UC Irvine
UC Irvine Previously Published Works

Title

First-principles anharmonic quantum calculations for peptide spectroscopy: VSCF calculations and comparison with experiments

Permalink

https://escholarship.org/uc/item/1xv4k3m1

Journal

Physical Chemistry Chemical Physics, 18(3)

ISSN

0956-5000

Authors

Roy, Tapta Kanchan
Sharma, Rahul
Gerber, R Benny

Publication Date

2016-01-21

DOI

10.1039/c5cp05979h

Copyright Information

Peer reviewed

CrossMark
click for updates

First-principles anharmonic quantum calculations for peptide spectroscopy: VSCF calculations and comparison with experiments†

Tapta Kanchan Roy,[ab] Rahul Sharma[c] and R. Benny Gerber*[ade]

First-principles quantum calculations for anharmonic vibrational spectroscopy of three protected dipeptides are carried out and compared with experimental data. Using hybrid HF/MP2 potentials, the Vibrational Self-Consistent Field with Second-Order Perturbation Correction (VSCF-PT2) algorithm is used to compute the spectra without any *ad hoc* scaling or fitting. All of the vibrational modes (135 for the largest system) are treated quantum mechanically and anharmonically using full pair-wise coupling potentials to represent the interaction between different modes. In the hybrid potential scheme the MP2 method is used for the harmonic part of the potential and a modified HF method is used for the anharmonic part. The overall agreement between computed spectra and experiment is very good and reveals different signatures for different conformers. This study shows that first-principles spectroscopic calculations of good accuracy are possible for dipeptides hence it opens possibilities for determination of dipeptide conformer structures by comparison of spectroscopic calculations with experiment.

Introduction

Biological functions of peptides and proteins largely depend on their 3D structures. The high conformational diversity in biomolecules sets a challenge to the high-resolution experiments and theoretical approaches to predict their accurate structures and the intramolecular force-fields. The gas phase high-resolution IR spectroscopy along with theoretical calculations can reliably reveal the intrinsic structure of small biomolecules. However, with the increasing size and conformational complexity, structural identifications become more challenging. In recent years, considerable advancements have been made in high-resolution measurements of mid-size peptides by using cryogenic cooling of biomolecular ions,[1–6] resonance two-photon IR,[7] fluorescence-dip spectroscopy,[8] and by other techniques.[9–13] These advanced methods are capable of yielding the IR spectroscopy of individual conformers which are free from interference

from one another. The vibrational spectra obtained by these experiments provide many characteristic features and constraints which have to be met by theoretical treatments. This challenges computational approaches to investigate the vibrational spectroscopy of biomolecules with adequate accuracy but reasonable computational effort. Several theoretical methods are available to compute the IR spectra. The first-principles approaches, which use suitable quantum chemical potentials and vibrational methods, have the importance of being free of adjustable parameters and of fitting to experiment. The simplest common approach in this direction is the harmonic oscillator (HO) approximation. While it is easily applicable for large systems, its accuracy is inadequate. Therefore, anharmonic treatment is essential for better accuracy. As an alternative approach, scaling corrections are used for the harmonic frequencies to match with experimental results. However, this approach is empirical and does not reflect on the potential energy surface (PES) underlying the system. On the other hand, rigorous vibrational methods are not feasible for large systems due to very high computational cost. Therefore, a suitable approximate method is needed which can be applied to large systems like peptides and proteins with adequate accuracy. One such inexpensive yet practical method is the VSCF approximation.[14–21] It has been successfully used for small to medium sized biomolecules.[22–26] Other alternatives such as perturbation theory[27,28] and classical MD simulations[29] are also available. The former has been applied with good accuracy for small biomolecules.[30] The Born–Oppenheimer MD (BOMD) simulations are applied for biomolecules where

[a] Institute of Chemistry and The Fritz Haber Research Center, The Hebrew University of Jerusalem, 91904, Israel. E-mail: taptakanchan@gmail.com, benny@fh.huji.ac.il; Tel: +972-2-6585732

[b] Department of Chemistry, Central University of Rajasthan, Rajasthan 305817, India

[c] Department of Chemistry, Indian Institute of Technology, Roorkee, Uttarakhand 247667, India

[d] Department of Chemistry, University of California, Irvine, CA 92697, USA

[e] Department of Chemistry, University of Helsinki, PO Box 55, 00014 Helsinki, Finland

This journal is © the Owner Societies 2016

Phys. Chem. Chem. Phys., 2016, **18**, 1607–1614 | **1607**

the nuclei are treated classically and the electrons are treated quantum mechanically using a DFT method.[31,32] However, at low temperature where the quantum effects are significant the classical MD calculations may result large errors especially for intensities.

