.org/10.1063/1.472960

19. Carter S, Culik SJ, Bowman JM (1997) Vibrational self-consistent field method for many-mode systems: A new approach and application to the vibrations of co adsorbed on cu(100). J Chem Phys 107(24):10548–10469. https://doi.org/10.1063/1.474210

20. Bulik IW, Frisch MJ, Vaccaro PH (2017) Vibrational self-consistent field theory using optimized curvilinear coordinates. J Chem Phys 147(4):044110. https://doi.org/10.1063/1.4995440

21. Whitehead RJ, Handy NC (1975) Variational calculation of vibration–rotation energy levels for triatomic molecules. J Mol Spectrosc 55(1):356–373. https://doi.org/10.1016/0022-2852(75)90274-X

22. Dunn KM, Boggs JE, Pulay P (1986) Vibrational energy levels of hydrogen cyanide. J Chem Phys 85(10):5838–5846. https://doi.org/10.1063/1.451545

23. Cassam-Chenaï P, Liévin J (2006) The VMFCI method: a flexible tool for solving the molecular vibration problem. J Comput Chem 27(5):627–640. https://doi.org/10.1002/jcc.20374

24. Carrington T (2017) Perspective: Computing (ro-) vibrational spectra of molecules with more than four atoms. J Chem Phys. https://doi.org/10.1063/1.4979117

25. Davidson ER (1975) The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. J Comput Phys 17(1):87–94. https://doi.org/10.1016/0021-9991(75)90065-0

26. Sleijpen G, Van der Vorst H (2000) A Jacobi–Davidson iteration method for linear eigenvalue problems. SIAM Rev 42(2):267–293. https://doi.org/10.1137/S0036144599363084

27. Ribeiro F, Iung C, Leforestier C (2005) A Jacobi–Wilson description coupled to a Block-Davidson algorithm: an efficient scheme to calculate highly excited vibrational levels. J Chem Phys 123(5):054106. https://doi.org/10.1063/1.1997129

28. Petrenko T, Rauhut G (2017) A new efficient method for the calculation of interior eigenpairs and its application to vibrational structure problems. J Chem Phys 146(12):124101. https://doi.org/10.1063/1.4978581

29. Handy NC, Carter S (2004) Large vibrational variational calculations using 'multimode' and an iterative diagonalization technique. Mol Phys 102(21–22):2201–2205. https://doi.org/10.1080/00268970410001728870

30. Rauhut G (2007) Configuration selection as a route towards efficient vibrational configuration interaction calculations. J Chem Phys 127(18):184109. https://doi.org/10.1063/1.2790016

31. Carrington T Jr (2021) Using collocation to study the vibrational dynamics of molecules. Spectrochim Acta A 248:119158. https://doi.org/10.1016/j.saa.2020.119158

32. Manzhos S, Wang X, Carrington T Jr (2018) A multimode-like scheme for selecting the centers of gaussian basis functions when computing vibrational spectra. Chem Phys 509:139–144. https://doi.org/10.1016/j.chemphys.2017.10.006

33. Császár AG, Fábri C, Szidarovszky T, Mátyus E, Furtenbacher T, Czakó G (2012) The fourth age of quantum chemistry: molecules in motion. Phys Chem Chem Phys 14(3):1085–1106. https://doi.org/10.1039/C1CP21830A

34. Mátyus E, Gábor C, Császár AG (2009) Toward black-box-type full- and reduced-dimensional variational (ro)vibrational

computations. J Chem Phys 130(13):134112. https://doi.org/10.1063/1.3076742

35. Lauvergnat D, Nauts A (2002) Exact numerical computation of a kinetic energy operator in curvilinear coordinates. J Chem Phys 116(19):8560–8570. https://doi.org/10.1063/1.1469019

36. Lauvergnat D, Baloïtcha E, Dive G, Desouter-Lecomte M (2006) Dynamics of complex molecular systems with numerical kinetic energy operators in generalized coordinates. Chem Phys 326(2–3):500–508. https://doi.org/10.1016/j.chemphys.2006.03.012

