



Cite this: *Phys. Chem. Chem. Phys.*,  
2018, 20, 4926

# Insight into vibrational circular dichroism of proteins by density functional modeling†

Jiří Kessler,<sup>a</sup> Valery Andrushchenko,<sup>id</sup><sup>a</sup> Josef Kapitán<sup>b</sup> and Petr Bouř<sup>id</sup>★<sup>a</sup>

Vibrational circular dichroism (VCD) spectroscopy is an excellent method to determine the secondary structure of proteins in solution. Comparison of experimental spectra with quantum-chemical simulations represents a convenient and objective way to extract information on the structure. This has been difficult for such large molecules where approximate theoretical models have to be used. In the present study we applied the Cartesian-coordinate based tensor transfer (CCT) making it possible to extend the density functional theory (DFT) and model spectral intensities of large globular proteins nearly at quantum-chemical precision. Indeed, comparison with experiment provided a better understanding of the dependence of VCD spectral shapes on the geometry, their sensitivity to fine structural details and interactions with the environment. On a model set of globular proteins the simulated spectra correlated well with experimental data and revealed which structural information can (and cannot) be obtained from this kind of spectroscopy. Although the VCD technique has been regarded as being rather insensitive to side-chain variations, we found that the spectra of human and hen lysozyme differing by a few amino acids only are quite distinct. This has been explained by long-distance coupling of the amide vibrations. Likewise, the modeling reproduced some spectral changes caused by protein deuteration even when the protein structure was conserved.

Received 29th November 2017,  
Accepted 22nd January 2018

DOI: 10.1039/c7cp08016f

rsc.li/pccp

## 1. Introduction

Vibrational circular dichroism (VCD) spectroscopy measures the absorption difference between left- and right-circularly polarized infrared light. The measurement is more difficult than for plain infrared absorption (IR), but the spectra are more sensitive to molecular structure. Available vibrational transitions are often more numerous and convey more information than the electronic ones. VCD is, for example, one of the very few methods by which the absolute configuration of small chiral molecules can be reliably determined.<sup>1,2</sup> It has been also widely applied to biopolymers, such as nucleic acids and proteins, in most cases to monitor their structure and interaction in solutions.<sup>3–7</sup>

Although molecular VCD was observed already in the 1970s,<sup>8</sup> rigorous quantum-chemical formulation appeared much later.<sup>9,10</sup> Nowadays, Stephen's magnetic field perturbation formulation<sup>10,11</sup> implemented within the density functional theory (DFT)<sup>12</sup> is a widely accepted standard for medium-sized molecules, well-balancing the accuracy and computational cost. An important

component in the theory development was also an origin-independent algorithm for computing VCD intensities, such as that based on the gauge-invariant atomic orbitals.<sup>13</sup>

For molecules composed of hundreds and thousands of atoms, even the DFT cannot be used, or its direct application is very inefficient in terms of required computer time and memory. Fragmentation schemes, such as the “atom in molecules” approach, are thus increasingly popular to simulate VCD and other vibrational optical activity spectra of such systems.<sup>14</sup> In the present study, we apply the Cartesian-coordinate tensor transfer (CCT)<sup>15–17</sup> to a set of globular proteins. The method has been recently automated so that it can be applied to molecules of up to ~10 000 atoms (the most serious limitation being the diagonalization of the harmonic force field) and used to simulate Raman optical activity (ROA).<sup>18</sup>

The computational methodology thus significantly increases the application potential of VCD spectroscopy. The technique is widely used to understand not only isolated proteins<sup>19</sup> but also their aggregates, including amyloidal fibrils implicated in various neurodegenerative diseases,<sup>20</sup> as well as other biopolymers<sup>21</sup> and new bio-inspired materials.<sup>22</sup> However, the origin of spectral signals is often poorly understood. The globular proteins investigated in the present study are reasonably rigid so that they can be used as benchmarks where specific spectral features can be more reliably assigned to a given structural motif. Of particular interest is the observation that proteins with rather similar

<sup>a</sup> Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Flemingovo náměstí 2, 16610 Prague, Czech Republic. E-mail: bour@uochb.cas.cz

<sup>b</sup> Department of Optics, Palacký University, 17. listopadu 12, 77146 Olomouc, Czech Republic

† Electronic supplementary information (ESI) available: Further computational tests and spectral analyses. See DOI: 10.1039/c7cp08016f



secondary structure often exhibit different spectral patterns. This paves the way for inspecting tertiary structure, *e.g.* in protein aggregates.

## 2. Method

### Measurement of VCD and IR spectra

Hen-egg-white and human lysozyme, human serum albumin, myoglobin, lactalbumin, insulin and concanavalin A (all from Sigma-Aldrich) were dissolved to a concentration of 300 mg mL<sup>-1</sup>. By default, MilliQ water was used; concanavalin A was dissolved in 100 mM phosphate buffer, pH = 6, and for insulin the pH was adjusted to 2.5 with HCl. VCD measurements were performed using a BioTools ChiralIR-2X instrument and the samples were placed in a demountable custom made (by Meopta, Inc.) CaF<sub>2</sub> cell devoid of any spacer<sup>23</sup> as it would be problematic for the 8 μm path length. High protein concentration and thin path length are needed to collect data in a wide spectral range of ~1200–2000 cm<sup>-1</sup>, which would be otherwise partially obscured by water absorption. Additional measurements were carried out in a BaF<sub>2</sub> cell using both D<sub>2</sub>O- and H<sub>2</sub>O-based solvents, with 50 and 15 μm spacers and protein concentrations of 30 and 150–250 mg mL<sup>-1</sup>, respectively. This reduced spectral noise but narrowed the usable wavenumber interval. Each spectrum was accumulated for 16 hours at 8 cm<sup>-1</sup> resolution. Absorption (IR) and VCD spectra of pure solvents were subtracted to correct for instrumental artifacts.