In this contribution, we calculate the first-principles based anharmonic IR spectra of three protected dipeptides by means of quantum mechanically calculated PES using quantum VSCF-PT2 approximation[33] of a level that turns out to be sufficiently accurate for comparison with experiments. So far, VSCF calculations were carried out for peptides and a protein using empirical[34] and semi-empirical[25] potentials. However, these potentials are unrealistic at least for the purpose of spectroscopy. Recently Roy *et al.*[35] showed the possibilities to calculate the anharmonic spectroscopy of a decapeptide antibiotic gramicidine S. Due to the very large size (176 atoms) of the decapeptide, a few subsets of important vibrational modes (12–16 modes per set) were considered to construct the pairwise coupling potential. The results showed very good accuracy in comparison with experiment. In the present study, we used quantum chemical full pair-wise coupling potentials considering all the vibrational modes of the three dipeptides. Until now, VSCF-PT2 calculations using quantum chemical potentials of suitable accuracy were carried out for amino acids,[22,36] their complexes with water molecules[37] and other biomolecules[38] of similar sizes. The present paper extends the state-of-the-art quantum VSCF-PT2 calculations to dipeptides. Most of the earlier works on vibrational calculations of dipeptides were based on the HO[39] or use of the empirical scaling factors[7,8,40,41] or on retaining up to cubic coupling terms of the Taylor series expansion of the vibrational potential.[6]

The three protected dipeptides chosen here contain no free acid (COOH) and/or basic (NH₂) groups (Fig. 1). Such protections are extensively used in experimental studies. As a result, these are suitable to consider as model systems in quest of understanding the 3D structures of polypeptides. The first two systems are the two lowest energy conformers of methyl-capped dipeptides: *N*-acetyl tryptophan amide (NATA) represented by two different structural motifs, say C5 and C7.[8] The C5 structure (NATA-C5) consists of extended peptide backbone and the C7 structure (NATA-C7) contains a hydrogen bond of ψ-amide NH to φ-amide carbonyl group forming a seven member ring. The third structure is a dipeptide of Ac-Val-Phe-OMe (AVPO), protected at both the terminal positions by an acetyl and methyl group.[7] All the three structures were measured in gas phase and a few characteristic spectral data were reported. In those studies[7,8] the harmonic IR spectra at HF and DFT levels were computed and scaled by empirical scaling factors to bring the calculated frequencies close to the experiment. For AVPO, different scaling factors were used for higher and lower frequency ranges.[7] It is often found that such multiple scaling factors are needed depending on the degree of anharmonic couplings in vibrational transitions. While use of such multiple scaling allows a good fit for many of the vibrational modes it does not provide any specific details of the molecular PES since the factors are empirical and not obtained from an anharmonic potential. The empirical nature
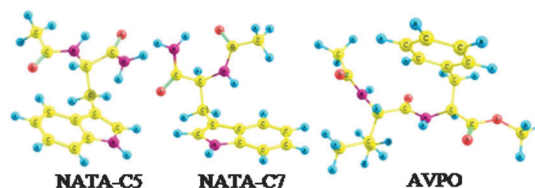

Fig. 1 Optimized structures of NATA-C5, NATA-C7 and AVPO at MP2/cc-pVDZ level.

of the anharmonic correction factors reduces the confidence in the vibrational prediction. In comparison with experiment, the 3D structure determinations thus may not be completely reliable. That motivates to compute the anharmonic vibrational spectra of large biomolecules using quantum mechanical VSCF approximation without any *ad hoc* adjustments.

Theory

The basic variant of VSCF theory uses a separable approximation.[14,15] It assumes that the full vibrational wave function is a product of single mode wave functions corresponding to each normal mode. Using a variational principle the final working equations for the wave functions and vibrational energy levels are derived. These involve mean field potentials that represent the average effects of the other modes on each mode. The corresponding equations are solved self-consistently using numerical techniques to get the vibrational states. The major challenge for this approach is the calculations of the multi-dimensional potentials which depend on its typical mathematical form. The VSCF potential can be written as a sum of one-mode, two-mode, three-mode terms and so on in mass weighted normal coordinate by (Q),

$$
\begin{aligned}
V(Q_1, \ldots, Q_N) \ = \ & \sum_{i=1}^{N} V_i^{\text{diag}}(Q_i) \\
& + \sum_{j}^{N} \sum_{i>j}^{N} V_{ij}^{2\text{coup}}(Q_i, Q_j) \\
& + \sum_{k=1}^{N} \sum_{j>k}^{N} \sum_{i>j}^{N} V_{ijk}^{3\text{coup}}(Q_i, Q_j, Q_k) + \cdots
\end{aligned}
\tag{1}
$$

The first term in eqn (1) is a sum of terms, where each is a single normal-mode (the so-called "diagonal approximation"). The second term consists of pair-wise couplings between different normal modes, *etc.* As a result of this form of the potential, the total number of grid points (NP) can be written as,

$$
\begin{aligned}
\text{NP} \ = \ & [\text{NV} \times \text{NG}] + \frac{[\text{NV}(\text{NV} - 1) \times \text{NG}^2]}{2} \\
& + \frac{[\text{NV}(\text{NV} - 1)(\text{NV} - 2) \times \text{NG}^3]}{6} + \cdots
\end{aligned}
\tag{2}
$$

where, NG is the number of grid points along a normal mode and NV is the number of vibrational modes. The first term gives one-dimensional grid points and yields intrinsic anharmonicity.