37. Scribano Y, Lauvergnat DM, Benoit DM (2010) Fast vibrational configuration interaction using generalized curvilinear coordinates and self-consistent basis. J Chem Phys doi 10(1063/1):3476468

38. Tennyson J (2016) Perspective: accurate ro-vibrational calculations on small molecules. J Chem Phys 145(12):124112. https://doi.org/10.1063/1.4962907

39. Sutcliffe BT, Tennyson J (1991) A general treatment of vibration-rotation coordinates for triatomic molecules. Int J Quantum Chem 39(2):183–196. https://doi.org/10.1002/qua.560390208

40. Yurchenko SN, Thiel W, Jensen P (2007) Theoretical rovibrational energies (TROVE): a robust numerical approach to the calculation of rovibrational energies for polyatomic molecules. J Mol Spectrosc 245(2):126–140. https://doi.org/10.1016/j.jms.2007.07.009

41. Petit AS, McCoy AB (2013) Diffusion monte Carlo in internal coordinates. J Phys Chem A 117(32):7009–7018. https://doi.org/10.1021/jp312710u

42. Bulik IW, Frisch MJ, Vaccaro PH (2018) Fixed-node, importance-sampling diffusion Monte Carlo for vibrational structure with accurate and compact trial states. J Chem Theory Comput 14(3):1554–1563. https://doi.org/10.1021/acs.jctc.8b00016

43. Yagi K, Keçeli M, Hirata S (2012) Optimized coordinates for anharmonic vibrational structure theories. J Chem Phys. https://doi.org/10.1063/1.4767776

44. Zimmerman PM, Smereka P (2016) Optimizing vibrational coordinates to modulate intermode coupling. J Chem Theory Comput 12(4):1883–1891. https://doi.org/10.1021/acs.jctc.5b01168

45. Thomsen B, Yagi K, Christiansen O (2014) Optimized coordinates in vibrational coupled cluster calculations. J Chem Phys. https://doi.org/10.1063/1.4870775

46. Arnaud Leclerc PST, Carrington T (2017) Comparison of different eigensolvers for calculating vibrational spectra using low-rank, sum-of-product basis functions. Mol. Phys. 115(15–16):1740–1749. https://doi.org/10.1080/00268976.2016.1249980

47. White SR (1992) Density matrix formulation for quantum renormalization groups. Phys Rev Lett 69:2863–2866. https://doi.org/10.1103/PhysRevLett.69.2863

48. Schollwöck U (2011) The density-matrix renormalization group in the age of matrix product states. Ann Phys 326(1):96–192. https://doi.org/10.1016/j.aop.2010.09.012

49. Baiardi A, Reiher M (2020) The density matrix renormalization group in chemistry and molecular physics: recent developments and new challenges. J Chem Phys 152(4):040903. https://doi.org/10.1063/1.5129672

50. Baiardi A, Stein CJ, Barone V, Reiher M (2017) Vibrational density matrix renormalization group. J Chem Theory Comput 13(8):3764–3777. https://doi.org/10.1021/acs.jctc.7b00329

51. Christiansen O (2003) Møller–Plesset perturbation theory for vibrational wave functions. J Chem Phys 119(12):5773–5781. https://doi.org/10.1063/1.1601593

52. Christiansen O (2004) Vibrational coupled cluster theory. J Chem Phys 120(5):2149–2159. https://doi.org/10.1063/1.1637579

53. Truhlar DG, Olson RW, Jeannotte AC, Overend J (1976) Anharmonic force constants of polyatomic molecules. Test of the procedure for deducing a force field from the vibration-rotation spectrum. J Am Chem Soc 98(9):2373–2379. https://doi.org/10.1021/ja00425a001

54. Gaw FJ, Willetts A, Handy NC, Green WH (1992) SPECTRO—a program for derivation of spectroscopic constants from provided quartic force fields and cubic dipole fields. In: Bowman JM (ed) Adv Mol Vib Collis Dyn, vol 1. JAI Press, Greenwich, CT