### Computations

The computation of vibrational protein properties from protein fragments closely followed that described in detail in ref. 18. Briefly, X-ray data from the Protein Data Bank (PDB) database (<http://www.rcsb.org>) were used as starting geometries. The simulations were performed for ionized species corresponding to the experimental conditions (neutral pH, except for insulin). The main peptide chain was split into overlapping fragments, each containing four amide (three amino-acid) residues and capped with methyl groups. For lactalbumin, shorter and longer fragments (containing 2–5 amide groups) were used to test the methodology (Fig. S1, ESI†). The four-amide approach appeared to be a reasonable compromise between accuracy and computational cost, although the limited fragment length can still lead to some variations of the amide I VCD intensity (around 1720 cm<sup>-1</sup>). In some cases, small contact fragments were added to model interactions between close peptide chains; these, however, did not significantly alter principal spectral features (Fig. S2, ESI†).

Each fragment geometry was partially optimized by restrained energy-minimization in vibrational normal mode coordinates<sup>24–26</sup> at the B3PW91<sup>27</sup>/CPCM(water)<sup>28,29</sup>/6-31++G\*\* level; the Gaussian 09 v D01 program was used for all quantum chemistry.<sup>30</sup> The B3PW91 functional previously provided excellent results for similar systems.<sup>31</sup> Then the force field and dipole derivatives (atomic polar and axial tensors)<sup>32,33</sup> were transferred from all fragments to the original protein using the Cartesian coordinate transfer (CCT) algorithm.<sup>15,16</sup>

For some proteins the effect of neglecting longer-range interactions within CCT was roughly estimated by adding the transition dipole coupling (TDC) correction<sup>34</sup> to the transferred force field. For each atom pair *i, j* not included in the fragmentation scheme the force constants (second energy derivatives) were obtained as

$$f_{ix,j\beta} = \frac{\partial^2 E}{\partial r_{ix} \partial r_{j\beta}} = \frac{1}{4\pi\epsilon_r\epsilon_0} \frac{\mathbf{P}_{ix} \cdot \mathbf{P}_{j\beta} r_{ij}^2 - 3\mathbf{P}_{ix} \cdot \mathbf{r}_{ij} \mathbf{P}_{j\beta} \cdot \mathbf{r}_{ij}}{r_{ij}^5} \quad (1)$$

where  $\epsilon_r$  and  $\epsilon_0$  are the relative and vacuum dielectric constants,  $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$  is the distance vector between the atoms,  $\mathbf{P}_{ix} = \frac{\partial \mathbf{p}}{\partial r_{ix}}$  is the atomic polar tensor<sup>35</sup> (dipole derivatives),  $\mathbf{p}$  is the electric dipole moment,  $\alpha$  and  $\beta$  denote Cartesian coordinates, and the dots indicate scalar products.

Absorption and VCD intensities were generated at the harmonic level<sup>33,36</sup> and smooth spectra were obtained by a convolution with Lorentzian functions of 10 cm<sup>-1</sup> bandwidth (full width at half maximum).

## 3. Results and discussion

### Characteristic features of protein IR and VCD spectra

Typical calculated IR and VCD spectra and their conventional subdivision by characteristic protein vibrations are documented in Fig. 1 for the example of lactalbumin. The spectra were simulated for the whole protein as well as for an analogue in which all side chains were replaced by alanine (methyl groups). In either case a “classical” protein IR spectrum is obtained,<sup>37,38</sup> it is dominated by C=O stretching signals (“amide I” region), followed by “amide II”, *i.e.* a combination of N–H bending and C(carbonyl)–N stretching motions.

The amide I band is simulated at a higher frequency (~1725 cm<sup>-1</sup>) than obtained experimentally (~1650 cm<sup>-1</sup>),<sup>39</sup> which is a notorious problem of such *ab initio* simulations, and

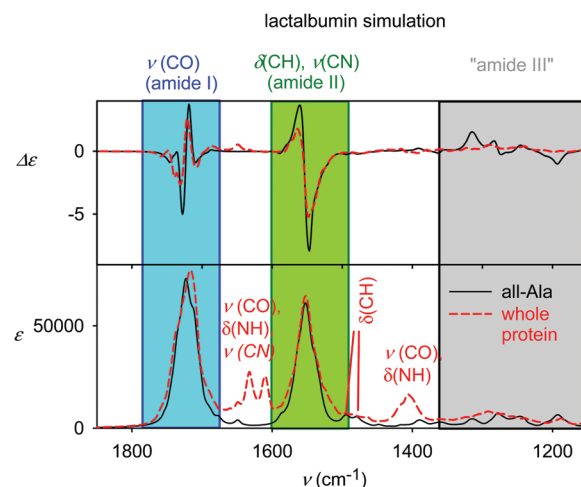


Fig. 1 VCD ( $\Delta\epsilon$ ) and absorption ( $\epsilon$ ) spectra of bovine  $\alpha$ -lactalbumin, simulated for the whole protein and an alanine analogue where all side chains were replaced by methyl groups. The backbone geometry remained unaltered.



cannot be easily resolved unless less-rigorous approaches (*e.g.*, frequency “scaling”) are applied. As discussed in previous studies, the error is mostly due to the approximate CPCM dielectric solvent model, only partially capable of mimicking hydrogen bonds between the solute and solvent (water).<sup>40–42</sup> Anharmonic interactions, the DFT error and an incomplete basis set are also part of the problem,<sup>43–45</sup> but tackling of the factors for large proteins would be prohibitively computationally expensive. Fortunately, the amide I vibration is quite isolated from the other ones, the signal is easily recognizable, and the band shape (as the best indicator of the structure) is not much affected by the frequency shift.<sup>46</sup> For lactalbumin, the “–/+” (when viewed from left to right, that is from higher to lower wavenumbers) VCD intensity pattern corresponds to its high contents of  $\alpha$ -helix, 43% according to the Protein Data Bank. This pattern has been observed in a large ensemble of  $\alpha$ -helical proteins in the past.<sup>47</sup>

The amide II band ( $\sim 1550\text{ cm}^{-1}$ ) and also a weak broad absorption band around  $1300\text{ cm}^{-1}$  conventionally attributed to “extended amide III” modes (mostly NH bending, C–N and C–C stretching coupled with backbone vibrations) are simulated at approximately the same positions as found experimentally.<sup>48,49</sup> Additional intensity due to the methyl groups (CH bending) is apparent for the all-Ala analogue around the amide III signal. The amide III region also comprises  $^2\text{C–H}$  bending, and the corresponding ROA signal was found useful to monitor the secondary structure of peptides and proteins;<sup>50,51</sup> in VCD, it plays a minor role only.<sup>48</sup>