1608 | *Phys. Chem. Chem. Phys.*, 2016, **18**, 1607–1614

This journal is © the Owner Societies 2016

The next term gives the total number of two-dimensional grid points and hence pair-wise coupling and so on. It can be easily seen that the number of grid points increases very rapidly with the number of normal modes, number of grid points per normal modes and the dimensionality of the potential. In this work the full PES is simplified as a sum of single-mode plus a sum of pair-wise coupling interactions between the normal modes to carry out the calculations for the dipeptides of 93 and 135 vibrational modes in a reasonable time scale. It is already found that this pair-wise coupling approximation leads to good accuracy in many applications.[22,23,38] To bring in effect of non-separability, the vibrational levels can be further corrected by second-order perturbation theory (VSCF-PT2).[33,38] This method is typically more accurate than the basic VSCF approximation. Any perturbative methods may suffer from degeneracy issues which are discussed later. The corresponding vibrational band intensities (in km mol^{-1}) are computed by standard expression using anharmonic wave functions and the transition moments are calculated directly from the given electronic structure method. The value of the dipole moments of the three components are considered at the same points as the potential. Using an Eckart frame the integrated absorption coefficients for vibrational band intensities were computed as,

$$I_i = \frac{8\pi^3 N_A}{3hc}\omega_i \left| \left\langle \psi_i^{(0)}(Q_i) \middle| \vec{u}(Q_i) \middle| \psi_i^{(m)}(Q_i) \right\rangle \right|^2 (n_0 - n_m) \quad (3)$$

In this equation it is considered that all the initial population is in the vibrational ground state. Ignoring the thermal effect on population for simplicity the $(n_0 - n_m)$ is set equal to one, where n_0 and n_m are fractions of molecules in the initial and final states corresponding to zero temperature. $\psi_i^{(0)}$ and $\psi_i^{(m)}$ are VSCF wave functions for the ground and excited vibrational states. The vector $\vec{u}$, is a function of the nuclear coordinates and evaluated from electronic structure theory. Considering the equilibrium position as the origin, $\vec{u}$ can be expanded in powers of normal coordinates. It is a good approximation for fundamental and overtone transitions, at least for the low lying states, to consider only $\vec{u}(0,\ldots,Q_i,\ldots,0)$ which we denote by $\vec{u}(Q_i)$. The intensities are evaluated by numerical integration using the numerical anharmonic wavefunctions computed from the VSCF equations. More details of the VSCF and VSCF-PT2 theories can be found in a recent review.[38]

As the main computational bottleneck of the VSCF-PT2 calculations is the construction of the pair-wise coupling potentials which require calculations of the multi-dimensional grid points, the choice of the quantum mechanical methods to construct the potential is very crucial. It was found that both the DFT and post HF methods (such as MP2) are suitable for the VSCF calculations to get good accuracy. While the DFT based methods are more practical to apply for large systems than MP2, the latter gives slightly better results than the former.[22] However, calculations of the full pair-wise couplings by MP2 or DFT based methods are computationally intensive for such large systems. The harmonic part of the potential requires the greatest accuracy, since the harmonic contribution to each frequency is much larger than the anharmonic one.

Considering this fact we used a hybrid potential scheme by "upgrading" PESs of a lower level to produce more accurate anharmonic frequencies.[36] This improved potential can be written as

$$V^{\text{ULM}}(Q_1,\ldots,Q_N) = V^{\text{LM}}(\lambda_1 Q_1,\ldots,\lambda_N Q_N),$$

$$\text{where } \lambda_i = \frac{w_i^{\text{HM}}}{w_i^{\text{LM}}} \quad (4)$$

Here V^{LM} and V^{ULM} are the potentials for N-mode systems at the low-level method (LM) and upgraded low-level method (ULM), respectively. The scaling coefficients λ_i are the ratio of the harmonic frequencies of the LM (w_i^{LM}) to the high-level method (w_i^{HM}). These scaling coefficients are chosen such that the upgraded potential reproduces the harmonic frequencies of the high level potential. It was applied successfully for several small systems.[26,36] The upgraded potentials constructed here are calculated from the much faster HF method while the harmonic potentials at this level are upgraded by much more accurate MP2 method. We used double-zeta Dunning type cc-pVDZ basis and 8 grid points per normal modes throughout this study and calculated the fundamental bands for all the vibrational modes. All the calculations are performed using locally modified GAMESS program.[42]

Results and discussions

The structures of the two conformers of NATA and AVPO are taken from the previous studies.[7,8] It is found that, similar to the preceding studies, the NATA-C7 is the global minima and it is more stable than NATA-C5 by 1.2 kcal mol^{-1} (with ZPE) at MP2/cc-pVDZ level of theory. Geometry optimizations were performed using gradient convergence tolerance of 0.0001 Hartree per Bohr and subsequent harmonic frequency calculations were performed to confirm that each structure is in the true minimum.

Peptides or proteins give rise to characteristic IR absorption bands namely N–H stretching, C–H stretching, C=O stretching (amide I) and N–H bending (amide II). These are most important IR bands to reveal the conformational changes of proteins and peptides. The first two modes are found generally above 2700 cm^{-1} and the last two modes are found below 1700 cm^{-1}.