55. Van Vleck JH (1929) On σ-type doubling and electron spin in the spectra of diatomic molecules. Phys Rev 33(4):467–506. https://doi.org/10.1103/PhysRev.33.467

56. Pulay P, Meyer W, Boggs JE (1978) Cubic force constants and equilibrium geometry of methane from Hartree–Fack and correlated wavefunctions. J Chem Phys 68(11):5077–5085. https://doi.org/10.1063/1.435626

57. 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(6):1301–1324. https://doi.org/10.1021/acs.jpca.0c09526. (**PMID: 33506678**)

58. Krasnoshchekov SV, Dobrolyubov EO, Syzgantseva MA, Palvelev RV (2020) Rigorous vibrational Fermi resonance criterion revealed: two different approaches yield the same result. Mol Phys 118(11):1743887. https://doi.org/10.1080/00268976.2020.1743887

59. 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. https://doi.org/10.3389/fspas.2021.665232

60. Martin JML, Taylor PM (1997) Accurate ab initio quartic force field for trans-HNNH and treatment of resonance polyads. Spectrochim Acta A 53(8):1039–1050. https://doi.org/10.1016/S1386-1425(96)01869-0

61. Kuhler KM, Truhlar DG, Isaacson AD (1996) General method for removing resonance singularities in quantum mechanical perturbation theory. J Chem Phys 104(12):4664–4670. https://doi.org/10.1063/1.471161

62. 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(3):1015–1036. https://doi.org/10.1021/ct200814m

63. Plíva J (1990) Anharmonic constants for degenerate modes of symmetric top molecules. J Mol Spectrosc 139(2):278–285. https://doi.org/10.1016/0022-2852(90)90065-X

64. 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(15):948–982. https://doi.org/10.1002/qua.24931

65. Mendolicchio M, Bloino J, Barone V (2021) 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. https://doi.org/10.1021/acs.jctc.1c00240

66. Gong JZ, Matthews DA, Changala PB, Stanton JF (2018) Fourth-order vibrational perturbation theory with the Watson Hamiltonian: Report of working equations and preliminary results. J Chem Phys 149(11):114102. https://doi.org/10.1063/1.5040360

67. Tew DP, Handy NC, Carter S, Irle S, Bowman J (2003) The internal coordinate path Hamiltonian; application to methanol and malonaldehyde. Mol Phys 101(23–24):3513–3525. https://doi.org/10.1080/0026897042000178079

68. Bowman JM, Huang X, Handy NC, Carter S (2007) Vibrational levels of methanol calculated by the reaction path version of

MULTIMODE, using an ab initio, full-dimensional potential. J Phys Chem A 111(31):7317–7321. https://doi.org/10.1021/jp070398m

69. Miller WH, Handy NC, Adams JE (1980) Reaction path Hamiltonian for polyatomic molecules. J Chem Phys 72(1):99–112. https://doi.org/10.1063/1.438959

70. Page M, McIver JW (1988) On evaluating the reaction path Hamiltonian. J Chem Phys 88(2):922–935. https://doi.org/10.1063/1.454172

71. Jackels CF, Gu Z, Truhlar DG (1995) Reaction-path potential and vibrational frequencies in terms of curvilinear internal coordinates. J Chem Phys 102(8):3188–3201. https://doi.org/10.1063/1.468630

72. Baiardi A, Bloino J, Barone V (2017) Simulation of vibronic spectra of flexible systems: Hybrid DVR-harmonic approaches. J Chem Theory Comput 13(6):2804–2822. https://doi.org/10.1021/acs.jctc.7b00236

73. Mendolicchio M, Bloino J, Barone V (2022) Perturb-then-diagonalize vibrational engine exploiting curvilinear internal coordinates. J Chem Theory Comput 18(12):7603–7619. https://doi.org/10.1021/acs.jctc.2c00773

74. Quade CR (1976) Internal coordinate formulation for the vibration–rotation energies of polyatomic molecules. J Chem Phys 64(7):2783–2795. https://doi.org/10.1063/1.432577