Between the amide I and amide II signal, the whole-peptide IR spectrum exhibits two relatively strong sharp peaks at  $1630$  and  $1608\text{ cm}^{-1}$  of side chain origin, in particular due to out of phase C=O stretching of the glutamic acid  $\text{COO}^-$  group and NH bending of glutamate and arginine  $\text{NH}_2$  groups. Visualization of the normal-mode motion revealed that they partially couple with the amide II C–N stretching vibration. In experimental spectra, these bands are hidden within the amide II absorption band. However, an analogous band, at least for the C=O stretching, is observable in  $\text{D}_2\text{O}$  experiments at  $\sim 1585\text{ cm}^{-1}$  (Fig. S3, ESI†). As the groups are not chiral and relatively isolated, the corresponding VCD intensity is very weak. The same groups also give rise to the  $1400\text{ cm}^{-1}$  band.

By comparing the whole-peptide and all-alanine simulations (Fig. 1) we can see that a direct contribution of longer amino acid side chains to IR and VCD intensities is rather small. Nevertheless, their inclusion still causes up to 50% variations in amide I and amide II VCD intensities, due to the vibrational coupling. In both simulations, we can also recognize the signals of the  $\text{CH}_3$  and  $\text{CH}_2$  umbrella and scissoring CH bending modes ( $1495$  and  $1480\text{ cm}^{-1}$ ). These motions are associated with small changes of the electric dipole moment and are visible in IR spectra only because they are so numerous. Their VCD signal is negligible, which can also be seen in Fig. S4 (ESI†) where it was simulated by deleting the intensity parameters (atomic axial and atomic polar tensors) for the backbone.

On the other hand, we realize that our model is too crude to account for all aspects of the backbone–side chain interactions, such as the side chain electric field (charge state) and

conformational averaging, which would require an excessive computational time. These were found to be important for the fine structure of absorption spectra<sup>52,53</sup> and are likely to be included for more faithful VCD models in the future.

### Dependence of the spectra on protein secondary structure, comparison to experiment

Before comparing the simulated spectra to measured ones, one has to take into account both theoretical and instrumental limitations. As pointed out above, the amide I band is calculated to be about  $60\text{ cm}^{-1}$  higher than in experiment; the error for the amide II (experimentally  $\sim 1540\text{--}1550\text{ cm}^{-1}$ ) and other bands is typically a few  $\text{cm}^{-1}$  only. Although the globular proteins studied here are supposed to be relatively rigid, their solution structure might deviate from the X-ray geometry used in the simulations. The experimental bands are affected by inhomogeneous broadening, *i.e.* system dynamics and solute–solvent interactions, and the instrumental resolution ( $8\text{ cm}^{-1}$ ). The experimental intensity scale is only approximate; the real optical path length (nominally  $8\text{ }\mu\text{m}$ ) strongly depends on sample preparation, such as the force applied to keep the  $\text{CaF}_2$  windows together. Small sample volumes ( $\sim 20\text{ }\mu\text{L}$ ) and baseline subtraction also affect the experimental accuracy.

Having in mind these restrictions, we find the agreement between the theory and the experiment very good. Fig. 2 compares theoretical and experimental results for proteins of rather distinct secondary structures. These include those with a higher content of  $\alpha$ -helix, human serum albumin (46%) and myoglobin (77%), characterized by the principal “–/+” amide I couplet. While for both proteins negative sign prevails, the couplet becomes slightly more conservative for myoglobin, *i.e.* integrated intensities of the negative and positive lobes are about equal. This trend is overestimated by the simulations, providing almost an ideal couplet for myoglobin. The negative amide II signal remains approximately the same for the two “ $\alpha$ -helical” proteins, so does the weaker but still well-visible positive amide III signal.

Comparison of the  $\alpha$ -helical *versus*  $\beta$ -sheet proteins reveals the “power” of VCD spectroscopy. Indeed, unlike for IR, the spectra of the predominantly  $\beta$ -sheet concanavalin-A (46%  $\beta$ -sheet according to PDB) are quite different (Fig. 2, bottom). The helical “–/+” amide I signal changes to approximately opposite “+/-” one; however, the positive part is further split ( $1732/1715$  and  $1681/1659\text{ cm}^{-1}$  in simulation and experiment, respectively). The nature of the amide I splitting can be seen in Fig. 3 where typical vibrational modes from these two bands are plotted (other examples of normal modes can be found in Fig. S5, ESI†). The relatively regular  $\beta$ -sheet strand enables a longer-distance coupling of individual C=O stretching vibrations with distinct frequencies. The  $1732\text{ cm}^{-1}$  band is thus dominated by the “+ + +”/“– – –” patterns where three consecutive C=O bonds become longer/shorter when vibrating. The other  $1715\text{ cm}^{-1}$  calculated positive band is dominated by “+ + –” or “– – +” phasing of the amide I vibrations. The positive  $1621\text{ cm}^{-1}$  amide I signal encountered in the measured spectrum might be related to the small calculated couplet at  $\sim 1630\text{ cm}^{-1}$ . It is known that the  $\beta$ -sheet amide II VCD signal is often weaker



