



Cite this: *Phys. Chem. Chem. Phys.*,  
2024, 26, 29732

## Tailored anharmonic potential energy surfaces for infrared signatures†

Janine Hellmers, Pascal Czember and Carolin König \*

Accurately calculated infrared spectra are essential for supporting experimental interpretation, yet full-space anharmonic vibrational structure calculations are only feasible for a limited number of degrees of freedom. Fortunately, characteristic spectroscopic signatures are often dominated by a few key vibrations. We propose a computational protocol specifically tailoring high dimensional anharmonic potential energy surfaces for the accurate and efficient calculation of such spectral signatures with vibrational coupled cluster response theory. Our protocol focuses on the selection of appropriate coordinates for the relevant degrees of freedom and the identification of specific mode-coupling terms for the potential energy surface that require more thorough treatment. This includes applying different levels of electronic structure theory and selecting a restricted set of higher mode-coupling terms (> mode pairs). We validate this protocol on two spectral regions: the fundamental C=O stretching vibrations in uracil and the fundamental OH stretchings in catechol. Our findings indicate that the convergence behaviour towards harmonic frequencies in the so-called FALCON algorithm is an effective indicator for the locality character of the relevant degrees of freedom. We find that the C=O stretchings in uracil are better described using normal coordinates, while the description with local FALCON coordinates of the OH-stretching vibrations in catechol showed superior performances in VCC spectra calculations. Overall, our protocol offers valuable guidelines for accurate and efficient anharmonic calculation of vibrational spectral signatures.

Received 23rd July 2024,  
Accepted 20th November 2024

DOI: 10.1039/d4cp02916j

rsc.li/pccp

### 1 Introduction

Vibrational spectroscopic tools have been proven indispensable to reveal biomolecular processes on an atomistic level. For instance, difference infrared (IR) spectroscopy can provide unique insights into phenomena such as temporary water wires in rhodopsin.<sup>1–4</sup> Also, the proton transfer in the blue light sensor using flavin (BLUF) upon light stimuli can be detected by changes in hydrogen-bonding patterns through IR spectroscopy.<sup>5</sup> Often, the interpretation of experimental results requires computational aid. However, accurately calculating the vibrational properties of large biomolecular systems poses a formidable computational challenge, especially when anharmonicity or nuclear quantum effects emerge. This is, for instance, evident for spectral regions where hydrogen-bonded

stretching vibrations absorb.<sup>6,7</sup> Vibrational structure methods offer a route to obtain accurate theoretical descriptions of vibrational wave functions accounting for these effects.<sup>8,9</sup> However, full-space treatments of larger molecular systems are impossible by today's means. Fortunately, full interpretation of experimental results is often not necessary: spectral signatures are usually of primary interest, as these and their changes in chemical reactions are often characteristic, such as C=O stretching vibrations that are affected by hydrogen bonding. Since the spectral signatures are presumably influenced by only a few degrees of freedom (DOFs), tailoring the computational setup and focusing the computational effort on the region of interest, while applying less accurate but more efficient treatments for the remaining system, is highly desirable.<sup>10</sup> The computation of anharmonic IR spectra is rather complex, and finding the best strategy for tailoring the computational setup is not trivial. A successful computation involves several key components: (i) a suitable description of the vibrational DOFs and the selection of the vibrational (sub-)space, (ii) an efficient generation of the potential energy and dipole surfaces, and (iii) the subsequent vibrational wave function calculation. In our work, we focus on tailoring strategies for the first two points. An overview of tailoring strategy ideas for correlated vibrational wave functions is given in ref. 10.

*Institut of Physical Chemistry and Electrochemistry, Leibniz University Hannover, Germany. E-mail: carolin.koenig@pci.uni-hannover.de*

† Electronic supplementary information (ESI) available: A list of calculated underlying surfaces for the calculated spectra along with the visualization of modes for the applied vibrational spaces in this work and chain length checks ensuring converged damped linear response calculations with the band-Lanczos module. Furthermore, it contains a detailed discussion about the stability of certain potential energy surfaces of catechol and additional information about the multilevel approaches. See DOI: <https://doi.org/10.1039/d4cp02916j>



An important component of tailoring strategies is the choice of vibrational coordinates. One relatively simple, yet previously successful<sup>11–14</sup> approach, involves targeting only a subset of vibrational DOFs. For the coordinates, spanning these reduced vibrational spaces, alternative rectilinear coordinates with a more local nature than traditional normal coordinates are advertised.<sup>10</sup> These coordinates are known to achieve faster convergence of the *n*-mode expanded<sup>15</sup> potential energy surface (PES).<sup>16–22</sup> For instance, some employ so-called localized modes,<sup>11,23</sup> while others utilize local normal modes obtained from partial Hessian analysis.<sup>12–14</sup> As a prerequisite, localized modes require a full normal mode analysis,<sup>22</sup> which can be tedious to generate for large biomolecular systems. In comparison, the partial Hessian analysis requires only diagonalizations of blocks referring to disjoint subsets of atoms. However, the resulting vibrational DOFs include artificial contributions of rotational and translational DOFs and usually lack the description of relative motion.<sup>10</sup> In contrast, the so-called flexible adaption of local coordinates of nuclei (FALCON)<sup>24</sup> are free of any translational or rotational contributions and can provide coordinates that describe the relative motions of groups of atoms. As in the local monomers approach,<sup>22</sup> the construction of FALCON coordinates does not require a full normal mode analysis as a prerequisite. Moreover, the FALCON algorithm can be easily applied to covalently bounded systems and offers a so-called “growing scheme” that successively increases the vibrational subspace. Conceivable starting points for the growing scheme can be a group of atoms, a single atom, or several independent atoms in the molecular system, *e.g.* (groups of) atoms relevant to the spectroscopic signature of interest. This growing scheme is highly attractive for tailored approaches as it offers a systematic expansion of FALCON-generated vibrational subspaces.<sup>10</sup> In this work, this systematic expansion of FALCON-generated vibrational subspaces is pursued for anharmonic vibrational structure calculations for the first time.

The generation of the PES remains, despite the incremental *n*-mode expansion,<sup>15</sup> a large computational bottleneck for systems with many DOFs.<sup>16</sup> Besides using local coordinates, a conceivable tailoring strategy for PESs can be realized through multilevel potentials. The exploration of multilevel potentials has remained an ongoing field of research.<sup>8,13,16,25–48</sup> In most of these studies, the low-order coupling potentials of the *n*-mode expansion are described through higher resolutions, *e.g.* higher electronic structure levels and accurate functional forms. Higher-coupling terms are represented with a lower resolution, *e.g.* lower electronic structure levels or more approximate functions, for example, with a quartic force field (QFF).<sup>28,49</sup> Some studies further describe selected higher-coupling terms on the high level by identifying them with several techniques, *e.g.* screening measures on low-cost PESs up to four-mode coupling terms.<sup>16,35,40,41</sup> For example, Yagi and co-workers<sup>28</sup> suggested an inexpensive calculation of indices to determine important mode-coupling terms. In their case, less important coupling terms are recycled at lower resolution, while important coupling terms on the same level

are described at higher resolution. In the following, we refer only to the term multilevel instead of multiresolution. In our case, all coupling orders are approximated by the same potential energy function form, where the high and low level rather define the applied level of electronic structure theory. We concentrate on multilevel property surfaces but instead of assigning one electronic structure method to one mode-coupling level, we employ high- and lower-level descriptions for selected sets of modes or mode-coupling terms within the same mode-coupling level. We, hence, do not aim at a good representation of the full vibrational spectrum, but rather of vibrational signatures. For this, a reduced number of mode-coupling terms relevant to the spectroscopic signature of interest are calculated with a higher level of electronic structure theory, while remaining terms with the same coupling order are treated with a lower level of theory.

Another approach to accelerate the PES generation involves the use of fragmented PES calculations.<sup>22,50–53</sup> For instance, the Steele group<sup>54</sup> recently studied the impact of many-body terms in fragmented PES generations of rather large clathrate-like alkali metal cation hydrates (comprising 20 water molecules). Their findings revealed that higher-order mode-coupling terms are predominantly influenced by one-body contributions. This understanding of vibrational coupling behaviour offers valuable physical insights and presents a promising way to reduce computational costs for larger biomolecular systems. However, this approach lies beyond the scope of the current investigation.

In another study, the same group<sup>25</sup> demonstrated the efficacy of combining localized-mode methods<sup>23</sup> with multilevel approaches to accelerate the PES generation (non-fragmented) on vibrational-self-consistent field (VSCF) calculations of zero-point and fundamental excitation energies. They first employed a distance-based criterion<sup>55</sup> to determine the coupling terms described in the low level. Subsequently, they recalculate the significant mode combinations with a more elaborate high-level electronic description. For the less significant terms, the low-level description is reused. They restrict their approach to mode pair (2MC) terms and do not concentrate the computational effort on a specific spectral region. In contrast to that, we combine local-mode FALCON representations with multilevel techniques to develop a robust strategy for generating PESs for accurate spectroscopic signals. Here, we also take higher-coupling terms (3MC, 4MC) into account and investigate the performance on intensities obtained with correlated mode calculations by vibrational coupled cluster (VCC) theory. By this, we present a reliable protocol for tailored multilevel vibrational structure calculations for spectroscopic signatures. Our protocol involves the investigation of systematically increasing subspaces, obtained through the FALCON growing scheme. Additionally, we systematically enhance the accuracy of property surfaces (PESs and dipole moment surfaces (DMSs)) by incorporating selected higher coupling terms (> mode pairs) to the initial low-level property surface and describe selected coupling terms in the high-level electronic structure theory. We assess our protocol for pyrimidine-2,4(1*H*,3*H*)-dione



(uracil) within the frequency range of 1600–1900  $\text{cm}^{-1}$ , where  $\text{C}=\text{O}$  stretchings of amid-groups occur, and at the frequency range of 3500–3900  $\text{cm}^{-1}$  for 1,2-dihydroxybenzen (catechol), where OH-stretching vibrations absorb. These spectral signatures serve as representative signatures with anharmonic characters, akin to characteristic spectral signatures in biomolecules. Due to their small system size, uracil and catechol are excellent test models for studying the spectral features common in biomolecular environments. For both molecules, we also compare to experimental spectra. We note, however, that theory–experiment comparison at this accuracy is challenging. The first assessment of the developed protocol in this work assumes molecules in vacuum at 0 K. Obviously, no corresponding experimental results are available. The experiments that are closest to these conditions are presumably vibrational spectra in noble gas matrices, which we have chosen as experimental reference for uracil.<sup>56</sup> Also here, however, concentration,<sup>57</sup> temperature/annealing<sup>58</sup> and matrix effects<sup>59</sup> can have a significant impact on the signal's shape, which leads to insecurities in the comparison of the spectra.<sup>56</sup> These insecurities are even more pronounced when compared to gas phase spectra, as in the case of catechol. In this work, we, hence, do not expect to exactly reproduce the experimental spectra. The experiment is rather used as a guideline for the validation of the methods developed.