In Table 1, the computed HO frequencies in HF and MP2 potentials and VSCF-PT2 along with VSCF frequencies in HF/MP2 potentials are presented and compared with the experiment for NATA-C5. In Fig. 2a, the corresponding VSCF-PT2 spectra are given and compared with experiment. The VSCF values are given in the table to show the extent of corrections due to PT2 approximation. As stated earlier, any perturbative method occasionally suffers from resonance problem particularly for low frequency modes where the spectral congestion is very high. Considering the large size of the three peptides, it is expected that these must contain several low frequency vibrations which ought to be more vulnerable as far as resonance is concerned, leading to numerical instabilities. For example, we found numerical convergence problem in VSCF-PT2 calculations

Table 1 Comparisons of calculated vibrational fundamental stretching frequencies with experiment for NATA-C5

Mode type	Exp.[a]	HO (HF)	HO (MP2)	VSCF[b]	VSCF-PT2[b]
Asym-NH$_2$	3538	3942	3726	3551	3537
Indole NH	3523	3911	3680	3551	3568
φ-Amide NH	3430	3869	3612	3421	3424
Sym-NH$_2$	3417	3802	3573	3455	3428
Phenyl CH	3074	3367	3239	3105	3071
Phenyl CH	3059	3356	3227	3078	3054
Phenyl CH	—	3344	3215	3074	3050
Alkyl CH	3004	3293	3217	3030	3002
Alkyl CH	—	3287	3194	3021	2993
Alkyl CH	2956	3255	3149	2991	2966
Alkyl CH	—	3202	3098	2965	2917
Alkyl CH	2933	3201	3080	2912	2889

[a] Ref. 8. [b] Calculated using hybrid HF/MP2 potential with cc-pVDZ basis.
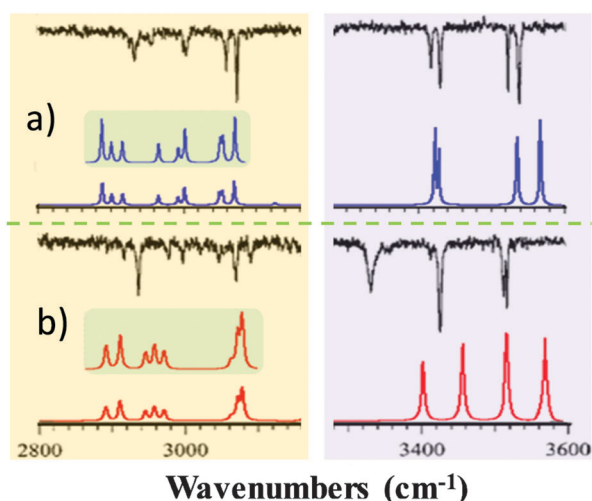
Fig. 2 Comparisons of calculated and experimental IR spectra of NATA. (a) Experiment (black trace) *vs.* calculated (blue trace) of NATA-C5; (b) experiment (black trace) *vs.* calculated (red trace) of NATA-C7. A Lorentzian width function with HWHM of 5 cm^{-1} is used for the computed spectra (inset: enlarged computed spectra).

for the second lowest mode (torsional mode of $\sim$25 cm^{-1}) of the NATA-C5 mostly due to resonance. Keeping this mode as harmonic the convergence problem is removed.[43] However, the level of accuracy in the peak positions showed that the resonances contributed weakly at least to the peaks we assigned. NATA has four NH stretching transitions, two from NH$_2$ group, one from indole group and one from φ-amide NH. Only those VSCF-PT2 transitions are given in the tables for which the experimental data were observed (see ESI† for the full list of computed vibrations and intensities). The comparison between theoretical and experimental transitions is made for each type of vibrations considering peak positions and intensity pattern. The theoretical spectra of NATA-C5 in comparison to experiment are given in the Fig. 2a. Note that, VSCF algorithm can compute frequencies and intensities but not the spectral line shapes. It requires to treat spectral width and line broadening which yield spectral line shapes. The Lorentzian broadening

function with HWHM of 5 cm^{-1} is used to calculate the spectral line shapes in this study. As it can be seen, very good agreements are found for peak positions of the NH stretches. The bands located at 3538 and 3417 cm^{-1} corresponding to asymmetric and symmetric NH$_2$ stretching modes match extremely well in terms of positions and band shapes with errors of only 1 and 11 cm^{-1}, respectively for the band positions. The calculated φ-amide NH stretching mode at 3424 cm^{-1} is accurately located at the experimental band position (3430 cm^{-1}) with similar intensity. Only the indole NH stretching is blue shifted by 45 cm^{-1}. It may happen due to the fact that indole NH stretch has hardly any interaction with the different dipeptide backbones and hence coupled weakly with other modes. That reduces the degree of anharmonic corrections to retain this peak blue of the experiment. Two adjacent peaks of φ-amide NH and symmetric NH$_2$ stretches which are separated by 13 cm^{-1} in experiment are found to be separated by 4 cm^{-1} in the computed spectra.