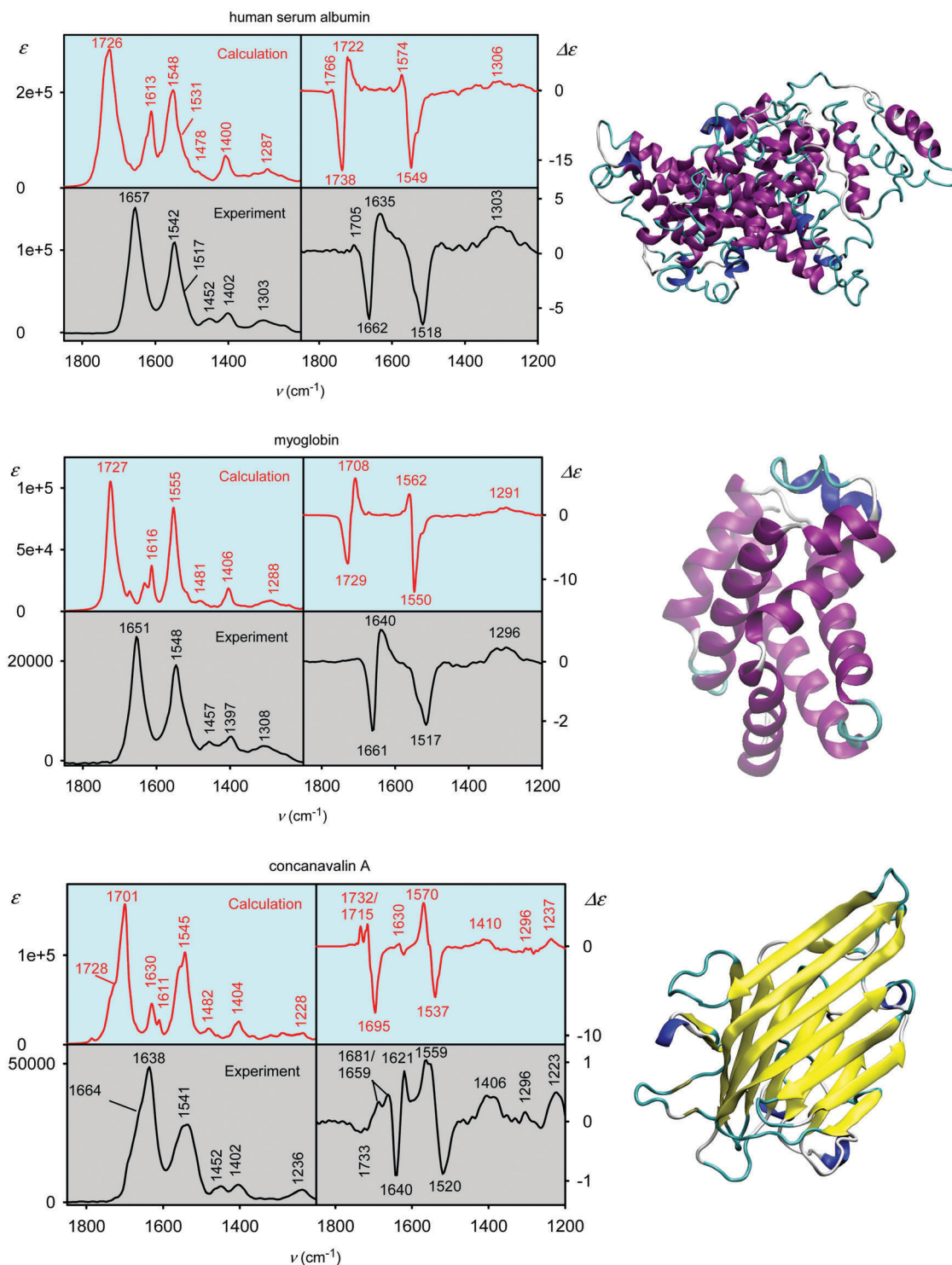


Fig. 2 Calculated and experimental IR and VCD spectra of mostly  $\alpha$ -helical (human serum albumin and equine myoglobin) and  $\beta$ -sheet (concanavalin-A) proteins.

than for  $\alpha$ -helix, or becomes a “+/-” couplet. For concanavalin, the experimental couplet at  $1559/1520\text{ cm}^{-1}$  is reproduced by the computations at  $1570/1537\text{ cm}^{-1}$ . The extended amide III region changes, too, giving rise to positive bands at  $1406$ ,  $1296$  and  $1223\text{ cm}^{-1}$ , all of which are reasonably well reproduced by theory.

The spectroscopic behavior of proteins with mixed secondary structure is more complex; the simulated and experimental spectra of bovine  $\alpha$ -lactalbumin and also hen egg-white lysozyme are shown in Fig. 4. These two proteins have similar secondary structure, and their spectra are also similar. The hen lysozyme spectrum in some details differs from that published previously,<sup>54</sup>





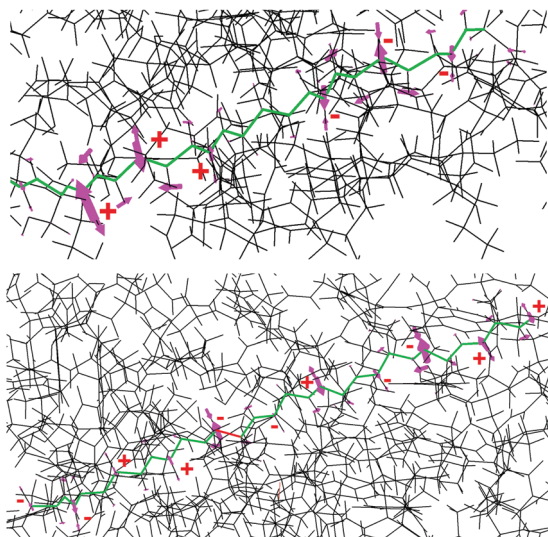


Fig. 3 Example of two amide I “ $\beta$ -sheet” modes in concanavalin. In the first mode ( $1731\text{ cm}^{-1}$ , top) the C=O bonds vibrate locally in phase (“+ + +” or “- - -”, etc.) while in the second one ( $1715\text{ cm}^{-1}$ , bottom) the typical pattern is “+ + -” or “- - +”.

most probably because of the partial aggregation at high concentrations.

Unlike for the “pure”  $\alpha$ -helical or  $\beta$ -sheet proteins discussed above, a band to band assignment within the amide I region is difficult, most probably because of the computational error stemming from the limited fragment size (Fig. S1, ESI†). In  $\alpha$ -lactalbumin, for example, the lower experimental resolution prevents the up-going simulated VCD component at  $1722\text{ cm}^{-1}$  to be discernible, while the predominantly negative signal around  $1661\text{ cm}^{-1}$  corresponds to the simulation (Fig. 4). The experimental VCD spectra of predominantly negative amide II are simulated reasonably well except for the positive  $1564\text{ cm}^{-1}$  peak not resolved in experiment, and perhaps only manifested as a steeper onset of the dominant  $1521\text{ cm}^{-1}$  negative band. The amide III vibrations provide a broad weak VCD centered around  $1300\text{ cm}^{-1}$ , also positive but weaker than for the purely  $\alpha$ -helical proteins, both in theory and experiment.