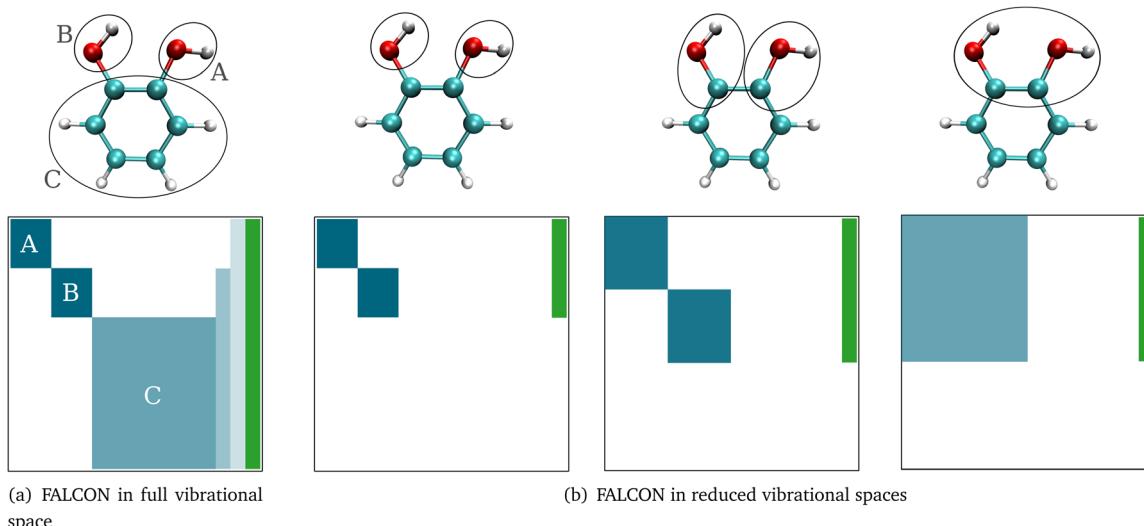
After a brief introductory background on the methodology of FALCON coordinates and how multilevel property surfaces are generated in this work, we introduce our computational protocol designed to enhance the efficiency of anharmonic vibrational spectra calculations. This protocol is specifically tailored to capture distinct spectroscopic signatures. After reporting the computational details, we demonstrate the efficiency of our protocol on  $\text{C}=\text{O}$  stretching signatures of the uracil molecule.

We continue with the investigation of the spectroscopic signatures originating from OH stretches in catechol. Here, we perform a comprehensive evaluation of FALCON-generated subspaces for subsequent anharmonic calculation and assess the suitability of different choices of underlying rectilinear coordinates for multilevel spectra.

## 2 Methodology

### 2.1 FALCON coordinates

Here, we briefly introduce FALCON coordinates and the related terminology used in this work. For details on the FALCON algorithm, see ref. 24. FALCON coordinates are iteratively constructed rectilinear coordinates with well-defined locality on certain groups of atoms. They are purely vibrational coordinates, free from global translational and rotational contributions. Two different types of local coordinates can be generated: (i) the first option is to employ the FALCON algorithm on the full vibrational space, resulting in spatially local coordinates that describe the intra-group motion, and inter-group coordinates describing relative motion between rigid groups with fixed structures. Generally, intra-group vibrational modes involve only non-zero displacement of the nuclei within this group, while the remaining nuclei have zero displacement. To further illustrate the nature of intra and inter-group FALCON coordinates, the character of these FALCON modes is shown in Fig. 1a. Here, FALCON coordinates spanning the full vibrational space are generated, where the coordinates are kept local to the two OH groups (A) and (B), and the remaining benzene ring (C). The intra-group FALCON coordinates of the two OH groups exhibit the highest degree of locality, followed by the intra-group coordinates of the benzene ring. Further,



**Fig. 1** Sketch illustrating the locality character of FALCON coordinates spanning the (a) full vibrational space or (b) reduced vibrational spaces of the molecule catechol. Circles denote groups with defined intra-group local modes. Matrices below provide a schematic representation of the complete set of full quasi-harmonic displacement vectors. White, blue, and green regions represent no contribution from the nuclei, purely (semi-)local vibrational coordinates, and translational/rotational contributions of the corresponding FALCON coordinate, respectively. Deeper shades of blue indicate a higher degree of localization. Further information can be found in the main text.



inter-group coordinates describing the relative motion of (B)  $\leftrightarrow$  (C) and (BC)  $\leftrightarrow$  (A) are created with rigid structures of (A)–(C).

The second option in the FALCON algorithm is to generate (ii) reduced vibrational spaces with the so-called growing scheme. It systematically expands the vibrational space, starting from user-defined growing seeds. More than one growing seed can be defined as a starting point. The growing seeds refer to an initial group, which are input-defined disjoint (groups of) atoms. The vibrational subspace is expanded iteratively through the fusion of (initial) groups into the groups containing the growing seeds. This fusion is determined by a coupling estimate, for instance, based on the full Hessian. If the full Cartesian Hessian proves to be the bottleneck for the investigated system, a distance-based criterion can be applied, which has been found to yield similar results for small molecules.<sup>24</sup> The growing process can be stopped at any iteration step, creating reduced vibrational spaces. The diagonalization of the corresponding reduced mass-weighted Hessian results in quasi-harmonic frequencies and quasi-harmonic modes and are unlike harmonic frequencies and normal modes not directly obtained from the eigenvalues and eigenvectors of the full mass-weighted Hessian. This is exemplarily shown in Fig. 1b. The expansion of the vibrational space starts from the two OH-groups of catechol as growing seeds and each remaining atom of the benzene ring represents one group. In the first fusion step, the two carbon atoms of the benzene ring are fused into the growing group, resulting in a reduced space with six FALCON coordinates and a high degree of locality. In a subsequent fusion, the two groups are merged into one vibrational subspace (right in Fig. 1b), resulting in already rather delocalized twelve FALCON coordinates. If the iteration process is not halted, the growing scheme converges to the full vibrational space with coordinates equivalent to normal coordinates.

## 2.2 Multilevel potential energy and dipole moment surfaces

The PES  $V(q_1, q_2, q_3, \dots, q_M)$  as a function of  $M$  vibrational coordinates  $q_1, q_2, q_3, \dots, q_M$  can be expressed in terms of a set of approximations<sup>8,9,15,60–65</sup>

$$V^{(1)}, V^{(2)}, \dots, V^{(M)}. \quad (1)$$

Here,  $V^{(1)}$  refers to uncoupled anharmonic one-mode oscillations,  $V^{(2)}$  to two-mode coupling terms, and so on. This series enables the numerical representation of the PES through lower-dimensional potential energy functions  $\bar{V}^{\mathbf{m}}$  within the restricted mode-coupling<sup>60</sup> approximation

$$V(q_1, q_2, q_3, \dots, q_M) = \sum_{\mathbf{m} \in \text{MCR}[V]} \bar{V}^{\mathbf{m}}, \quad (2)$$

where  $\mathbf{m}$  describes a set of modes, also referred to as mode combination (MC), with the dimensionality of  $\mathbf{m} = (m_1, m_2, \dots, m_M)$ . The bar-potentials  $\bar{V}^{\mathbf{m}}$  (or sub-potentials) are functions of only the modes present in the respective MC  $\mathbf{m}$ . The final set of the MCs included for the PES representation are collected in the mode-combination range (MCR). If all one-, two-, ...,  $M$ -mode couplings are included in the MCR, the

expansion becomes exact. For example, the sub-potentials for one- and two-mode couplings of mode  $m_1$  and  $m_2$  read

$$\bar{V}^{m_1} = V(q_{m_1}, 0, \dots) = V^{m_1}, \quad (3)$$

$$\begin{aligned} \bar{V}^{m_1 m_2} &= V(q_{m_1}, 0, \dots, 0, q_{m_2}, \dots) \\ &\quad - V(q_{m_1}, 0, \dots, 0, q_{m_2} = 0, \dots) \\ &\quad - V(q_{m_1} = 0, 0, \dots, 0, q_{m_2}, \dots) \\ &= V^{m_1 m_2} - V^{m_1} - V^{m_2}, \end{aligned} \quad (4)$$

respectively. For a comprehensive overview of the restricted mode-coupling approach, we refer the reader to ref. 8 and references therein. The introduction of the MCR offers a highly flexible and general formulation of the PES, as selected MC terms can be included or omitted.<sup>16</sup> Nonetheless, to obtain a correct expression for every included MC, all subsets of this MC have to be included in the MCR. Hence, the final MCR has to be closed on forming subsets.

The MCR formalism further enables a multilevel description  $V^{\text{ML}}$  of the underlying potential of the form

$$\begin{aligned} V^{\text{ML}} \approx & \sum_{\mathbf{m} \in \text{MCR}^{\text{HL}}[V]} \bar{V}^{\mathbf{m}, \text{HL}} + \sum_{\mathbf{m} \in \text{MCR}^{\text{LL}}[V]} \bar{V}^{\mathbf{m}, \text{LL}} \\ & - \sum_{\mathbf{m} \in \text{MCR}^{\text{HL}}[V]} \bar{V}^{\mathbf{m}, \text{LL}}, \end{aligned} \quad (5)$$

where the multilevel (ML) PES is obtained through a summation of the sub-potential terms described in the high level (HL) and the sub-potential terms described by a lower level (LL). Low-level terms that are also present in the high-level MCR are subtracted from the expression. Different (also more than two) levels for the electronic structure or orders of coupling (one-mode, two-mode, three-mode, ...,  $M$ -mode couplings) are viable for the low and high levels. In this work, the corresponding terms for the high and low levels in eqn (5) are obtained individually in a sum-over-product (SOP) form. The SOP Hamiltonian for each level is defined as a sum over  $T$  products

$$H = \sum_{t=1}^T \prod_{m \in \mathbf{m}^t} h^{m, t}, \quad (6)$$

where  $h^{m, t}$  refers to one-mode operators and  $\mathbf{m}^t$  refers to the MCs for term  $t$ . To form the multilevel SOP Hamiltonian, non-existing low-level terms are added to the high-level SOP Hamiltonian.

An attractive tool for an efficient black-box calculation of the lower-dimensional sub-potentials in the SOP form represents the adaptive density guided approach (ADGA).<sup>26,66,67</sup> Grid points for the sub-potentials are efficiently placed, guided by subsequent vibrational wave function calculations. In comparison to previous studies employing the multilevel variant of the ADGA scheme,<sup>16,26,27,68–71</sup> we do not restrict low and high levels to the respective mode-coupling order but rather restrict high- and low-level descriptions to sets of modes tailored to a spectroscopic signature. Our preceding formulation was limited to the energy surface but is equally valid for dipole moment surfaces.



### 2.3 Computational protocol

Our computational protocol for generating tailored property surfaces for spectral signatures is displayed in Fig. 2. The overall procedure to generate anharmonic IR spectra is presented in the right panel of Fig. 2. Firstly (i), the structure of the investigated system is optimized, followed by the (ii) generation of vibrational coordinates and vibrational (reduced) space. With these modes (iii), the potential energy surface (PES) and dipole moment surface (DMS) operators are generated in a SOP form, followed by the (iv) vibrational wave function calculation. The two steps in the focus are step (ii) and (iii), which are elaborated on the right panel of Fig. 2. These two steps will be explained in detail in the following.

**2.3.1 Selection of suitable coordinates.** Selecting appropriate coordinates spanning the vibrational space at truncated  $n$ -mode expansion poses a challenge, as emphasized in prior research.<sup>16</sup> As a selection guideline, we suggest the FALCON<sup>24</sup> algorithm's growing scheme, which provides a user-friendly way for systematic expansions of vibrational space and coordinate representations. The growing seeds usually start from groups representing the spectral signature, *e.g.* C=O or OH groups. As explained above, the growing scheme creates reduced vibrational spaces with corresponding quasi-harmonic frequencies. The convergence behaviour of the quasi-harmonic frequencies, referring to the modes of interest, with increasing vibrational space size can guide the selection of the vibrational space and coordinates: rapid convergence for increasing vibrational space size suggests that these FALCON-generated reduced spaces are likely suitable for subsequent anharmonic calculations. For applications of larger molecules, the use of the full vibrational space is unfeasible due to the

large number of DOFs. Here, the growing scheme can provide an initial estimate of the vibrational subspace size and assist in identifying which atoms should be considered for spanning the vibrational space.