75. Isaacson AD (2006) Including anharmonicity in the calculation of rate constants. 1 the HCN/HNC isomerization reaction. J Phys Chem A 110(2):379–388. https://doi.org/10.1021/jp058113y

76. Harris DO, Engerholm GG, Gwinn WD (1965) Calculation of matrix elements for one-dimensional quantum-mechanical problems and the application to anharmonic oscillators. J Chem Phys 43(5):1515–1517. https://doi.org/10.1063/1.1696963

77. Dickinson A, Certain P (1968) Calculation of matrix elements for one-dimensional quantum-mechanical problems. J Chem Phys 49(9):4209–4211. https://doi.org/10.1063/1.1670738

78. Light JC, Hamilton IP, Lill JV (1985) Generalized discrete variable approximation in quantum mechanics. J Chem Phys 82(3):1400–1409. https://doi.org/10.1063/1.448462

79. Bačić Z, Light J (1986) Highly excited vibrational levels of "floppy'' triatomic molecules: a discrete variable representation-distributed gaussian basis approach. J Chem Phys 85(8):4594–4604. https://doi.org/10.1063/1.451824

80. Choi SE, Light J (1990) Determination of the bound and quasibound states of Ar-HCl van der Waals complex: discrete variable representation method. J Chem Phys 92(4):2129–2145. https://doi.org/10.1063/1.458004

81. Karabulut H, Sibert EL III (1997) Trigonometric discrete variable representations. J Phys B: At Mol Opt Phys 30(15):513. https://doi.org/10.1088/0953-4075/30/15/001

82. Colbert DT, Miller WH (1992) A novel discrete variable representation for quantum mechanical reactive scattering via the S-matrix Kohn method. J Chem Phys 96(3):1982–1991. https://doi.org/10.1063/1.462100

83. Light JC, Carrington T Jr (2000) Discrete-variable representations and their utilization. Adv Chem Phys 114:263–310. https://doi.org/10.1002/9780470141731.ch4

84. Carrington T, Miller WH (1984) Reaction surface Hamiltonian for the dynamics of reactions in polyatomic systems. J Chem Phys 81(9):3942–3950. https://doi.org/10.1063/1.448187

85. Koch A, Billing GD (1997) The reaction volume Hamiltonian model: further development and application. J Chem Phys 107(18):7242–7251. https://doi.org/10.1063/1.474965

86. Watson JKG (1968) Simplification of the molecular vibration–rotation Hamiltonian. Mol Phys 15(5):479–490. https://doi.org/10.1080/00268976800101381

87. Schuurman MS, Allen WD, von Ragué Schleyer P (2005) Schaefer III HF (2005) The highly anharmonic BH_5 potential energy surface characterized in the ab initio limit. J. Chem. Phys. 122(10):104302. https://doi.org/10.1063/1.1853377

88. Krasnoshchekov SV, Isayeva EV, Stepanov NF (2014) Criteria for first- and second-order vibrational resonances and correct evaluation of the Darling-Dennison resonance coefficients using the canonical Van Vleck perturbation theory. J Chem Phys 141(23):234114. https://doi.org/10.1063/1.4903927

89. Fermi E (1931) Über den ramaneffekt des kohlendioxyds. Z. Phys. A 71(3–4):250–259. https://doi.org/10.1007/BF01341712

90. Martin JML, Lee TJ, Taylor PM, François J-P (1995) The anharmonic force field of ethylene, C_2H_4, by means of accurate ab initio calculations. J Chem Phys 103(7):2589–2602. https://doi.org/10.1063/1.469681

91. Yang Q, Bloino J (2022) An effective and automated processing of resonances in vibrational perturbation theory applied to spectroscopy. J Phys Chem A 126(49):9276–9302. https://doi.org/10.1021/acs.jpca.2c06460

92. 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(2):548–556. https://doi.org/10.1021/acs.accounts.7b00603

93. Demaison J, Boggs JE, Császár AG (2016) Equilibrium molecular structures from spectroscopy to quantum chemistry. CRC Press, Boca Raton. https://doi.org/10.1201/b10374