The spectra of hen egg white lysozyme (containing 41% of  $\alpha$ -helix) differ in minor features only. Instead of the positive ( $1722\text{ cm}^{-1}$ ) and negative ( $1709\text{ cm}^{-1}$ ) low wavenumber amide I VCD features in the  $\alpha$ -lactalbumin spectrum, a positive band at  $1712\text{ cm}^{-1}$  is predicted. Indeed, the band is now also detectable at

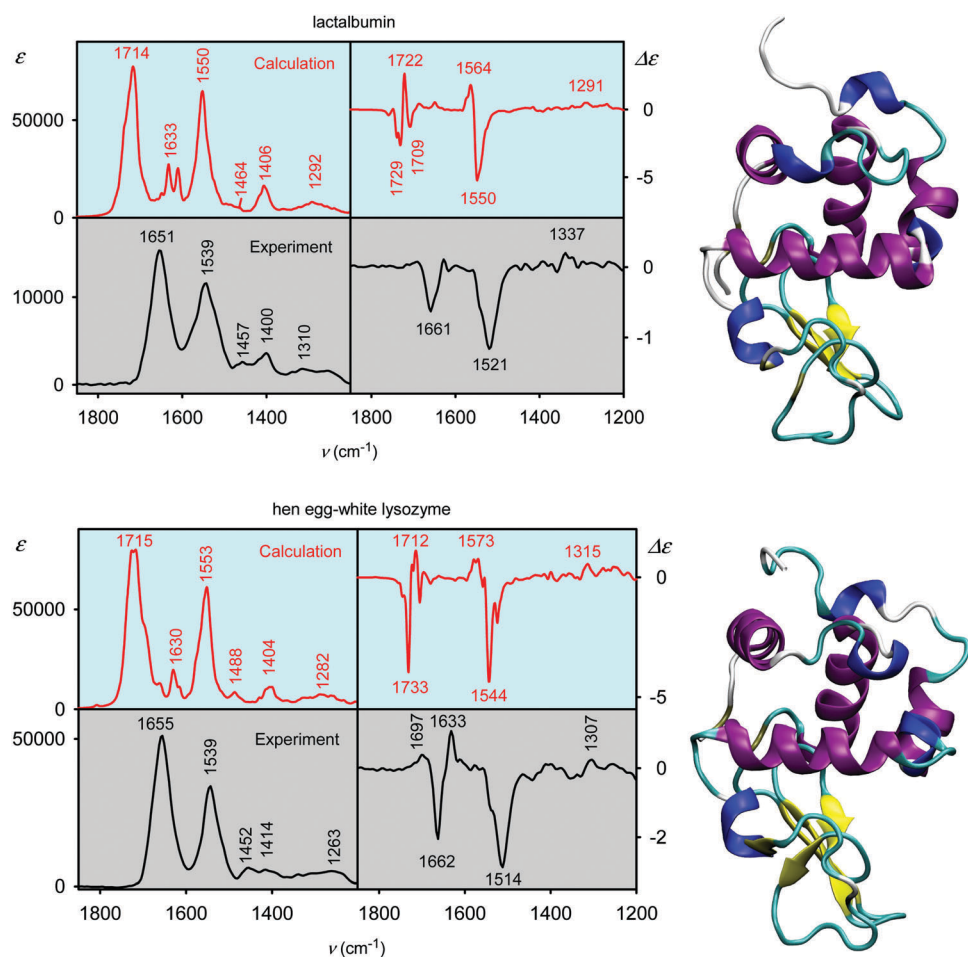


Fig. 4 Calculated and experimental IR and VCD spectra of proteins with mixed secondary structure, bovine  $\alpha$ -lactalbumin and hen egg-white lysozyme.



1633  $\text{cm}^{-1}$  in experiment. The small amide I positive experimental band at 1697  $\text{cm}^{-1}$  is not reproduced.

### Amide I signal of deuterated proteins

Due to the high water absorption at the amide I region, IR and VCD protein measurements are often performed in deuterated samples, in  $\text{D}_2\text{O}$ . The deuteration, however, leads to medium to large variations of the amide I signal (often referred to as amide I' in  $\text{D}_2\text{O}$ );<sup>55</sup> these represent a complicated product of solvent effects and protein dynamics.<sup>40,41,56</sup> The static DFT computation cannot fully explain this effect, it can nevertheless estimate the contribution of the  $\text{NH} \rightarrow \text{ND}$  exchange stemming from vibrational coupling within the protein molecule. In Fig. 5 experimental IR and VCD  $\alpha$ -lactalbumin amide I spectra are compared to the simulations. While deuteration shifts the absorption maximum by just 9  $\text{cm}^{-1}$  to lower wavenumbers, the VCD pattern changes more dramatically. The main negative component at 1659  $\text{cm}^{-1}$  loses intensity by  $\sim 40\%$ , shifts right, and splits. The 1628  $\text{cm}^{-1}$  positive band vanishes and a new negative signal at 1616  $\text{cm}^{-1}$  appears. The simulation faithfully reproduces these changes, although the precision is clearly limited in providing a perfect fit. A normal mode analysis (Fig. 6) gives a qualitative insight into the deuteration by revealing that while  $\text{C}=\text{O}$  stretching is dominant, the amide I vibrations also involve NH bending significantly affected by the isotopic substitution.

### Explicit solvent model included in the calculations

The effect of “explicit” solvent on the spectra was estimated for the smallest molecule investigated, the insulin dimer. Water molecules closer than 3.6 Å to the protein were kept with the fragments, so that the solvent–solute hydrogen bonds could be included at the same computational level as the protein parts. By comparing the spectra with those obtained using the default CPCM dielectric (“implicit”) environment (Fig. 7) it is obvious that the “explicit” solvent not only conserves the principal spectral features, but makes the simulations more realistic.