**2.3.2 Multilevel potential energy and dipole moment surface generation.** We suggest a series of steps to tailor the property surfaces according to the needs of the investigated spectral signatures: (1) once the type of vibrational coordinates (local or normal) and the vibrational space (reduced or full) are assessed, we start with the low-level generation of the two-mode coupled (2MC) property surface within the considered vibrational space. For this, we apply the ADGA scheme.<sup>26,66,67</sup> (2) Subsequently, higher coupling terms of a reduced number of modes that are relevant to the spectroscopic signatures of interest are selected and computed in the lower level of electronic structure theory. For example, if the first estimate to the PES includes two-mode couplings (2MC) of 12  $M$ , we add four-mode (4MC) couplings for a smaller space, such as 6  $M$ . The choice of these selected higher coupling terms is, for instance, inspired by smaller spaces generated in the FALCON growing scheme. In case, normal coordinates are chosen as vibrational coordinates, the reduced number of modes relevant to the region of interest can be selected in a way that they exhibit similar vibrational behaviour to the FALCON coordinates of the smaller space. Alternatively, higher-coupling terms relevant to the region of interest can be selected with (pre)-screening estimates<sup>8,16,25,28,32,35,40,46,72,73</sup> assessing the importance of certain mode-mode couplings. In our work, we utilize a Hamiltonian-based screening measure,<sup>16,72</sup> which describes an upper bound to a maximal contribution of a particular MC to any matrix element. Calculating the screening measure

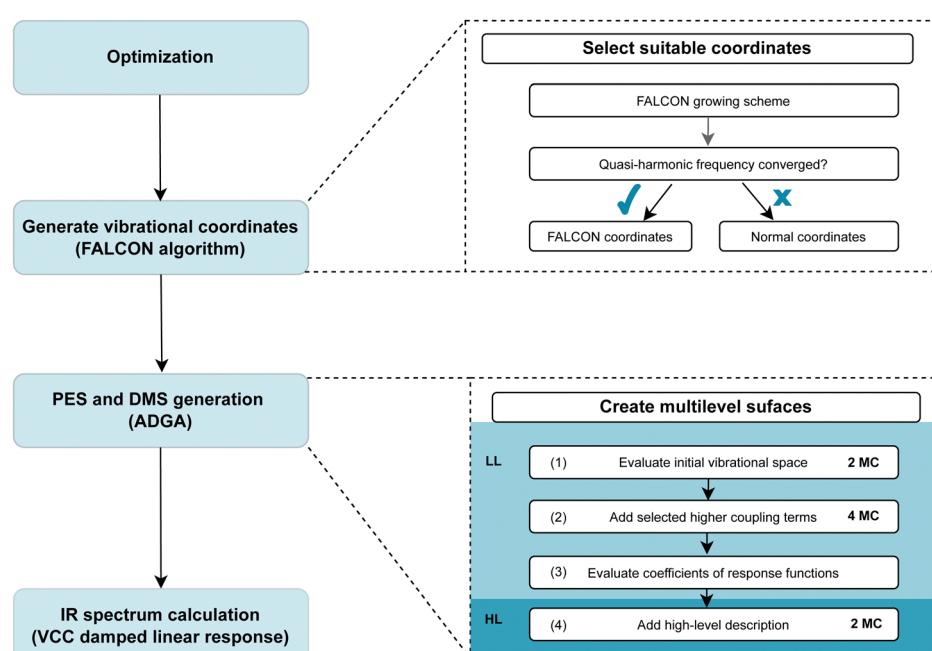


Fig. 2 Visualization of the computational protocol suggested for generating tailored multilevel property surfaces for spectral signatures. For more details see explanations in the main text.



necessitates a vibrational-self-consistent field (VSCF) calculation, which, in turn, relies on potential operators. So far, this screening estimate has been applied on higher-coupled potentials obtained on a lower level of theory to select the most important MC. Here, we assess the importance of higher-order MCs based on estimates for 2MC coupled PESs. (3) To determine the modes most relevant to the vibrational signatures of interest, we suggest performing an anharmonic vibrational calculation on the low-level property surfaces created so far. The anharmonic calculation can be performed with vibrational coupled cluster (VCC)<sup>8,9</sup> damped linear response functions.<sup>68–70</sup> By evaluating the coefficients of the response functions, we select the modes that exhibit the largest contribution to the spectral region of interest. (4) Subsequently, we conduct a high-level description, which involves generating 2MC terms with a higher level of electronic structure theory. Certainly, more than two levels are applicable within our protocol for tailored multilevel property surfaces.

### 3 Computational details

The investigated molecules uracil and catechol (Fig. 3) were optimized using the Gaussian16<sup>74</sup> program package employing the B2PLYP<sup>75</sup> functional and a Dunning-type correlation consistent triple- $\zeta$  basis set<sup>76</sup> with augmented<sup>77</sup> basis functions (aug-cc-pVTZ). Furthermore, a dispersion correction<sup>78</sup> (D3) with Becke-Johnson damping<sup>79</sup> (BJ) was added. The SCF convergence threshold was set to tight. The same setting was used for the analytic Hessian calculation. With the Hessian as input, the vibrational coordinates were generated with the program package MidasCpp2022.04.0.<sup>80</sup> For the growing scheme of the FALCON algorithm, the initial groups are represented by each atom. As initial growing seeds, we chose for uracil the two oxygen atoms [*cf.* O7 and O8 in Fig. 3a] of the C=O groups. For catechol, the two hydrogen atoms [*cf.* H1 and H2 in Fig. 3b] from the respective OH-groups were selected. The Hessian-based coupling estimate was chosen as a measure for the order of groups to be fused to the initial growing seeds. The

degeneracy threshold was set to 0.0001 a.u. For the catechol molecule, also FALCON coordinates representing the full space were generated. Here, the coordinates are local to three subsets (as depicted in Fig. 1a), representing the two OH groups and the benzene ring, respectively. For all generated FALCON coordinates, the input and output files are provided in the link given in the section ‘‘Data availability’’. For the uracil molecule, we also investigated reduced spaces spanned by normal coordinates. The coordinates were selected in a way that they capture a similar vibrational behaviour of the equivalent reduced space obtained through FALCON coordinates.

PESs and DMSs were generated with a locally modified version of the MidasCpp2022.04.0<sup>80</sup> chemistry program package employing the ADGA.<sup>26,67,81</sup> As fit-basis functions polynomial functions are used with a maximum order of eight. For the average vibrational density in the iterative ADGA procedure, a number of six modals per mode were chosen. The basis set number was set to ten B-Spline<sup>82</sup> functions with a density of 0.8. Default convergence criterion’s of the MidasCpp2022.04.0<sup>80</sup> version for the ADGA  $\varepsilon_{\text{rel}}$ ,  $\varepsilon_{\text{abs}}$  and  $\varepsilon_{\rho}$  were chosen. For the multilevel surfaces, the individual surface operator terms were calculated for each step in the protocol separately and merged afterwards by an in-house python3 library following eqn (5). All single-point (SP) calculations were performed employing ORCA5.0.3.<sup>83–85</sup> Two different levels of electronic structure theory were applied. The high level of theory in this work is represented by the B2PLYP-D3BJ/aug-cc-pVTZ level of theory. This choice was inspired as the double hybrid functional B2PLYP<sup>75</sup> in conjunction with triple- $\zeta$  basis set sizes has shown to yield harmonic frequencies more accurate than other density-functional theory (DFT) approaches while being less computationally demanding as, for instance, coupled cluster calculations.<sup>86,87</sup> For anharmonic vibrational structure calculations on proteins, the combination of B3LYP and MP2 has previously been found to be a pragmatic choice.<sup>88</sup> Hence, we consider the double hybrid functional B2PLYP that directly combines the named approaches promising for anharmonic calculations on biomolecular systems. The resolution of identity approximation for Coulomb integrals (RI-J)<sup>89</sup> and the COSX numerical integration for HF exchange (RI-COSX)<sup>90</sup> was used with suiting auxiliary basis sets (def2/J and aug-cc-pVTZ/C for B2PLYP).<sup>91</sup> For the low level of theory, we employed r<sup>2</sup>SCAN-3c<sup>92</sup> due to its recommended good-cost accuracy performance.<sup>93</sup> For all SP calculations, the SCF convergence threshold was set to tight. All calculated surfaces are listed in the ESI† in Table S1.

The anharmonic vibrational calculations were performed with MidasCpp2023.10.0<sup>94</sup> targeting all fundamental vibrational excitations, applying VCC damped linear response functions. For the underlying VSCF ground-state wave function calculations a B-Spline basis of 10th order using a primary basis density of 0.8 was chosen. Furthermore, the six lowest VSCF modals per mode were included for the correlation treatment. For the damped linear response calculations, the band-Lanczos algorithm<sup>68</sup> was chosen. One technical parameter of the band-Lanczos algorithm is the so-called chain

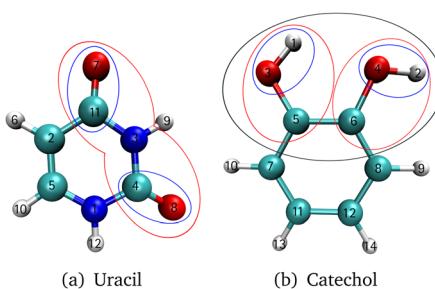


Fig. 3 First fusion groups for (a) pyrimidine-2,4(1H,3H)-dione (uracil) starting from the two oxygen atoms of the carbonyl group (O7 and O8) with indication of vibrational subspaces spanned by two (blue) and nine (red) FALCON coordinates, and (b) 1,2-dihydroxybenzen (catechol) starting the FALCON growing algorithm from the two hydrogen atoms of the hydroxygroups (H1 and H2) with indication of vibrational subspaces spanned by two (blue), six (red) and 12 (black) FALCON coordinates. The numbers on the atoms refers to the initial fusion group, respectively.



length, which describes the excitation space of the correlated vibrational wave function. Since the full chain length is often not computationally feasible, the user has to ensure a convergence of the chain length. For each VCC IR spectrum in the result section, the convergence of the chain length is ensured (cf. Fig. S2 and S3 in the ESI<sup>†</sup>). Different excitation levels for the VCC wave function were chosen, ranging from VCC[1] to VCC[4]. The damped linear response functions were calculated at every inverse centimeter for the spectral range of 1600 to 1900  $\text{cm}^{-1}$  and 3500 to 3900  $\text{cm}^{-1}$  for uracil and catechol, respectively. In order to compare the theoretical spectra with the experimental ones, the damping factors for uracil and catechol calculations were set to  $\gamma = 2 \text{ cm}^{-1}$  and  $\gamma = 13 \text{ cm}^{-1}$ , respectively.

All spectra shown in this study are normalized to the maximum peak height of the respective frequency range investigated. The experimental spectrum of uracil<sup>56</sup> was extracted using a plot digitizer.<sup>95</sup> Presented modes are visualized with the JSindo<sup>96</sup> software. Information on explicit runtimes refers to calculations performed on Intel(R) Xeon(R) silver 4316 CPU (2.30 GHz).