94. Demaison J (2007) Experimental, semi-experimental and ab initio equilibrium structures. Mol Phys 105(23–24):3109–3138. https://doi.org/10.1080/00268970701765811

95. Flygare W (1974) Magnetic interactions in molecules and an analysis of molecular electronic charge distribution from magnetic parameters. Chem Rev 74(6):653–687. https://doi.org/10.1021/cr60292a003

96. Penocchio E, Mendolicchio M, Tasinato N, Barone V (2016) Structural features of the carbon–sulfur chemical bond: a semi-experimental perspective. Can J Chem 94(12):1065–1076. https://doi.org/10.1139/cjc-2016-0282

97. Mendolicchio M, Penocchio E, Licari D, Tasinato N, Barone V (2017) Development and implementation of advanced fitting methods for the calculation of accurate molecular structures. J Chem Theory Comput 13(6):3060–3075. https://doi.org/10.1021/acs.jctc.7b00279

98. Licari D, Fusè M, Salvadori A, Tasinato N, Mendolicchio M, Mancini G, Barone V (2018) Towards the smart workflow system for computational spectroscopy. Chem Phys 20:26034–26052. https://doi.org/10.1039/C8CP03417F

99. Ye H, Mendolicchio M, Kruse H, Puzzarini C, Biczysko M, Barone V (2020) The challenging equilibrium structure of HSSH: another success of the rotational spectroscopy/quantum chemistry synergism. J Mol Struct 1211:127933. https://doi.org/10.1016/j.molstruc.2020.127933

100. Melli A, Melosso M, Tasinato N, Bosi G, Spada L, Bloino J, Mendolicchio M, Dore L, Barone V, Puzzarini C (2018) Rotational and infrared spectroscopy of ethanimine: a route toward its astrophysical and planetary detection. ApJ 855(2):123. https://doi.org/10.3847/1538-4357/aaa899

101. Obenchain DA, Spada L, Alessandrini S, Rampino S, Herbers S, Tasinato N, Mendolicchio M, Kraus P, Gauss J, Puzzarini C, Grabow J-U, Barone V (2018) Unveiling the sulfur-sulfur bridge: accurate structural and energetic characterization of a Homochalcogen intermolecular bond. Angew Chem Int Ed 57(48):15822–15826. https://doi.org/10.1002/anie.201810637

102. Sim G, Sutton L, Bartell L, Romenesko D, Wong T (1975) Augmented analyses: method of predicate observations. In: Molecular structure by diffraction methods, pp 72–80

103. Belsley DA (1991) Conditioning diagnostics. Wiley Online Library, New York

104. Champion J, Robiette AG, Mills I, Graner G (1982) Simultaneous analysis of the ν_1, ν_4, $2\nu_2$, $\nu_2 + \nu_5$, and $2\nu_5$ infrared bands of ^{12}CH$_3$F. J Mol Spectrosc 96(2):422–441. https://doi.org/10.1016/0022-2852(82)90207-7

105. Baker J (1997) Constrained optimization in delocalized internal coordinates. J Comput Chem 18(8):1079–1095. 10.1002/(SICI)1096-987X(199706)18:8<1079::AID-JCC12>3.0.CO;2-8

106. Baker J, Kessi A, Delley B (1996) The generation and use of delocalized internal coordinates in geometry optimization. J Chem Phys 105(1):192–212. https://doi.org/10.1063/1.471864

107. Bakken V, Helgaker T (2002) The efficient optimization of molecular geometries using redundant internal coordinates. J Chem Phys 117(20):9160–9174. https://doi.org/10.1063/1.1515183

108. Paizs B, Fogarasi G, Pulay P (1998) An efficient direct method for geometry optimization of large molecules in internal coordinates. J Chem Phys 109(16):6571–6576. https://doi.org/10.1063/1.477309

109. Jensen F, Palmer DS (2011) Harmonic vibrational analysis in delocalized internal coordinates. J Chem Theory Comput 7(1):223–230. https://doi.org/10.1021/ct100463a

110. Podolsky B (1928) Quantum-mechanically correct form of Hamiltonian function for conservative systems. Phys Rev 32(5):812. https://doi.org/10.1103/PhysRev.32.812