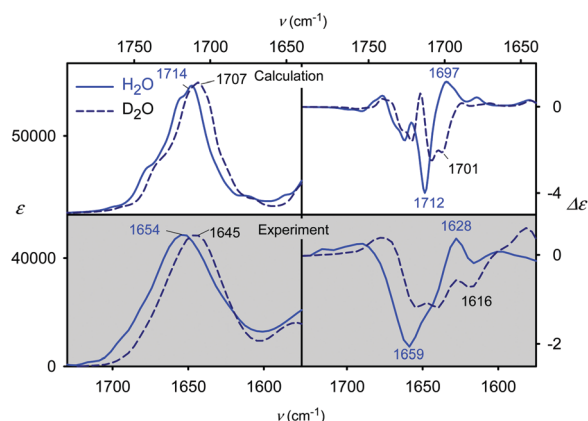


Fig. 5 Simulated and experimental amide I IR and VCD spectra of  $\alpha$ -lactalbumin, for deuterated and natural environment. The experimental spectra were normalized to absorption maxima, and the 5-amide based simulation was used in this case, instead of the default 4-amide fragmentation.

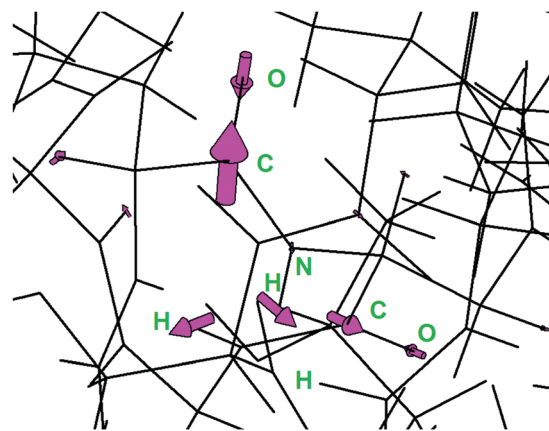


Fig. 6 An amide I mode of  $\alpha$ -lactalbumin; visualization of the normal mode movement reveals mutual coupling of  $\text{C}=\text{O}$  stretching vibrations, but also the involvement of  $\text{N-H}$  bending.

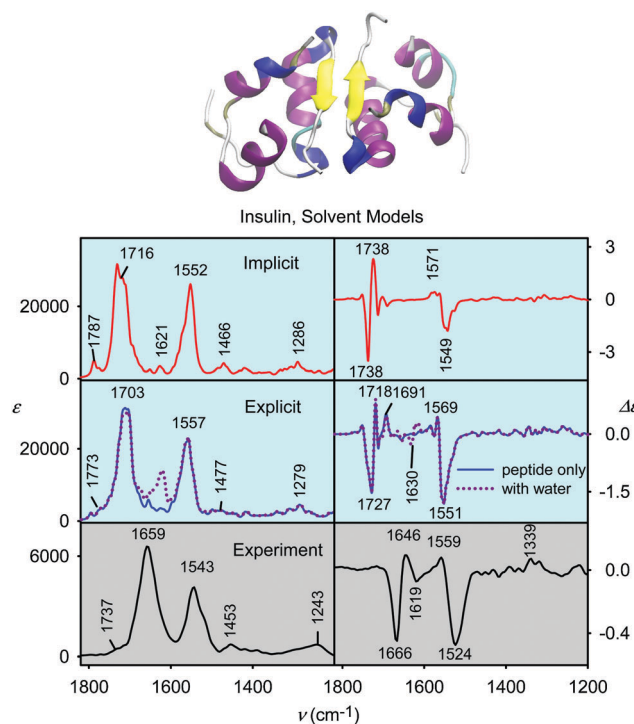


Fig. 7 Insulin, its absorption and VCD spectra simulated with an implicit solvent environment (CPCM) and when water molecules from the first hydration shell were retained. For the “explicit” case, the spectra were simulated with and without water intensity tensors while vibrational frequencies remained the same. The experiment is at the bottom.

The IR  $\text{C}=\text{O}$  stretching signal (originating mostly from the  $\text{COOH}$  glutamic groups) simulated at 1787  $\text{cm}^{-1}$  within CPCM shifts to lower wavenumbers and becomes broader, which better corresponds to the experimental shoulder at 1737  $\text{cm}^{-1}$ . The central amide I CPCM band ( $\sim 1716 \text{ cm}^{-1}$ ) also becomes smoother and shifts by 13  $\text{cm}^{-1}$  closer to the experimental value. Changes in other spectral parts are almost negligible. Interestingly, the amide II band goes slightly up in frequency (1552  $\rightarrow$  1557  $\text{cm}^{-1}$ ) when the explicit model is used. This is



farther from the experimental band ( $1543\text{ cm}^{-1}$ ), most probably due to the DFT error.

Even greater changes are observed in VCD spectra. The negative  $1738\text{ cm}^{-1}$  amide I band (CPCM) becomes broader, with a similar intensity as the amide II band ( $1551\text{ cm}^{-1}$ ), and its position shifts to  $1727\text{ cm}^{-1}$ , closer to the experimental one. On the other hand, the explicit solvent model still fails to reproduce the fine vibrational band splitting in the amide I region. While the positive band at  $1738\text{ cm}^{-1}$  splits into two ( $1718$  and  $1691\text{ cm}^{-1}$ ), only a single positive band is detected experimentally at  $1646\text{ cm}^{-1}$ . Rather minor changes occur at the amide II and lower-wavenumber spectral parts. Interestingly, when the water contribution to spectral intensities is included (dashed line in Fig. 7) some VCD intensity arises around  $1630\text{ cm}^{-1}$ . This “transfer” of chirality from the protein to water could explain the negative feature around  $1619\text{ cm}^{-1}$  in the experimental spectrum. Visualization of the normal mode displacement suggests that amide carbonyls making two (“bifurcated”) hydrogen bonds and also  $\text{NH}_2$  group bending vibrations of asparagines and glutamines contribute to it.

The explicit consideration of water molecules can thus account for some inconsistencies in the simulated spectral shapes; similar conclusions have been drawn also for nucleic acids.<sup>57,58</sup> Solvent molecules should be certainly included in future modeling of protein spectra, which is currently very demanding for computer resources. Typically, the first solvation shell is kept, and the geometry is based on the classical MD force field. In addition to the larger size of such clusters that need to be treated quantum-chemically, water/solvent molecules can contribute to the chiral signal. Therefore, the spectra of many MD snapshot geometries need to be averaged so that the solvent signal becomes realistic.<sup>59</sup> On the other hand, the sensitivity of the VCD spectra to the solvent–solute interactions makes them an excellent tool to study biologically relevant molecules and their interactions.