## 4 Results and discussion

### 4.1 C=O stretchings signatures in uracil

**4.1.1 Quasi-harmonic frequencies.** We first investigate the convergence behaviour of quasi-harmonic C=O stretching vibrations resulting from reduced vibrational subspaces generated with the FALCON growing scheme. We start from the two oxygen atoms O7 and O8 (Fig. 3a). The fused atoms representing the respective vibrational subspaces are listed in the ESI<sup>†</sup> in Table SII. The resulting quasi-harmonic frequencies for the two C=O stretching vibrations are displayed in Fig. 4, with the respective harmonic frequencies indicated by dotted lines. Is the reduced space represented by two FALCON coordinates (blue circles in Fig. 3), the quasi-harmonic frequencies deviate strongly from the full harmonic frequency about 36.61  $\text{cm}^{-1}$

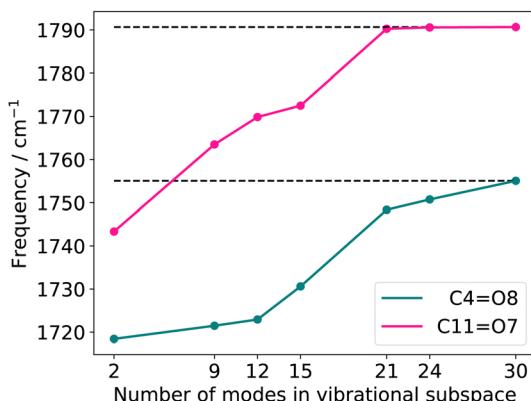


Fig. 4 Quasi-harmonic frequencies for the two C=O stretching vibrations with increasing vibrational subspace (B2PLYP-D3/aug-cc-pVTZ). The atoms considered in each vibrational subspace are shown in the ESI<sup>†</sup> in Table SII.

and 47.33  $\text{cm}^{-1}$  for the C4=O8 and C11=O7 stretching vibration, respectively. Increasing the subspace size, we obtain a converging behaviour towards the full harmonic vibration. However, the harmonic frequencies at normal modes is recovered by a vibrational subspace including 21 modes for the C=O frequency, which is higher in energy (C11=O7). For the quasi-harmonic frequency of C4=O8, full convergence is only reached when all vibrational modes are taken into account. In the harmonic case, tailored FALCON subspaces, hence, appear unsuitable for an adequate alternative representation of the vibrational motion of the two C=O stretching vibrations. Therefore, the full vibrational space (normal modes) is preferable for describing the vibrational motion of the C=O stretchings. Hence, we performed anharmonic reduced-space calculations for FALCON and normal coordinates to test our hypothesis.

**4.1.2 Anharmonic spectra of reduced vibrational spaces.** Fig. 5 displays VCC damped linear response spectra for 2MC,

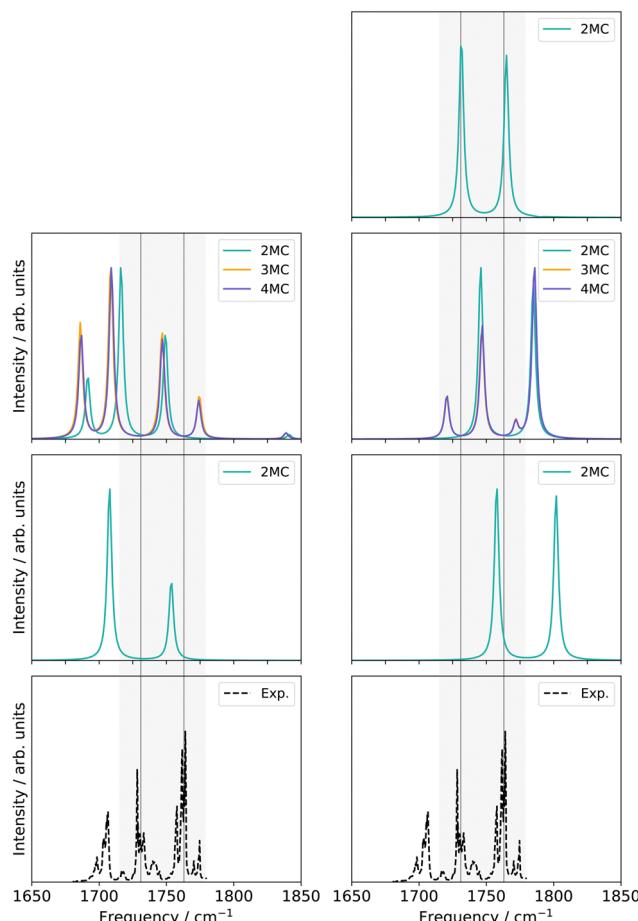


Fig. 5 Damped linear response spectra of uracil for the spectral range of 1650–1850  $\text{cm}^{-1}$  on two-mode reduced spaces on a full coupled PES (FVCI), nine-mode reduced space on 2MC, 3MC, and 4MC property surfaces (VCC[4]), and on a 30-mode full space (VCC[2]) for FALCON and normal modes, respectively. The experiment is obtained in an Ar-matrix at  $T = 10 \text{ K}$  and taken from ref. 56. The property surfaces are calculated on the  $r^2\text{SCAN-3c}$  level. All peak position values for the calculations can be found in the ESI<sup>†</sup> in Table SIII.

3MC, and 4MC property surfaces in both the two-mode and nine-mode (reduced) spaces, as well as the 30-mode (full) space, using normal and FALCON coordinates, where applicable. The normal modes of the respective reduced spaces were selected as they capture a similar vibrational behaviour of the equivalent reduced FALCON space. A visualization of the normal modes is found in the ESI,† Fig. S1. The respective normal modes of the two-mode space consist of the two C=O stretching vibrations (Q24 and Q25 in Fig. S1, ESI†) and the nine-mode space is extended with the modes Q0, Q5, Q9, Q11, Q13, Q18, and Q20 (Fig. S1a, f, j, l, n, s and u, ESI†). All SP calculations were carried out with the low-level method  $r^2$ SCAN-3c. Further, evaluations of different VCC excitation levels applied on the nine normal mode space (*cf.* ESI,† Section S2.3.1) suggest, that the coupling order of the potential exerts a higher influence on the calculated infrared spectrum than higher excitations (beyond VCC[2]) in the correlated wave function. Therefore the performance on the 30-mode space and subsequent multilevel calculations are assessed with a VCC[2] level of excitations in the wave function. For comparison, the experimental spectrum in Ar-matrix at 10 K is shown on the bottom panel of Fig. 5. The experimental spectrum exhibits three significant signals within the frequency range of 1600 to 1900  $\text{cm}^{-1}$ . According to ref. 56, the first signal around 1703.5–1706.4  $\text{cm}^{-1}$  refers to a combination band, whereas the peak at 1728.2–1733.2  $\text{cm}^{-1}$  and 1761.4–1763.7  $\text{cm}^{-1}$  can be assigned to the fundamental C=O stretching vibrations, with peak centres at 1731  $\text{cm}^{-1}$  and 1763  $\text{cm}^{-1}$ .

For the two-mode space in FALCON coordinates on a fully coupled potential (Fig. 5 left panel), the calculated peak positions and intensities (FVCI) significantly diverge from the experimental results: both peaks are red-shifted with a contrary calculated intensity compared to the experiment. Upon expanding the reduced space to nine FALCON modes, the resulting VCC[4] spectrum on 2MC shows a red shift compared to the experimental frequencies, accompanied by the emergence of a third signal around 1692  $\text{cm}^{-1}$ . However, the calculated peak heights remain inaccurate in comparison to the experiment. Increasing the coupling level does not improve the agreement with the experiment. Examining the convergence behaviour with growing subspace sizes from the two-mode space to the nine-mode space, we observe a blue shift for the C4=O8 peak, while a red shift is observed for the C11=O7 fundamental anharmonic peak. Concluding from these results, such FALCON-generated reduced spaces do not appear suitable for appropriately describing the C=O stretches in uracil.

In the case of the two-mode space spanned by two normal modes (Fig. 5 right panel), the calculated excitation energies obtained from a FVCI calculation show discrepancies of 27  $\text{cm}^{-1}$  and 39  $\text{cm}^{-1}$  when compared to the experimental values for C4=O8 and C11=O7, respectively. Increasing the vibrational space to nine modes (Fig. 5 right panel), the two peaks of the two-mode space are shifted towards lower frequencies and thereby in closer agreement to the experimental peaks. However, the intensities are still mismatched to the experiment. Increasing the coupling order of the potential to 3MC

and 4MC (orange and purple spectrum in Fig. 5), the spectra on these potentials with different coupling orders exhibit identical shapes. Further, we observe a good agreement to the intensity relation, shown in the experiment: a third peak around 1720  $\text{cm}^{-1}$  with a similar height to the experimental peak around 1705  $\text{cm}^{-1}$  appears. This peak refers to a resonance between a combination band and the fundamental C4=O8 stretching vibration. Increasing the vibrational space to the full vibrational space (30 normal modes), the excitation energies of the two fundamental C=O stretches agree with the experiment for a 2MC potential. Comparing the peak positions for 2MC coupled potentials of the two-mode, nine-mode, and 30-mode space, we obtain a convergence towards the experiment, but convergence is only reached for the full vibrational space (30 normal modes). We can, hence, conclude, that normal coordinates, in particular the space of all normal coordinates, are a suitable choice for the initial surface, as already indicated by the convergence behaviour of the quasi-harmonic frequencies (Fig. 4). Further, these calculations showed, that higher coupling terms ( $>2\text{MC}$ ) have a high impact on the intensities without impacting the peak positions. Further, the inclusion of three-body terms reveals the appearance of the combination band peak. This can be explained by the fact, that the resonance between the combination band and the fundamental band is itself an interaction of three different modes that necessitates a proper description through a 3MC-coupled PES. Unfortunately, the generation of 3MC coupling terms for 30-modes and the subsequent wave function calculation is already a computational challenge and tailored multilevel property surfaces for these C=O stretchings appear as a promising route.