The CH stretching vibrations of NATA include pyroll, aromatic and alkyl CH stretches. The experimental spectra showed that this region suffers from spectral congestion with mixing of Fermi resonances due to CH bending modes. For such cases the computed spectra can provide support to assign the measured spectra more clearly. For example, two aromatic CH stretches were experimentally measured at 3074 and 3059 cm^{-1} in NATA-c5. In theoretical calculations it is found that four such transitions are possible with three intense peaks and one very weak peak (Table 1 and Table S1, ESI†). Comparing the experimental values with the theoretical ones, very good agreement was found for both the peaks with an error of only 3 and 5 cm^{-1}, respectively. The intensity pattern is also good. A third computed peak at 3050 cm^{-1} with similar intensity of 3054 cm^{-1} indicates that one more experimental peak very close to 3059 cm^{-1} may be present which might be overlapped with the former to give a single peak instead. In the alkyl CH stretching region ($\sim$2930–3010 cm^{-1}) three peaks were experimentally identified out of six possible peaks. Looking at the experimental intensity pattern we found that the two extreme peaks can be characterized by their higher intensity than the others. Our computed intensities nicely agree with the same trend. The overall match in this region is very good where the first two peaks result in good agreement and the last peak is red shifted with an error of 1.5%. Though the intensity pattern in this region matches very well the experiment, the overall line heights of the computed CH intensities are relatively smaller than the experiment. As stated earlier, we obtained the spectral line shapes by fitting to Lorentzian function. This is arbitrary in nature and consequently often yields unsatisfactory result compared to experiment. In this study we also found that the line shapes are unsatisfactory in some cases while the intensities and particularly the peak positions are in good agreement with experiment. Overall good agreement is found for NATA-C5 with mean absolute percentage error (MAPE) of only 0.46% (see Table S1, ESI†) for the peak positions.

Next, we turn to the NATA-C7 conformer and its comparisons with experiment and also with NATA-C5. At first glance

Table 2 Comparisons of calculated vibrational fundamental stretching frequencies with experiment for NATA-C7

| Mode type	Exp.[a]	HO (HF)	HO (MP2)	VSCF[b]	VSCF-PT2[b]
Asym-NH$_2$	3516	3921	3720	3534	3521
Indole NH	3521	3911	3679	3562	3573
φ-Amide NH	3429	3861	3610	3453	3461
Sym-NH$_2$	3334	3773	3520	3427	3406
Phenyl CH	3092	3366	3239	3110	3076
Phenyl CH	3072	3353	3226	3095	3082
Phenyl CH	3048	3341	3214	3079	3067
Alkyl CH	2999	3281	3174	2986	2975
Alkyl CH	2979	3310	3212	2989	2961
Alkyl CH	2938	3211	3101	2940	2915
Alkyl CH	2919	3237	3084	2906	2895
Alkyl CH	—	3199	3173	3003	2962

[a] Ref. 8. [b] Calculated using HF/MP2 potential with cc-pVDZ basis.

(Fig. 2a and b), it can be seen that the spectral pattern of the two NATA conformers show distinct features in particular for NH stretch regions. However, comparing with experiment, NATA-C7 results less satisfactory agreement particularly in the NH stretching regions. In NATA-C7, the measured asymmetric and symmetric stretching modes for NH$_2$ are found at 3516 and 3334 cm^{-1} while the computed values are located at 3521 and 3406 cm^{-1}, respectively (Table 2). In comparison to NATA-C5, large experimental red shift in the symmetric NH$_2$ stretch is due to the intra-molecular hydrogen bonding. While the corresponding computed asymmetric mode matches very well with the deviation of only 5 cm^{-1}, the symmetric mode yields large blue shift with an error of 2.16%. This high error is probably due to the inaccuracy in the electronic structure method itself (HF in this case) which may be inadequate to describe this hydrogen bond and hence the anharmonicity accurately. As a simple test to assess this error, we compared the corresponding NH and the adjacent hydrogen bond distances at HF and MP2 optimized structures. While the NH and hydrogen bond distances by MP2 are 1.02 and 1.98 Å, HF yields 1.00 and 2.15 Å, respectively. Clearly, HF shows more localized interactions by binding the hydrogen more tightly towards the nitrogen and consequently underestimates the hydrogen bond interactions. As a result the corresponding NH vibration shows large blue shift. The indole NH stretch is once again blue shifted, similar to NATA-C5, with an error of 1.47%. These two large blue shifts in NATA-C7 lead to the final spectral pattern rather unsatisfactorily. The remaining φ-amide NH stretch shows reasonable accuracy with a deviation of 32 cm^{-1} to the blue of the experiment. Due to such blue shifts the final computed spectra in this region (Fig. 2b) show some inconsistencies in comparison with experiment. The separation between the experimental asymmetric NH$_2$ and indole NH stretching peaks is only 5 cm^{-1}. However, the large blue shift in the computed indole NH stretch increases this splitting by 52 cm^{-1}. Similarly, due to the large blue shift in the computed symmetric NH$_2$ mode, the extent of separation between it and the indole NH stretching is narrowed down in comparison with measured spectra. However, the overall intensity pattern is good in this region.

The aromatic CH stretches of NATA-C7 show very good agreement with the measured spectra. The deviations are less than 20 cm^{-1} for all the three transitions. The same good agreement is found for the alkyl CH stretches where the errors are within 25 cm^{-1} for all the cases. Although there are some large errors in the NH stretching regions, the overall good accuracy is found with the MAPE only 0.83% (see Table S2, ESI†) due to good accuracy in rest of the modes. It is found that the VSCF and the VSCF-PT2 values of the second lowest mode of NATA-C7 (torsional mode, about 45 cm^{-1}) converge to negative frequency values, which are obviously unphysical. The harmonic and diagonal values are reasonable. In cases such as this, the VSCF and VSCF-PT2 values for this frequency should be disregarded. This is a case where the VSCF algorithm fails to converge to a physical value, but frequencies of other modes, if physical can be retained.[38] The intensities, as can be seen, agree nicely in comparison with the experimental vibrational pattern.