### Variation of the $\alpha$ -helical VCD signal

Although  $\alpha$ -helical VCD “signatures” are very characteristic in various proteins<sup>49</sup> one may ask if minor variations of  $\alpha$ -helical geometry caused by different amino acid sequences or interactions with the environment can distort them. We thus separately simulated the spectra for eight helical myoglobin segments. Their geometries were kept unchanged, without any treatment of the broken bonds, and the same spectral parameters (force field, atomic polar and axial tensors) as for the full protein were used. As apparent from Fig. 8 the main VCD sign-pattern is conserved in all the helices; the minor variations are thus unlikely to cause a major error in quantitative structural predictions.<sup>47,60</sup> On the other hand, spectral details do vary, including different widths and positions of the absorption peak. Interestingly, the shortest  $\alpha$ -helix (number 3 in the figure) has a very distinct pattern in the amide II region ( $\sim 1550\text{ cm}^{-1}$ ), implying that this structural motif is likely to be more easily identified by VCD spectroscopy. Because the torsion angles of the third helix ( $(\phi, \psi) \sim (-62^\circ, -30^\circ)$ ) do not significantly differ from those found in the other helices and are close to standard values ( $-57^\circ, -47^\circ$ ),<sup>61</sup> we explain the specific signal of the shortest helix by its length and

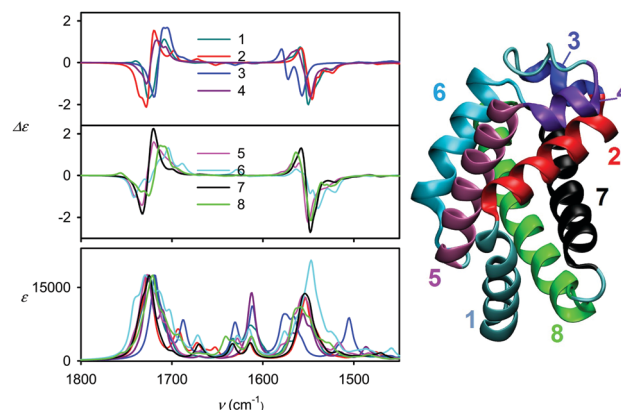


Fig. 8 Simulated VCD and absorption spectra for individual helical myoglobin parts. The spectra were normalized to the highest amide I absorption ( $\sim 1724\text{ cm}^{-1}$ ).

strong “end-effects” caused by the absence of vibrational coupling to other protein parts.

### Additivity of the spectra

Another implication of stronger dependence of VCD than IR on a long-range vibrational coupling is different additivity of the spectra. The spectra of deuterated  $\alpha$ -lactalbumin were simulated for the whole molecule and the  $\alpha$ -helical and  $\beta$ -sheet protein parts (Fig. S6, ESI†). The sum of  $\alpha$ -helical and  $\beta$ -sheet IR spectra add up to approximately 70% of total intensity, with the rest coming from the connecting loop and “unordered” peptide chains. For VCD, the situation is more complex. Within the amide II’ region (around  $1470\text{ cm}^{-1}$ ) the spectrum of the whole protein is nearly the same as the  $\alpha$ -helical one, *i.e.*, contributions of the  $\beta$ -sheet and unordered parts are negligible. This explains the similar bandshape of the amide II VCD signal in all proteins containing a substantial amount of  $\alpha$ -helix but variable amounts of other structures (myoglobin, HSA, lysozyme and lactalbumin, *cf.* Fig. 2 and 4). The higher-frequency component of the amide I’ signal ( $\sim 1730\text{ cm}^{-1}$ ) behaves in a similar way. On the other hand, the sum of the  $\alpha$ -helical and  $\beta$ -sheet contributions accounts for only  $\sim 40\%$  intensity of the lowest frequency amide I’ band ( $\sim 1705\text{ cm}^{-1}$ ). This is in agreement with the higher sensitivity of these vibrations to deuteration as mentioned above. Apparently, the lower-frequency amide I and amide I’ modes are “less local”, *i.e.* they are better coupled to other vibrations with similar frequency, even if they occur in different protein parts.

This is also consistent with experimental experience (mostly  $\text{D}_2\text{O}$  measurements). The higher-frequency amide I lobe comes mainly from  $\alpha$ -helix, while the lower-frequency one from a combination of random coil and  $\beta$ -sheet. The random coil signal is usually stronger than that from  $\beta$ -sheet, approximately in the calculated ratio 60/40.<sup>62</sup>

### Electric-dipolar origin of protein chirality and simpler models

For any molecule, both the “irreducible” magnetic (local parts of the atomic axial tensors, AAT) and electric dipolar (atomic polar tensors, APT) parts contribute to VCD intensities.<sup>11,63</sup>





In the past, neglecting AAT was suggested as a fast way to approximately simulate the spectra, but in general both contributions are needed for the results to be reliable.<sup>64,65</sup> In this context, it is interesting to note that except for some minor discrepancies the “APT-only” model nearly matches the exact results for proteins (the two approaches are compared in Fig. S7 (ESI†) for  $\alpha$ -lactalbumin). In other words, protein VCD is dominated by electric dipole interactions. The transition electric dipole moments are mostly localized on the amide groups that are nearly planar, *i.e.* not locally chiral. However, for an accurate simulation also containing features such as the low-wavenumber amide I VCD component ( $1709\text{ cm}^{-1}$ ) in  $\alpha$ -lactalbumin, it is necessary to include the contributions from the AAT part as well.