**4.1.3 Multilevel spectra.** Multilevel setups and their nomenclature are summarized in Table 1. As can be seen in Fig. 5, the 30-mode 2MC potential ( $r^2$ SCAN-3c) proves to be a promising initial low-level potential for a multilevel approach, as it results in the correct peak position in the spectrum (*cf.* Fig. 5 upper right). We will refer from now on to this surface as LL1. The anharmonic spectrum of a 3MC (or higher) coupled potential of the nine-mode reduced space in normal coordinates exhibits qualitatively correct agreement of intensities, and it reveals the third peak analogously to the experimental data (*cf.* Fig. 5). Therefore, we added all 4MC coupling terms of these nine modes as selected higher coupling terms (step 2 in the multilevel surface creation procedure in our protocol) to the initial low-level property surface. The resulting VCC[2]

**Table 1** Nomenclature of multilevel property surfaces generated for the description of the C=O stretching vibrations in uracil.  $M$  refers to the number of modes considered within one electronic structure level. All modes refer to normal coordinates

	High level	Low level
LL1		30 $M$ 2MC
LL2		9 $M$ 4MC; LL1
ML1	2 $M$ 2MC	LL2
ML2	6 $M$ 2MC	ML1



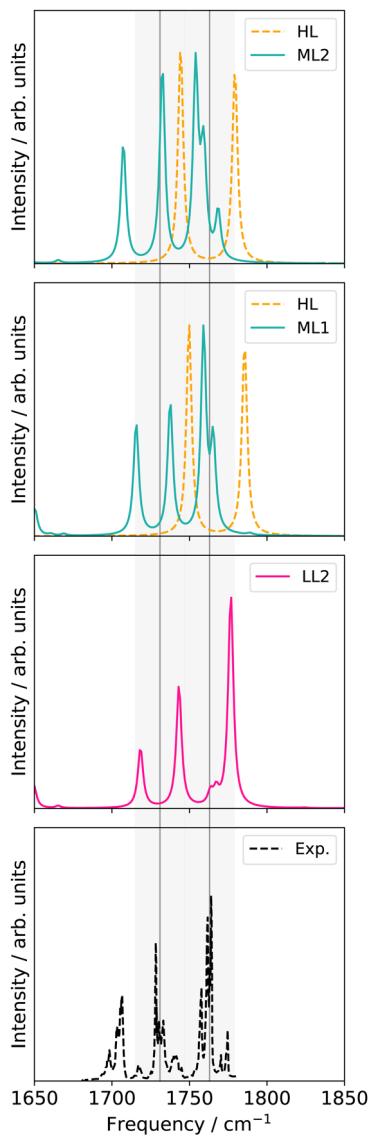


Fig. 6 VCC[2] damped linear response spectra of different levels of the employed property surfaces on 30 normal modes of uracil. For further explanations see main text and Table 1.

spectrum is displayed in Fig. 6 (LL2). The excitation energies are shifted towards higher frequencies compared to the experiment by  $12\text{ cm}^{-1}$  and  $14\text{ cm}^{-1}$  for the  $\text{C}4=\text{O}8$  and  $\text{C}11=\text{O}7$  in comparison to LL1, respectively. Nonetheless, the spectrum correctly represents the intensity ratio, and a third peak around  $1718\text{ cm}^{-1}$  is captured. The resulting spectrum of the merged potential (LL2) lies, hence, between the two single-level spectra [*cf.* LL1 potential and nine-mode 4MC potential (*cf.* Fig. 5 right panel)].

Replacing the most important operator terms through operator terms that are obtained with a more expensive electronic-structure method, appears as a logical next step for an appropriate description of the potential. The two most important modes for the spectral signatures of the  $\text{C}=\text{O}$  spectral range are the normal coordinates of the  $\text{C}=\text{O}$

stretching vibrations. Therefore, we generated a two-mode 2MC coupled high-level potential, where the SPs were calculated on the B2PLYP-D3/aug-cc-pVTZ level and combined it with non-existing terms to the LL2 potential. The merged potential is denoted ML1. The resulting VCC[2] spectrum is depicted in Fig. 6. In comparison to the calculated spectrum on the LL2 potential, we observe a red shift of the ML1. We further observe a good agreement with the experimental data, assuming some uncertainties also in the experiment. Following our protocol, the response functions of the latest low-level potential are analyzed to identify which modes are most significant for the  $\text{C}=\text{O}$  stretching signals and necessitate a high-level description. Both peaks exhibit the strongest mixing with the modes shown in Fig. 7 besides the two  $\text{C}=\text{O}$  stretching vibrations (Q24 and Q25 in Fig. 7). The contribution of these modes to the peaks of the response functions at  $1743\text{ cm}^{-1}$  and  $1777\text{ cm}^{-1}$  of the LL2 are displayed in Table 2. For these additional four modes (Q5, Q11, Q13, Q18), we generated a 2MC potential in B2PLYP-D3/aug-cc-pVTZ level of theory (high level) and replaced the respective low-level operator terms in the ML1 potential. The resulting VCC[2] spectrum on this new multilevel potential is labeled with ML2 in Fig. 6. Compared to the ML1 spectrum the intensity of the  $\text{C}4=\text{O}8$  stretching peak increased, and all peaks are slightly red-shifted. Further, a third peak next to the peak referring to  $\text{C}4=\text{O}8$  appears, which is also shown by the experimental spectrum. This third signal around  $1707\text{ cm}^{-1}$  in the ML2 spectrum originates from a resonance of a combination band from mode Q11 and Q13 (*cf.* Fig. 7) with the fundamental  $\text{C}=\text{O}$  stretch Q24. Q11 and Q13 are a symmetric and asymmetric stretch of the uracil ring. It becomes evident, that an accurate description of selected low-order coupling terms significantly improves the spectral shape. Our protocol offers guidance on effectively selecting these terms tailored to the spectroscopic signatures while managing computational costs.

To estimate the computational costs, we examine the number of SP calculations required for the generation of the ML2 PES: for the LL2 PES, which forms the low-level component of the ML2 PES, 211 914 SP calculations were required, while the

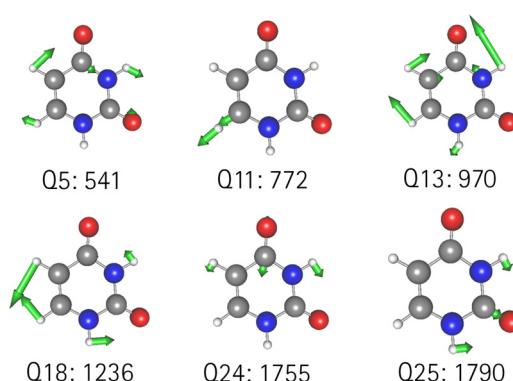


Fig. 7 Visualization of the normal modes described in the high level for uracil. The frequency refers to the harmonic frequency in  $\text{cm}^{-1}$  obtained with B2PLYP-D3/aug-cc-pVTZ.

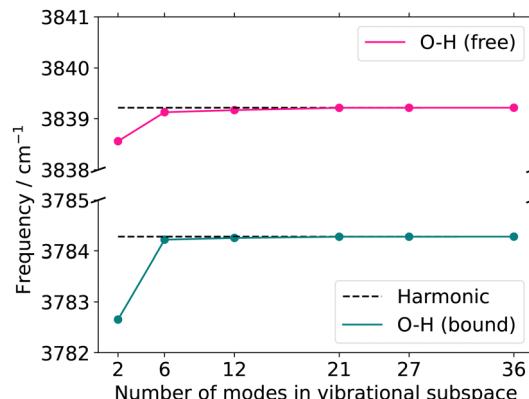
**Table 2** Coefficients to the fundamental vibrational excitations of the two  $\text{C}=\text{O}$  stretching vibrations obtained in the VCC[2] response framework on the LL2 PES. For further information see main text

1743 $\text{cm}^{-1}$		1777 $\text{cm}^{-1}$	
Coefficient	Mode(s)	Coefficient	Mode(s)
0.403	$\text{Q24}_1$	0.780	$\text{Q25}_1$
0.231	$\text{Q11}_1 \text{Q13}_1$	0.033	$\text{Q11}_1 \text{Q13}_1$
0.120	$\text{Q5}_1 \text{Q18}_1$	0.019	$\text{Q5}_1 \text{Q18}_1$

high-level component required 2736 SPs. On average, a single SP calculation for the uracil molecule takes 35 s of CPU time at the low level ( $r^2\text{SCAN-3c}$ ) and 373 s of CPU time at the high level (B2PLYP-D3/aug-cc-pVTZ) of theory. Based on this, the estimated total CPU time for these calculations amounts to approximately 100 days. Since these SP calculations can be run in parallel, this results in a total wall time of approximately 2.5 days using 40 cores. In comparison, generating a PES for the full vibrational space of uracil, including 4MC coupling terms, requires an estimated number of 2.7 million SP calculations. A detailed explanation of how this number was determined is available in Section S4 of the ESI.† For the low-level calculations, this corresponds to approximately 1000 days of CPU time, leading to a wall time of around 25 days with 40 cores. For a generation of the PES at the high level only, the immense amount of computing time is even more pronounced, requiring approximately 250 days of wall time with 40 cores. This is 10 times and 100 times the amount of walltime compared to the walltime required by the protocol for a calculation in the low or high level, respectively. This thought experiment highlights that for a high-level PES generation of the full vibrational space of uracil is prohibitively resource-intensive and thereby underlines the efficiency of the proposed protocol.

## 4.2 O–H stretching signatures of catechol

**4.2.1 Quasi-harmonic frequencies.** The quasi-harmonic frequencies of the two OH-stretching vibrations of catechol for growing vibrational subspace sizes are presented in Fig. 8 in comparison to the corresponding harmonic frequencies. The FALCON growing scheme starts from the two hydrogen atoms, where the created subspace sizes consisting of two, six, and twelve FALCON modes are highlighted with blue, red, and black in Fig. 3b, respectively. The remaining fused groups are listed in the ESI† in Table SIV. We obtain a fast convergence of the quasi-harmonic frequencies for growing sizes of tailored subspaces: the subspace that is spanned by two FALCON coordinates (blue circles in Fig. 3b) exhibits quasi-harmonic frequencies deviating marginally by  $1.63 \text{ cm}^{-1}$  and  $0.66 \text{ cm}^{-1}$  from the harmonic frequency for the bound ( $\text{O}_3\text{--H}_1$ ) and the free OH stretch ( $\text{O}_4\text{--H}_2$ ), respectively. With an increased subspace size to six FALCON modes (red circles in Fig. 3b), the deviation of the two quasi-harmonic frequencies is decreased to  $0.06 \text{ cm}^{-1}$  and  $0.09 \text{ cm}^{-1}$  for the bound and free OH stretch, respectively. This rapid convergence of the quasi-harmonic frequencies indicates a local nature of the OH-stretching vibrations, unlike the  $\text{C}=\text{O}$  stretchings of uracil. Hence, these



**Fig. 8** Quasi-harmonic frequencies for the two OH-stretching vibrations with increasing vibrational subspace size (B2PLYP-D3/aug-cc-pVTZ). O–H (bound) refers to the hydrogen-bonded stretching vibration of  $\text{O}_3\text{--H}_1$ , whereas O–H (free) refers to the hydrogen-bond-accepting stretching vibration of  $\text{O}_4\text{--H}_2$ . The atoms considered in each vibrational subspace are listed in the ESI† in the Table SIV. The dotted line represents the harmonic frequency at normal modes.

growing-scheme-generated subspaces are promising for the selection of suitable vibrational coordinates used in subsequent anharmonic structure calculations.

**4.2.2 Anharmonic spectra of reduced vibrational spaces.** As the vibrational subspace defined by six FALCON coordinates demonstrates nearly converged results for the two OH stretchings in the harmonic case, our investigation focuses on anharmonic calculations on vibrational subspaces spanned by two-, six-, and twelve-FALCON modes. Visualizations of the underlying modes for the six-mode space are shown in Table 3 (upper row) and for the twelve-mode space in the ESI† in Fig. S6. For each subspace, we performed VCC-damped linear response calculations, employing 2MC, 3MC, and 4MC coupling levels, if applicable. All SP calculations for the PES were carried out with the low-level method  $r^2\text{SCAN-3c}$ . The performance of the anharmonic VCC spectra is assessed with a VCC[4] wave function for the six-mode space. In contrast, the twelve-mode space is assessed with VCC[3] wave function calculations, due to computational feasibility. Performance checks (*cf.* ESI† in Fig. S9) for VCC[1] to VCC[4] calculations on the six-mode space suggest that triple excitations (VCC[3]) exhibit nearly identical spectral shape to the VCC[4] reference spectrum, and can thus be considered converged. The resulting anharmonic spectra are shown in Fig. 9 and are discussed in detail in the following.