111. Schaad LJ, Hu J (1989) The Schrödinger equation in generalized coordinates. J Mol Struct (Thoechem) 185:203–215. https://doi.org/10.1016/0166-1280(89)85014-6

112. Wilson EB (1939) A method of obtaining the expanded secular equation for the vibration frequencies of a molecule. J Chem Phys 7(11):1047–1052. https://doi.org/10.1063/1.1750363

113. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ (2016) Gaussian 16 Revision A.03. Gaussian Inc. Wallingford

114. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 98(7):5648–5652. https://doi.org/10.1063/1.464913

115. 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(10):3027–3034. https://doi.org/10.1021/ct200106a

116. 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(24):5129–5143. https://doi.org/10.1021/acs.jpca.9b03157

117. Møller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. Phys Rev 46(7):618–622. https://doi.org/10.1103/PhysRev.46.618

118. 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(15):154104. https://doi.org/10.1063/1.3382344

119. Grimme S, Ehrlich S, Goerigk L (2011) Effect of the damping function in dispersion corrected density functional theory. J Comput Chem 32(7):1456–1465. https://doi.org/10.1002/jcc.21759

120. McCarthy MC, Cheng L, Crabtree KN, Martinez O Jr, Nguyen TL, Womack CC, Stanton JF (2013) The simplest Criegee intermediate (H$_2$C=O-O): isotopic spectroscopy, equilibrium structure, and possible formation from atmospheric lightning. J Phys Chem Lett 4(23):4133–4139. https://doi.org/10.1021/jz4023128

121. Almlöf J, Taylor PR (1987) General contraction of gaussian basis sets. I. Atomic natural orbitals for first- and second-row atoms. J Chem Phys 86(7):4070–4077. https://doi.org/10.1063/1.451917

122. 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(10):4689–4707. https://doi.org/10.1021/acs.jctc.5b00622

123. Orr VL, Ichikawa Y, Patel AR, Kougias SM, Kobayashi K, Stanton JF, Esselman BJ, Woods RC, McMahon RJ (2021) Precise equilibrium structure determination of thiophene (c-C$_4$H$_4$S) by rotational spectroscopy-structure of a five-membered heterocycle containing a third-row atom. J Chem Phys 154(24):244310. https://doi.org/10.1063/5.0055267

124. Krasnoshchekov SV, Craig NC, Stepanov NF (2013) Anharmonic vibrational analysis of the gas-phase infrared spectrum of 1,1-difluoroethylene using the operator van vleck canonical perturbation theory. J Phys Chem A 117(14):3041–3056. https://doi.org/10.1021/jp311398z

125. McKean DC, Law MM, Groner P, Conrad AR, Tubergen MJ, Feller D, Moore MC, Craig NC (2010) Infrared spectra of CF$_2$=CHD and CF$_2$=CD$_2$: scaled quantum-chemical force fields and an equilibrium structure for 1,1-difluoroethylene. J Phys Chem A 114(34):9309–9318. https://doi.org/10.1021/jp104498n

126. 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(15):2373–2387. https://doi.org/10.1021/acs.jpca.2c01419

127. Hollenstein H, Ha T-K, Günthard HH (1986) IR induced conversion of rotamers, matrix spectra, ab initio calculation of conformers, assignment and valence force field of trans glycolic acid. J Mol Struct 146:289–307. https://doi.org/10.1016/0022-2860(86)80300-3

128. Halasa A, Lapinski L, Reva I, Rostkowska H, Fausto R, Nowak MJ (2014) Near-infrared laser-induced generation of three rare conformers of glycolic acid. J Phys Chem A 118(30):5626–5635. https://doi.org/10.1021/jp5051589

129. Ahokas JM, Kosendiak I, Krupa J, Wierzejewska M, Lundell J (2018) High vibrational overtone excitation-induced conformational isomerization of glycolic acid in solid argon matrix. J Raman Spectrosc 49(12):2036–2045. https://doi.org/10.1002/jrs.5474