The third dipeptide AVPO is the largest system studied here and experimental data is available for both the high and mid-IR ranges. We tested a few other possible conformers by B3LYP/cc-pVDZ method and found several low energy structures, but the reported one remains as the global minima. The measured and calculated spectra are shown in Fig. 3 (corresponding spectral data are given in Table 3 and Table S3, ESI†). As it was found for NATA, two distinct high frequency regions are observed in experiment due to NH stretches around 3400–3450 cm^{-1} and CH stretch in the range of ∼2950–3100 cm^{-1}. There are two NH stretches located above 3400 cm^{-1} which indicates that the molecule is free from any hydrogen bonding (i.e. C=O···H–N). The computed spectra of two NH stretches due to Phe and Val sub-unit are slightly underestimated (∼30 cm^{-1}) from experiment with an error below 1%. Experimental spectra show four aromatic CH stretches in the range of 3000–3100 cm^{-1} with a maxima at 3038 cm^{-1}. The corresponding computed maxima is found at 3071 cm^{-1} which is blue shifted to the experiment by
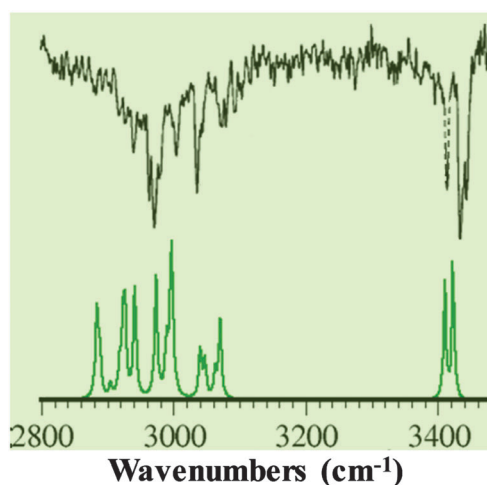
Fig. 3 Comparison of calculated (green trace) and experimental (black trace) IR spectra of AVPO. A Lorentzian width function with HWHM of 5 cm^{-1} is used for the computed spectra. The dotted peak at 3420 cm^{-1} in experiment is the overtone of acetyl C=O stretching. Due to VSCF convergence problem this peak is not shown in the calculated spectra.

Table 3 Comparisons of calculated vibrational fundamental stretching frequencies with experiment for AVPO

Mode type	Exp.[a]	HO (HF)	HO (MP2)	VSCF[b]	VSCF-PT2[b]
Phe NH str.	3451	3866	3607	3423	3423
Val NH str.	3441	3842	3592	3406	3411
Phenyl CH str.	3096	3370	3245	3111	3064
Phenyl CH str.	3076	3350	3219	3074	3048
Phenyl CH str.	3038	3359	3235	3095	3071
Phenyl CH str.	3006	3332	3202	3062	3040
Methyl CH str. of Val	2974	3278	3183	3019	2998
Methyl CH str. of Val	2965	3268	3180	3008	2974
Methyl CH str. of Val	2941	3259	3177	2975	2942
Phenyl CO str.	1765	1998	1818	1790	1789
Acetyl CO str.	1711	1950	1780	1755	1753
Val CO str.	1696	1913	1764	1731	1730

[a] Ref. 7. [b] Calculated using HF/MP2 potential with cc-pVDZ basis.

Fig. 4 Histogram for error analysis. Comparisons of % error by VSCF–PT2 against experiments (considering all the available experimental transitions).

33 cm^{-1}. Comparing with the intensities we found that the other three aromatic CH stretches match well with experiment and the error range is around 1.1%. Below 3000 cm^{-1}, in the region of aliphatic CH stretches, three transitions were measured at 2974, 2965 and 2941 cm^{-1} respectively, with a maximum of 2974 cm^{-1}. These vibrations attributed to intense CH stretches of the methyl group of Val. The corresponding calculated spectra are found at 2998, 2974 and 2942 cm^{-1} respectively, with a maximum of 2998 cm^{-1}. As it can be seen two out of three peaks match very well with the experiment. Intensity pattern is also good in this region. The experimental transitions in the mid-IR ranges are observed for AVPO. In this region three C=O stretches due to Phe, acetyl and Val were measured and located at 1765, 1711 and 1696 cm^{-1}, respectively. The corresponding calculated frequencies are found at 1789, 1753 and 1730 cm^{-1} where all of them are blue shifted with deviations ∼20–40 cm^{-1}. The MAPE for AVPO is within 1.1% where most of the transitions match well with the experiment (Fig. S1, ESI†).

Overall it is found that for most of the cases we have accuracy within 10–20 cm^{-1}. For some cases we have deviations up to 40–50 cm^{-1}. Ideally, if one increases the level of electronic structure theory (*i.e.*, basis set) the accuracy should increase by paying much higher computational cost. But practically that may not happen always as occasionally the cancellation/addition of errors nullify the expected improvement. On the whole the level of agreement with experiment found here seems to be roughly of the level of agreement obtained so far for much smaller molecules, *e.g.* amino acids. This is therefore a significant step forward in this field.