The FVCI spectrum on a 2MC potential of the two-mode reduced space (turquoise in Fig. 9) exhibits peaks at  $3618 \text{ cm}^{-1}$  and  $3677 \text{ cm}^{-1}$  for the bound and free OH stretch, respectively. In the FALCON six-mode space, a spectrum with a 2MC potential exhibits peak positions at  $3593$  and  $3655 \text{ cm}^{-1}$  for the bound and free OH stretch, respectively. These positions are red-shifted by  $25 \text{ cm}^{-1}$  and  $16 \text{ cm}^{-1}$  with respect to the two-mode space spectrum. The spectral shapes on the 3MC (orange) and 4MC (purple) surfaces yield identical results and are blue-shifted compared to the 2MC potential spectrum. Regarding the calculated peaks of the free OH-stretching vibration, a

**Table 3** Coupling levels and visualization of level-specific modes for the high (HL) and low level (LL) of multilevel surfaces applied for the reduced vibrational space (6 FC), and the full vibrational space in either FALCON (36 FC) or normal coordinates (36 NC). The quasi-harmonic and harmonic frequencies of the respective modes are given in parenthesis in  $\text{cm}^{-1}$ . A visualization of the remaining modes can be found in Fig. S12 and S13 (ESI)

Space	Level	MC	Modes of tailored spaces
6 FC	HL	2MC	
	LL	4MC	Analogous to HL
	HL	2MC	
	LL	4MC	
	3MC (group)	2MC	2MC terms of the full set of modes and 3MC terms limited to those incorporating Q40, Q41 or both
	HL	2MC	
36 NC	HL	2MC	Q25, Q27, Q40, Q41 from HL +
	LL	4MC	
	3MC (group)	2MC	2MC terms of the full set of modes and 3MC terms limited to those incorporating Q40, Q41 or both
	LL	4MC	

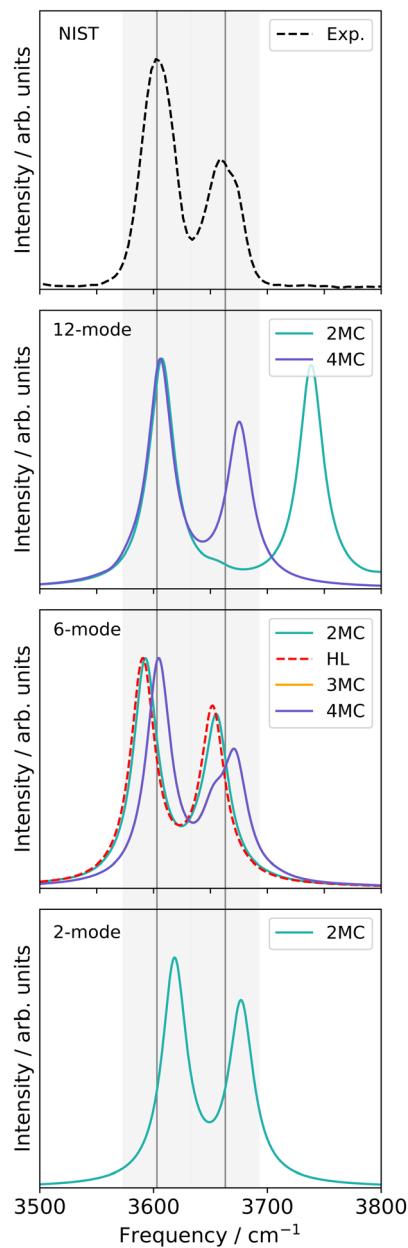
resonance peak with  $3652$  and  $3672\text{ cm}^{-1}$  is observed, as indicated by a left-sided shoulder of the peak.

For the twelve-mode space with a 2MC potential, the fundamental OH bound peak is located at  $3607\text{ cm}^{-1}$ . This is in agreement with the corresponding peak in the spectrum calculated in the six-mode space with 3MC/4MC coupling terms. In contrast, the peak associated with the free OH stretch (twelve-mode 2MC) shows a significant deviation from the spectra on the two- and six-mode spaces, appearing at  $3738\text{ cm}^{-1}$ . Additionally, the two peaks in this spectrum are of equal height, unlike in the spectral shapes of the two- and six-mode space. The spectrum calculation with a potential including 3MC terms resulted in profound convergence issues (*cf.* Fig. S7h, ESI†) and is, hence, not shown here. Introducing 4MC terms (purple) in the potential of the twelve-mode space, the spectral shape qualitatively aligns with the spectrum of the two- and six-mode space again. The peak position of the free OH stretch

appears at  $3675\text{ cm}^{-1}$  aligning with the second resonance peak of the 3MC/4MC six-mode space spectra.

Additionally, an experimental gas-phase spectrum<sup>97</sup> is shown on top of Fig. 9. It exhibits two peaks at  $3603\text{ cm}^{-1}$  and  $3663\text{ cm}^{-1}$ . The overall shape of the experimental IR spectrum is remarkably well reassembled by the theoretical infrared spectrum on the two-mode space, given the applied approximation in the PES. The peak positions still deviate from the experiment about  $15\text{ cm}^{-1}$  and  $14\text{ cm}^{-1}$  for the bound and free OH stretch, respectively. For the six-mode space, the spectral shapes of the 3MC and 4MC surfaces demonstrate semi-quantitative agreement with the experimental spectrum. The peak position assigned to the bound OH stretch deviates only marginally by  $2\text{ cm}^{-1}$  and  $1\text{ cm}^{-1}$  for the 3MC and 4MC PES, respectively. The experiment also indicates a slight shoulder, however, reversed to the shoulder in the six-mode space. Still, the local description of the OH stretching appears





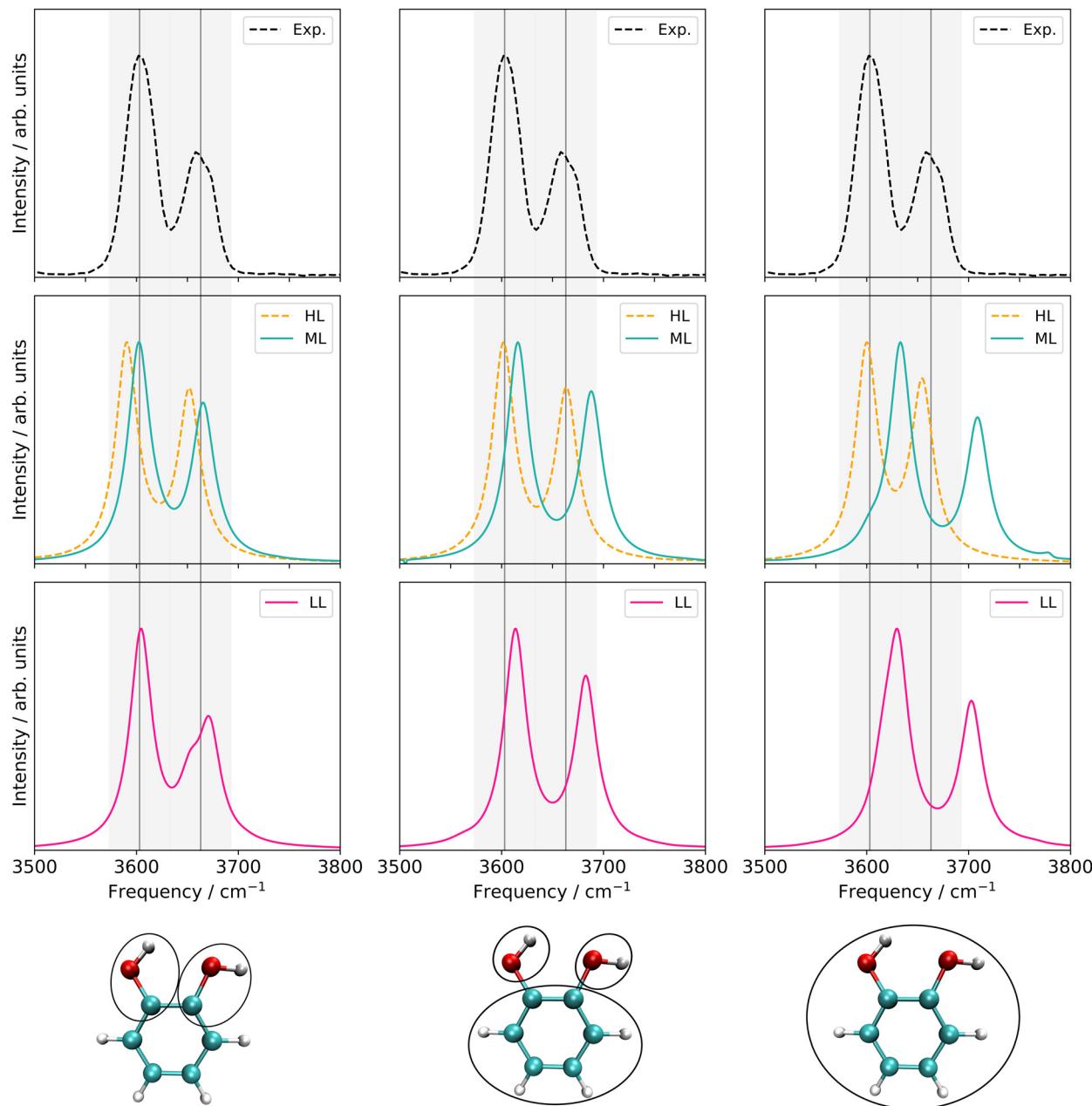
**Fig. 9** FVCI, VCC[4] and VCC[3] damped linear response spectra of catechol for the spectral range of  $3500\text{--}3800\text{ cm}^{-1}$  on the two-mode (FALCON) reduced-space on full-coupled property surfaces (FVCI), on the six-mode (FALCON) reduced-space on 2MC, 3MC and 4MC property surfaces, and on the twelve-mode (FALCON) reduced space on 2MC, and 4MC property surfaces in comparison to the experiment in gas phase.<sup>97</sup> The property surfaces are calculated on the  $r^2\text{SCAN-3c}$  level, if not stated otherwise in the legend (HL). All peak position values for the calculations can be found in Table SV in the ESI.<sup>†</sup>

to be remarkably suitable, given the applied level of electronic structure theory. To study the possibility of a fortunate error cancellation arising from the underlying approximation, we calculated the 2MC PES at a higher level using B2PLYP-D3/aug-cc-pVTZ (red in Fig. 9). Notably, we observe a close alignment between the HL and LL spectra, indicating the good performance of the  $r^2\text{SCAN-3c}$  method.

For the twelve-mode space, the inclusion of 4MC terms recovers the numerical stability. This could indicate, that the poor descriptions of the 2MC and 3MC spectra in the twelve-mode space simply can suffer from a non-converged n-mode expanded PES. Still, the twelve-mode space spectrum with 4MC terms performs worse than the spectra of the six-mode space in comparison to experimental data. In a thorough analysis of the 2MC PES, we found two 2MC potential cuts showing conspicuous sharp declines at already small displacements of the corresponding twelve-mode space along the OH-stretch coordinates in conjunction with low-frequency out-of-plane wagging modes (*cf.* Fig. S10, ESI<sup>†</sup>). The exclusion of these coupling terms stabilized the response calculations and further yielded frequencies that deviated only one inverse centimeter from both experimental peaks. However, to obtain this minor deviation, the inclusion of the 4MC terms was still necessary. We, hence, believe that these mode-coupling terms between the OH-stretching mode and the out-of-plane modes cause the numerical instabilities in the VCC calculations. This assumption is underlined by earlier work on malonaldehyde.<sup>98,99</sup> Here, a similar situation with strong anharmonicity coupling between the in-plane and out-of-plane vibrational motion of the OH groups has been found. These studies also employed rectilinear (normal) coordinates to describe the vibrational motion. Both studies offer the indication that the application of rectilinear stiff coordinates and neglecting rotational and vibrational coupling effects challenge the accurate description of the torsion angular motion of the OH-groups in such systems. Fortunately, such floppy-out-of-plane motions are probably not present in more complex biomolecular systems, as the rotation of the OH-rotation is likely hindered due to solvent molecules or the surrounding environment of the remaining biomolecule.

From the analysis above, we conclude that the OH-stretching modes are better characterized by a rather local nature in the two-mode or six-mode space as (1) their locality character stabilizes the calculation of OH-stretching infrared signals, and (2) they provide very good cost-accuracy ratios.

**4.2.3 Multilevel spectra.** As the two-mode space shows larger deviations from the experiment, we believe that the reduced six-mode space emerges as the most appropriate choice for the generation of multilevel potential. Additionally, we explore two alternative descriptions of the vibrational space. For that, we chose the full vibrational space by employing either FALCON or normal coordinates. The FALCON coordinates applied here keep the vibrations within two subsets corresponding to the individual OH groups and the benzene ring local, totaling in 36 FALCON modes including relative vibrational motion between the groups. All three fundamental space descriptions employed in this work are illustrated in the lower panel of Fig. 10. Fig. 10 further shows the VCC spectra obtained from the LL, HL, and ML property surfaces of each vibrational space. An overview of all created ML potentials of the three different vibrational spaces is given in Table 3. In the case of the vibrational reduced space spanned by six FALCON modes



**Fig. 10** Left panel: VCC[4] damped linear response spectra on different levels of the employed property surfaces of the six-mode FALCON space of catechol. Middle panel: VCC[2] damped linear response spectra of different levels of the employed property surfaces on 36 FALCON modes of catechol. Right panel: VCC[2] damped linear response spectra of different levels of the employed property surfaces on 36 normal modes of catechol. Information about the applied electronic structure method and coupling orders can be found in Table 3. Further, all peak position values can be found in Table SV in the ESI.†

(6 FC), the ML PES includes 4MC terms for all six modes described by the lower level of theory and 2MC terms described by the higher level of theory. The selection process for the ML PES of the 36 FC and 36 NC spaces is summarized in the following:

**LL PES:** the LL PES consists of 2MC terms for the full set of modes, 3MC terms limited to those incorporating at least one of the two OH stretches (Q40/Q41), and 4MC terms for a selected set of modes. The selection of these modes is based

on their coupling strength to the OH stretches, estimated on the 2MC LL potential. Due to strong coupling values shown in the space spanned by normal coordinates (*cf.* Table SVI in the ESI†), a rather large set of ten modes (including the two OH stretches) was chosen. The coupling strength estimates of the FALCON space are by two orders of magnitude smaller than those for normal coordinates (*cf.* Table SVII in the ESI†). Thus, for the FALCON coordinates only four modes plus the two OH-stretching modes are selected for 4MC terms in the LL of the 36



FC space. Noticeably, all four modes describe inter-group motions of the two OH-groups.

**HL PES:** the 2MC terms described in the HL were selected by evaluating the coefficients of the response functions referring to the two fundamental OH-stretching peaks. Interestingly, we observed for both spaces relatively small coefficients for mixing vibrational states (*cf.* Table SVIII in the ESI<sup>†</sup>). However, the 2MC potential cuts of the LL PES showed instabilities between the out-of-plane wagging and the OH-stretching vibration of the respective OH group (*cf.* Fig. S14 and S15 in the ESI<sup>†</sup>). For the normal coordinates, these are the 2MC terms Q7/Q41 and Q10/Q40, and for the FALCON space Q6/Q41 and Q11/Q41. Based on this observation, we included only the modes describing the bending behaviour of the OH groups in conjunction with the OH-stretching modes for the HL. For the FALCON space, these are the bendings Q28 and Q30, whereas for the normal mode space we chose Q24, Q25, Q27, and Q28.

Let us now discuss the performance of the VCC spectra of the LL, HL and ML potential originating from the three different spaces displayed in Fig. 10. For computational efficiency, all response calculations on the full vibrational space were conducted at the VCC[2] level. Analysis of VCC excitation levels in response functions for catechol on the 2MC, 3MC, and 4MC of the FALCON six-mode space (*cf.* Fig. S9 in the ESI<sup>†</sup>) reveals that VCC[2] spectra exhibit identical peak positions to VCC[4] spectra, but tend to overestimate the intensity of the free OH peak. The left panel of Fig. 10 shows the results for the reduced space with six FALCON modes as depicted below the graph in this panel. The VCC[4] spectra on LL and HL potential are already extensively discussed above but for better comparison displayed again. On the resulting multilevel spectrum (turquoise), we observe, that the shoulder of the free OH-peak of the LL spectrum disappears for the ML spectrum, resulting in one fundamental peak with semi-quantitative agreement to the experiment. The LL spectrum of the 36 FC space is shown in pink in Fig. 10 middle bottom. It exhibits a similar shape as the experimental data, but the OH-stretching peaks are blue-shifted by 10 cm<sup>-1</sup> and 20 cm<sup>-1</sup>. The ML spectrum almost aligns with the LL spectrum, while the HL spectrum alone (yellow dashed lines) shows surprisingly good agreement with the experiment. While the resulting LL and ML spectra with the 36 NC space exhibit similar shapes to the spectra on 36 FC coordinates, the calculations showed numerical instabilities in chain length convergence (*cf.* Fig. S7h in the ESI<sup>†</sup>). The spectrum on the HL potential has red-shifted peaks towards the experiment with frequency values of 3600 cm<sup>-1</sup> and 3655 cm<sup>-1</sup>.

Comparing all multilevel spectra, the spectra calculations on the normal coordinates are the only ones that mitigate numerical instability issues. They further show the largest deviations from the experiment. The reduced-space (6 FC) spectrum from the growing scheme matches the experimental spectrum best. Eye-catching is also the fact, that the HL spectrum of the 36 FC (*i.e.* four modes on a 2MC PES with B2PLYP-D3/aug-cc-pVTZ) alone achieves excellent agreement with the experiment, in contrast to the respective ML spectrum of 36 FC which deviates slightly more from the experiment. The HL spectrum alone only

includes two inter-group coordinates describing the OH-groups' relative bending motions and two local OH-stretching modes, while the ML PES considers all 36 FALCON coordinates. A larger vibrational space enhances the likelihood of encountering instabilities in the PES, as shown by some 2MC potential cuts. Hence, these pitfall mode combinations can cause these discrepancies.<sup>25</sup>

We can conclude, that finding a suitable description in rectilinear coordinates for the OH stretchings was not as straightforward as for the C=O stretches in the uracil case. Still, our protocol provides highly valuable guidance: we can summarize, that the OH-stretching vibrations are most likely better described by modes localized to a small subset – namely the OH groups themselves – as predicted by the quasi-harmonic frequencies of the growing scheme. This growing scheme also provided suitable FALCON-reduced spaces for the OH stretchings.

## 5 Conclusions

In this work, we proposed a computational protocol for tailoring efficient and at the same time highly accurate calculations of certain spectroscopic signatures in IR-spectroscopy. The protocol consists of two key steps: the selection of suitable coordinates and the generation of the multilevel property surfaces. We evaluated this protocol to two types of spectroscopic signatures: the spectral region of C=O fundamental stretchings in uracil and the OH stretchings in catechol.

For the uracil molecule, the protocol successfully yielded a calculated spectrum that closely matches experimental data: the multilevel property surface was generated as a function of normal modes. Reduced spaces in FALCON coordinates appeared unsuitable for anharmonic calculations, which has been correctly predicted by the convergence behaviour of the quasi-harmonic frequencies with increasing vibrational sizes. Moreover, the inclusion of selected higher coupling terms (in the low level) was necessary to reveal Fermi-resonance signals of a combination band and a fundamental C=O stretch.

For the calculation of the OH-stretching spectral signatures of catechol, the protocol suggested a local character of the two OH stretchings from the evaluation of quasi-harmonic frequencies. An underlying reduced-space provided by the FALCON growing scheme algorithm has proven to be well-suited and circumvented instability issues, which occurred in full-space representations. Moreover, we found that including higher coupling terms was more important than high-level treatment of certain coupling terms, highlighting the good performance of the fast  $r^2$ SCAN-3c semiempirical method.

Overall, we established a protocol that provides reliable guidance while balancing computational cost and accuracy for tailored spectroscopic signatures. However, further practical experience is necessary to develop a robust black-box procedure applicable to larger systems, such as biomolecules. Attractive theories and implementations exist, such as fragment-based potential energy surface generation<sup>51,54</sup> or an embedding



potential for the vibrational wave function,<sup>100</sup> that can be integrated into our protocol to further push the size limitations of these vibrational structure calculations.

## Author contributions

J. H. writing – original draft, review, and editing, project administration, data curation, formal analysis, investigation, methodology, validation, visualization, software P. C. writing – review and editing, investigation, validation, formal analysis, software C. K. conceptualization, funding acquisition, validation, supervision, writing – review and editing.

## Data availability

Data for this article, including the optimized structures of uracil and catechol, all underlying calculated potential energy and dipole moment surfaces, and input and output files for the generation of the FALCON coordinates are available at <https://doi.org/10.5281/10.5281/zenodo.14002766>.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) through the Emmy Noether Young Group Leader Programme (project 397897341) and an individual research grant (project 523893296). The computations were carried out on the cluster system at the Leibniz University Hannover, Germany, which is funded by the Leibniz University Hannover, the Lower Saxony Ministry of Science and Culture (MWK), and the DFG.