Conclusions

This work shows that the approximate first-principles calculations for anharmonic molecular vibrations using *ab initio* PESs offer very good agreement with high resolution experiments for three dipeptides of as large as 135 vibrational modes. To our knowledge, so far these are the largest systems where the full pair-wise couplings for all the modes are used in quantum mechanical VSCF-PT2 calculations using first-principles potentials.
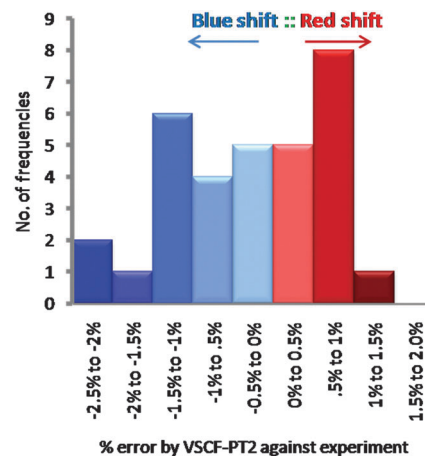
The overall comparison with experiment is very good for most cases and shows distinctive features for each conformer. The choice of potential (hybrid) plays the key role for accurate and affordable calculations of such large systems. Only for a very few transitions, comparatively larger errors are obtained due to inadequacy in the HF potential itself. However, keeping in mind of the large size of the peptides we have considered, very fast algorithm like HF is desirable to pursue the computations in reasonable time scale with the payment of some accuracy. Resonances are another possible setback which may invoke numerical instabilities in the VSCF-PT2 calculations and may yield errors in the final spectra.

A statistical analysis of % errors (Fig. 4) of the computed frequencies to the available experimental frequencies for all the three systems shows that more number of computed frequencies are slightly blue shifted than the red. For example, it is found that, most of the calculated hydrogenic stretches are distributed between −1.5 to 1.0% error range to the experiment. It shows that HF/MP2 hybrid potential is stiffer and binds stronger for such transitions. The same trend was found for pure MP2 potential[22] and it is no surprise that the hybrid HF/MP2 potential also reflects the same trend. It is also expected that HF, being a lower level electronic structure method than MP2, can be less accurate but more suitable to carry out the calculations for such large systems within reasonable computational time. We expect that quantum calculations, as the VSCF-PT2 done here, will be more accurate than the classical MD at low temperature. As it was found in some previous studies while the less anharmonic far-IR spectra showed good accuracy by MD, the mid-IR and high frequency region showed considerable deviations in peak positions as well as intensities.[31,32] It is anticipated that the commonly used BLYP potential for MD calculations should be less accurate than the B3LYP or MP2 and the latter methods are computationally costlier for MD. We also note that empirically scaled HO values can yield good agreement using different scaling factors to fit the experiment best. The present anharmonic method is a first-principles one, based on quantum calculations

1612 | *Phys. Chem. Chem. Phys.*, 2016, **18**, 1607–1614

This journal is © the Owner Societies 2016

for both the potentials and the vibrations, with no adjustable parameters. The approach is thus unbiased unlike any empirical scaling. The level of matching with experiment highly increases the confidence about the accuracy of the corresponding vibrational spectra and molecular structure. Thus such approach can be used for even more challenging systems *e.g.* peptide–water complexes. The method used leads to the determination of an accurate anharmonic part of the potential, as the test of spectroscopy establishes. Moreover, knowledge of the anharmonic part of the potential is important for a range of molecular properties, such as intramolecular vibrational energy flow.[44] In the harmonic approximation, there is no coupling between different normal modes, thus the effect of vibrational energy flow between different modes is all due to the anharmonic part of the potential. Finally, we show the limitations of the state-of-the-art anharmonic calculations with the increase in molecular size. This paper shows that the calculations for small peptides (dipeptides) can be done by anharmonic quantum calculations, using good enough potentials (the hybrid HF/MP2) almost as a matter of routine and hence is applicable for determining the geometry of conformers and interactions for such systems. This opens a way to study mid-size peptides in gas-phase, peptide–water complexes and other biomolecules.

Acknowledgements

References

1 T. R. Rizzo, J. A. Stearns and O. V. Boyarkin, *Int. Rev. Phys. Chem.*, 2009, **28**, 481–515.

2 S. Schwing, C. Reyheller, A. Schaly, S. Kubik and M. Gerhards, *ChemPhysChem*, 2011, **12**, 1981–1988.

3 A. M. Rijs, M. Kabeláč, A. Abo-Riziq, P. Hobza and M. S. de Vries, *ChemPhysChem*, 2011, **12**, 1816–1821.

4 N. S. Nagornova, T. R. Rizzo and O. V. Boyarkin, *J. Am. Chem. Soc.*, 2010, **132**, 4040–4041.

5 N. S. Nagornova, M. Guglielmi, M. Doemer, I. Tavernelli, U. Rothlisberger, T. R. Rizzo and O. V. Boyarkin, *Angew. Chem., Int. Ed.*, 2011, **50**, 5383–5386.

6 C. M. Leavitt, A. F. DeBlase, C. A. Johnson, M. v. Stipdonk, A. B. McCoy and M. A. Johnson, *J. Phys. Chem. Lett.*, 2013, **4**, 3450–3457.

7 C. Unterberg, A. Gerlach, T. Schrader and M. Gerhards, *J. Chem. Phys.*, 2003, **118**, 8296–8300.

8 B. C. Dian, A. Longarte, S. Mercier, D. A. Evans, D. J. Wales and T. S. Zwier, *J. Chem. Phys.*, 2002, **117**, 10688–10702.

9 A. Y. Ivanov, G. Sheina and Y. P. Blagoi, *Spectrochim. Acta, Part A*, 1999, **55**, 219–228.

10 A. Y. Ivanov, A. M. Plokhotnichenko, V. Izvekov, G. G. Sheina and Y. P. Blagoi, *J. Mol. Struct.*, 1997, **408**, 459–462.

11 L. C. Snoek, R. T. Kroemer and J. P. Simons, *Phys. Chem. Chem. Phys.*, 2002, **4**, 2130–2139.

12 E. Nir, L. Grace, B. Brauer and M. S. de Vries, *J. Am. Chem. Soc.*, 1999, **121**, 4896–4897.

13 M. S. de Vries and P. Hobza, *Annu. Rev. Phys. Chem.*, 2007, **58**, 585–612.

14 J. M. Bowman, *J. Chem. Phys.*, 1978, **68**, 608–610.

15 R. B. Gerber and M. A. Ratner, *Chem. Phys. Lett.*, 1979, **68**, 195–198.

16 G. Rauhut, *J. Chem. Phys.*, 2004, **121**, 9313–9322.

17 B. C. Garrett and D. G. Truhlar, *Chem. Phys. Lett.*, 1982, **92**, 64–70.

18 T. K. Roy and M. D. Prasad, *J. Chem. Sci.*, 2009, **121**, 805–810.

19 T. K. Roy and M. D. Prasad, *J. Chem. Phys.*, 2009, **131**, 174102.

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

21 M. Keceli and S. Hirata, *J. Chem. Phys.*, 2011, **135**, 134108.

22 T. K. Roy, T. Carrington Jr. and R. B. Gerber, *J. Phys. Chem. A*, 2014, **118**, 6730–6739.

23 K. Meng and J. Wang, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2001–2013.

24 B. Brauer, R. B. Gerber, M. Kabeláč, P. Hobza, J. M. Bakker, A. G. A. Riziq and M. S. de Vries, *J. Phys. Chem. A*, 2005, **109**, 6974–6984.

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

26 R. Knaanie, J. Sebek, J. Kalinowski and R. B. Gerber, *Spectrochim. Acta, Part A*, 2014, **119**, 2–11.

27 V. Barone, *J. Chem. Phys.*, 2005, **122**, 014108.

28 E. L. Sibert III, *J. Chem. Phys.*, 1988, **88**, 4378–4390.

29 M. P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3336–3359.

30 V. Barone, M. Biczysko, J. Bloino and C. Puzzarini, *Phys. Chem. Chem. Phys.*, 2012, **15**, 1358–1363.

31 S. Jaeqx, J. Oomens, A. Cimas, M.-P. Gaigeot and A. M. Rijs, *Angew. Chem., Int. Ed.*, 2014, **53**, 3663–3666.

32 A. Cimas, T. D. Vaden, T. S. J. A. de Boer, L. C. Snoek and M. P. Gaigeot, *J. Chem. Theory Comput.*, 2009, **5**, 1068–1078.

33 G. M. Chaban, J. O. Jung and R. B. Gerber, *J. Chem. Phys.*, 1999, **111**, 1823–1829.

34 A. Roitberg, R. B. Gerber, R. Elber and M. A. Ratner, *Science*, 1995, **268**, 1319–1322.

35 T. K. Roy, V. Kopysov, N. S. Nagornova, T. R. Rizzo, O. V. Boyarkin and R. B. Gerber, *ChemPhysChem*, 2015, **16**, 1374–1378.

36 B. Brauer, G. M. Chaban and R. B. Gerber, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2543–2556.

37 G. M. Chaban and R. B. Gerber, *J. Chem. Phys.*, 2001, **115**, 1340–1348.

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

39 C. J. Johnson, A. B. Wolk, J. A. Fournier, E. N. Sullivan, G. H. Weddle and M. A. Johnson, *J. Chem. Phys.*, 2014, **140**, 221101–221105.

This journal is © the Owner Societies 2016

Phys. Chem. Chem. Phys., 2016, **18**, 1607–1614 | **1613**

40 M. Z. Kamrath, E. Garand, P. A. Jordon, C. M. Leavitt, A. B. Wolk and M. J. Van Stipdonk, *J. Am. Chem. Soc.*, 2011, **133**, 6440–6448.

41 H. Fricke, A. Funk, T. Schrader and M. Gerhards, *J. Am. Chem. Soc.*, 2007, **130**, 4692–4698.

42 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347–1363.

43 P. Daněček, J. Kapitán, V. Baumruk, L. Bednárová, V. Kopecký Jr. and P. Bouř, *J. Chem. Phys.*, 2007, **126**, 224513.

44 R. B. Gerber and M. A. Ratner, *Adv. Chem. Phys.*, 1988, **70**, 97–132.