## Notes and references

- 1 T. Kottke, V. A. Lórenz-Fonfría and J. Heberle, *J. Phys. Chem. A*, 2017, **121**, 335–350.
- 2 S. Ito, H. E. Kato, R. Taniguchi, T. Iwata, O. Nureki and H. Kandori, *J. Am. Chem. Soc.*, 2014, **136**, 3475–3482.
- 3 E. Freier, S. Wolf and K. Gerwert, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 11435–11439.
- 4 K. Muroda, K. Nakashima, M. Shibata, M. Demura and H. Kandori, *Biochemistry*, 2012, **51**, 4677–4684.
- 5 T. Domratcheva, E. Hartmann, I. Schlichting and T. Kottke, *Sci. Rep.*, 2016, **6**, 1–14.
- 6 T. Fornaro, D. Burini, M. Biczysko and V. Barone, *J. Phys. Chem. A*, 2015, **119**, 4224–4236.
- 7 B. Temelso and G. C. Shields, *J. Chem. Theory Comput.*, 2011, **7**, 2804–2817.
- 8 O. Christiansen, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6672–6687.
- 9 O. Christiansen, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2942–2953.
- 10 C. König, *Int. J. Quantum Chem.*, 2021, **121**, e26375.
- 11 P. T. Panek and C. R. Jacob, *J. Phys. Chem. Lett.*, 2016, **7**, 3084–3090.
- 12 B. Thomsen, T. Kawakami, I. Shigemoto, Y. Sugita and K. Yagi, *J. Phys. Chem. B*, 2017, **121**, 6050–6063.
- 13 K. Yagi, K. Yamada, C. Kobayashi and Y. Sugita, *J. Chem. Theory Comput.*, 2019, **15**, 1924–1938.
- 14 K. Yagi and Y. Sugita, *J. Chem. Theory Comput.*, 2021, **17**, 5007–5020.
- 15 S. Carter, S. J. Culik and J. M. Bowman, *J. Chem. Phys.*, 1997, **107**, 10458–10469.
- 16 E. L. Klinting, O. Christiansen and C. König, *J. Phys. Chem. A*, 2020, **124**, 2616–2627.
- 17 A. Molina, P. Smereka and P. M. Zimmerman, *J. Chem. Phys.*, 2016, **144**, 124111.
- 18 P. T. Panek and C. R. Jacob, *J. Chem. Phys.*, 2016, **144**, 164111.
- 19 X. Cheng and R. P. Steele, *J. Chem. Phys.*, 2014, **141**, 104105.
- 20 B. Thomsen, K. Yagi and O. Christiansen, *J. Chem. Phys.*, 2014, **140**, 154102.
- 21 K. Yagi, M. Keçeli and S. Hirata, *J. Chem. Phys.*, 2012, **137**, 204118.
- 22 J. M. Bowman, Y. Wang, H. Liu and J. S. Mancini, *J. Phys. Chem. Lett.*, 2015, **6**, 366–373.
- 23 C. R. Jacob and M. Reiher, *J. Chem. Phys.*, 2009, **130**, 084106.
- 24 C. König, M. B. Hansen, I. H. Godtfiebsen and O. Christiansen, *J. Chem. Phys.*, 2016, **144**, 074108.
- 25 A. A. Zhanserkeev, E. L. Yang and R. P. Steele, *J. Chem. Theory Comput.*, 2023, **19**, 5572–5585.
- 26 M. Sparta, M. B. Hansen, E. Matito, D. Toffoli and O. Christiansen, *J. Chem. Theory Comput.*, 2010, **6**, 3162–3175.
- 27 D. Toffoli, M. Sparta and O. Christiansen, *Chem. Phys. Lett.*, 2011, **510**, 36–41.
- 28 K. Yagi, S. Hirata and K. Hirao, *Theor. Chem. Acc.*, 2007, **118**, 681–691.
- 29 K. Yagi, S. Hirata and K. Hirao, *Phys. Chem. Chem. Phys.*, 2008, **10**, 1781–1788.
- 30 K. Yagi and H. Otaki, *J. Chem. Phys.*, 2014, **140**, 084113.
- 31 J. Liu, J. Yang, X. C. Zeng, S. S. Xantheas, K. Yagi and X. He, *Nat. Commun.*, 2021, **12**, 1–10.
- 32 G. Rauhut, *J. Chem. Phys.*, 2004, **121**, 9313–9322.
- 33 G. Rauhut and B. Hartke, *J. Chem. Phys.*, 2009, **131**, 014108.
- 34 B. Ziegler and G. Rauhut, *J. Chem. Phys.*, 2016, **144**, 114114.
- 35 B. Ziegler and G. Rauhut, *J. Chem. Theory Comput.*, 2019, **15**, 4187–4196.
- 36 S. Carter and N. C. Handy, *Chem. Phys. Lett.*, 2002, **352**, 1–7.
- 37 J. A. Tan and J. L. Kuo, *J. Chem. Theory Comput.*, 2018, **14**, 6405–6416.
- 38 J. A. Tan and J. L. Kuo, *J. Chem. Phys.*, 2021, **154**, 134302.
- 39 J. A. Tan and J.-L. Kuo, *Molecules*, 2022, **27**, 3198.
- 40 D. M. Benoit, *J. Chem. Phys.*, 2004, **120**, 562–573.



41 M. Bounouar and C. Scheurer, *Chem. Phys.*, 2006, **323**, 87–101.

42 A. D. Boese and J. M. Martin, *J. Phys. Chem. A*, 2004, **108**, 3085–3096.

43 K. Pflüger, M. Paulus, S. Jagiella, T. Burkert and G. Rauhut, *Theor. Chem. Acc.*, 2005, **114**, 327–332.

44 V. Rodriguez-Garcia, S. Hirata, K. Yagi, K. Hirao, T. Taketsugu, I. Schweigert and M. Tasumi, *J. Chem. Phys.*, 2007, **126**, 124303.

45 T. Hrenar, H. J. Werner and G. Rauhut, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3123–3125.

46 L. Pele and R. B. Gerber, *J. Chem. Phys.*, 2008, **128**, 165105.

47 T. K. Roy and R. B. Gerber, *Phys. Chem. Chem. Phys.*, 2013, **15**, 9468–9492.

48 T. K. Roy and R. B. Gerber, *J. Chem. Theory Comput.*, 2020, **16**, 7005–7016.

49 K. Yagi, K. Hirao, T. Taketsugu, M. W. Schmidt and M. S. Gordon, *J. Chem. Phys.*, 2004, **121**, 1383–1389.

50 M. Riera, C. Knight, E. F. Bull-Vulpe, X. Zhu, H. Agnew, D. G. Smith, A. C. Simmonett and F. Paesani, *J. Chem. Phys.*, 2023, **159**, 054802.

51 D. G. Artiukhin, E. L. Klinting, C. König and O. Christiansen, *J. Chem. Phys.*, 2020, **152**, 194105.

52 D. Madsen, O. Christiansen and C. König, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3445–3456.

53 C. König and O. Christiansen, *J. Chem. Phys.*, 2016, **145**, 064105.

54 R. J. Spencer, A. A. Zhanserkeev, E. L. Yang and R. P. Steele, *J. Am. Chem. Soc.*, 2024, **146**, 15376–15392.

55 X. Cheng, J. J. Talbot and R. P. Steele, *J. Chem. Phys.*, 2016, **145**, 124112.

56 A. Y. Ivanov, A. M. Plokhotnichenko, E. D. Radchenko, G. G. Sheina and Y. P. Blagoi, *J. Mol. Struct.: THEOCHEM*, 1995, **372**, 91–100.

57 P. Larkin, *Infrared and Raman spectroscopy: principles and spectral interpretation*, Elsevier, 2017.

58 S. Coussan, Y. Bouteiller, J. P. Perchard and W. Q. Zheng, *J. Phys. Chem. A*, 1998, **102**, 5789–5793.

59 S. G. Stepanian and L. Adamowicz, *Low Temp. Phys.*, 2020, **46**, 155–164.

60 J. Kongsted and O. Christiansen, *J. Chem. Phys.*, 2006, **125**, 124108.

61 D. Toffoli, J. Kongsted and O. Christiansen, *J. Chem. Phys.*, 2007, **127**, 204106.

62 G. C. Schatz, *Rev. Mod. Phys.*, 1989, **61**, 669.

63 H. Rabitz and F. Alis, *J. Math. Chem.*, 1999, **25**, 197–233.

64 J. O. Jung and R. B. Gerber, *J. Chem. Phys.*, 1996, **105**, 10332–10348.

65 K. Yagi, T. Taketsugu, K. Hirao and M. S. Gordon, *J. Chem. Phys.*, 2000, **113**, 1005–1017.

66 M. Sparta, I. M. Høyvik, D. Toffoli and O. Christiansen, *J. Phys. Chem. A*, 2009, **113**, 8712–8723.

67 E. Lund Klinting, B. Thomsen, I. Heide Godtliebsen and O. Christiansen, *J. Chem. Phys.*, 2018, **148**, 64113.

68 I. H. Godtliebsen and O. Christiansen, *Phys. Chem. Chem. Phys.*, 2013, **15**, 10035–10048.

69 I. H. Godtliebsen and O. Christiansen, *J. Chem. Phys.*, 2015, **143**, 134108.

70 B. Thomsen, M. B. Hansen, P. Seidler and O. Christiansen, *J. Chem. Phys.*, 2012, **136**, 124101.

71 P. Seidler, M. Sparta and O. Christiansen, *J. Chem. Phys.*, 2011, **134**, 054119.

72 C. König and O. Christiansen, *J. Chem. Phys.*, 2015, **142**, 144115.

73 P. Seidler, T. Kaga, K. Yagi, O. Christiansen and K. Hirao, *Chem. Phys. Lett.*, 2009, **483**, 138–142.

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

75 S. Grimme, *J. Chem. Phys.*, 2006, **124**, 34108.

76 T. H. Dunning and T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.

77 R. A. Kendall, T. H. Dunning, R. J. Harrison and T. H. Dunning, *J. Chem. Phys.*, 1992, **96**, 6796–6806.

78 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

79 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

80 O. Christiansen, I. H. Godtliebsen, E. M. Gras, W. Györfi, M. B. Hansen, M. B. Hansen, E. L. Klinting, J. Kongsted, C. König, S. A. Losilla, D. Madsen, N. K. Madsen, P. Seidler, K. Sneskov, M. Sparta, B. Thomsen, D. Toffoli and A. Zoccante, *Molecular Interactions Dynamics And Simulation C++ (Mid-asCPP) Package*, <https://gitlab.com/midasCPP/midasCPP>, 2022.

81 M. Sparta, D. Toffoli and O. Christiansen, *Theor. Chem. Acc.*, 2009, **123**, 413–429.

82 D. Toffoli, M. Sparta and O. Christiansen, *Mol. Phys.*, 2011, **109**, 673–685.

83 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 73–78.

84 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2022, **12**, e1606.

85 F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.

86 V. Barone, M. Biczysko and J. Bloino, *Phys. Chem. Chem. Phys.*, 2014, **16**, 1759.

87 M. Biczysko, P. Panek, G. Scalmani, J. Bloino and V. Barone, *J. Chem. Theory Comput.*, 2010, **6**, 2115–2125.

88 T. K. Roy, V. Kopysov, A. Pereverzev, J. Šebek, R. B. Gerber and O. V. Boyarkin, *Phys. Chem. Chem. Phys.*, 2018, **20**, 24894–24901.

89 O. Vahtras, J. Almlöf and M. Feyereisen, *Chem. Phys. Lett.*, 1993, **213**, 514–518.

90 F. Neese, *J. Comput. Chem.*, 2003, **24**, 1740–1747.

91 F. Weigend, A. Köhn and C. Hättig, *J. Chem. Phys.*, 2002, **116**, 3175–3183.

92 S. Grimme, A. Hansen, S. Ehlert and J.-M. Mewes, *J. Chem. Phys.*, 2021, **154**, 064103.

93 M. Bursch, J.-M. Mewes, A. Hansen and S. Grimme, *Angew. Chem., Int. Ed.*, 2022, **61**, e202205735.

94 O. Christiansen, D. G. Artiukhin, F. Bader, I. H. Godtliebsen, E. M. Gras, W. Györffy, M. B. Hansen, M. B. Hansen, M. G. Højlund, N. Machholdt Hoyer, A. Buchgraitz Jensen, E. L. Klinting, J. Kongsted, C. König, D. Madsen, N. K. Madsen, K. Monrad, G. Schmitz, P. Seidler, K. Sneskow, M. Sparta, B. Thomsen, D. Toffoli and A. Zoccante, *Molecular Interactions Dynamics And Simulation C++ (MidasCPP) Package*, <https://gitlab.com/midasCPP/midasCPP>, 2023.

95 W. Digitizer, 2023, <https://apps.automeris.io/wpd/>.

96 K. Yagi, 2022, <https://tms.riken.jp/en/research/software/sindo/>.

97 NIST, 2023, <https://webbook.nist.gov/cgi/cbook.cgi?ID=C120809&Units=SI&Mask=80#IR-Spec>.

98 K. Yagi, T. Taketsugu and K. Hirao, *J. Chem. Phys.*, 2001, **115**, 10647–10655.

99 K. Yagi, G. V. Mil'nikov, T. Taketsugu, K. Hirao and H. Nakamura, *Chem. Phys. Lett.*, 2004, **397**, 435–440.

100 J. Hellmers and C. König, *J. Chem. Phys.*, 2023, **159**, 104108.