Another simplistic approach of simulating protein VCD used often in the past was the transition dipole coupling (TDC).<sup>66–68</sup> The amide groups (quite often only the C=O stretching) are then replaced by interacting dipoles.<sup>34,69–71</sup> TDC results are thus very dependent on input parameters; they well-capture the general dependence of the spectra on secondary structure, but fail to reproduce important details. As an example of the performance, we present IR and VCD  $\alpha$ -lactalbumin spectra calculated by the TDC method in Fig. S8 (ESI†). Electric transition dipoles (vectors and vibrational frequencies) were obtained for the amide I and II modes from a DFT calculation on *N*-methyl-acetylamine, using the same approximation level as for CCT. The TDC frequency-error is thus similar as for the all-atomic computation. For IR, the amide II band splitting is too large based on TDC, otherwise the spectral profile appears reasonable if compared to the experiment. The TDC VCD pattern is by definition conservative, *i.e.* integrated intensity over the whole spectral range is zero, which is in contrast to the predominantly negative experimental and CCT results (Fig. 3). Nevertheless the amide I TDC pattern exhibits some resemblance to the CCT one, and even more obvious is the correspondence for amide II, where both approaches predict the “+/-” couplet. We can thus interpret this similarity as confirmation of the importance of dipole-dipole interactions for protein VCD spectral pattern.

### Protein tertiary structure and long-distance interactions

Finally, we compare the spectra of structurally very similar hen egg white and human (recombinant) lysozyme (Fig. 9). Structural differences may be reflected in tiny variations of protein tertiary structure, flexibility, interactions mediated through amino acid side chains, protein response to the environment, *etc.* Whereas the absorption spectra are almost identical, larger variations are observable in VCD. This is rather surprising for the low-resolution technique, although some differences were previously observed in another form of vibrational optical activity, ROA,<sup>72</sup> and confirmed by the simulations as well.<sup>18</sup> While both lysozymes from different organisms exhibit a couplet (“+/-”) amide I VCD signal, the hen egg white lysozyme has a much smaller negative lobe at  $1662\text{ cm}^{-1}$ , and a third peak appears at  $1633\text{ cm}^{-1}$ . The shape of the amide II signal ( $\sim 1510\text{ cm}^{-1}$ ) remains about the same, but the band is narrowed and slightly shifted to a lower wavenumber.

The simulations reproduce these observations reasonably well, but only when the TDC long-range interaction correction

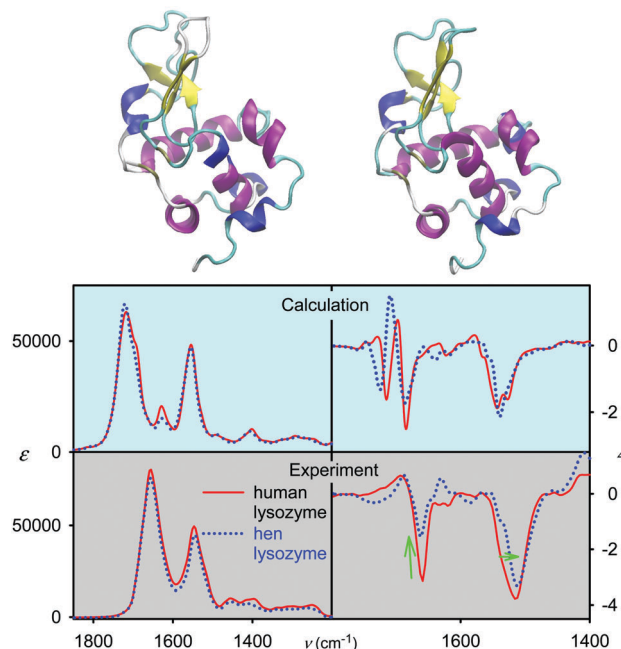


Fig. 9 IR and VCD spectra of human and hen egg white lysozyme, simulation (with TDC correction applied) and experiment.

is added to the DFT force field (Fig. 9). Without TDC, predicted differences between hen and human lysozyme VCD spectra are much smaller (Fig. S9, ESI†). Supposedly, the long-range coupling of the protein vibrational modes can thus explain part of the variations. We do realize that the X-ray geometries may differ from those in solution and the experimental spectra may depend on sample preparation,<sup>54</sup> which may further modulate the comparison. Nevertheless, measurements in  $D_2O$  where the signal to noise ratio in the amide I region is larger confirmed the results for non-deuterated proteins. In Fig. S10 (ESI†) the results of two independent measurements for each protein are plotted; the variations in the VCD shapes are clearly larger than the measurement error. Although the differences are not big, the high sensitivity of VCD spectroscopy to minute structural mutations is very appealing for further applications.

The importance of the longer-range vibrational interactions is also documented for concanavalin A (Fig. S11, ESI†). The TDC correction is gradually added to the DFT force field by changing the dielectric constant ( $\epsilon_r$ ). Some spectral features are improved by that, such as the split positive amide II band ( $1559\text{ cm}^{-1}$  in experiment) or a negative amide I high-frequency signal ( $1733\text{ cm}^{-1}$ ). In both IR and VCD, the amide I band broadens and shifts down in frequency, *i.e.* closer to experiment. On the other hand, it still deviates from the experiment in some features. Also, the excessive band splitting for  $\epsilon_r = 1$  suggests that the vacuum value of the permittivity constant in the TDC approximation overestimates the interactions of distant protein chromophores.

## 4. Conclusions

The CCT fragment approach and density functional computations helped us to significantly clarify the relation between vibrational





circular dichroism spectra of proteins and their structure. A very satisfactory agreement with the experiment has been achieved. The simulations explained the main differences observed among proteins differing in secondary structure, provided an insight into the nature of the underlying normal mode vibrations, made it possible to assign all principal vibrational bands, and could partially explain fine spectral changes due to deuteration. The results indicate that the high- and low-frequency amide I vibrations differ in their sensitivity to coupling with distant vibrations. To the best of authors' knowledge, this has not been previously addressed, and may be important to explain fine changes in spectral shapes caused by protein tertiary structure. Surprisingly, human and hen lysozymes provided rather different VCD spectra; this was explicable by long-range coupling of vibrational modes, and somewhat changes the concept of locality typically attributed to this kind of spectroscopy. Clearly, the combination of vibrational optical activity spectroscopy and computational chemistry appears to be quite useful for protein studies, and can prevent erroneous interpretation of experimental data. On the other hand, the accuracy of the simulations still needs to be improved, namely by better describing the solvent and dynamics effects.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by the Grant Agency (16-05935S to PB and 16-04902S to VA), the Technology Agency (TE01020229) and the Ministry of Education (EU COST CA15214/LTC17012 and OPVVV CZ.02.1.01/0.0/0.0/16\_019/0000729).

## References

- 1 T. B. Freedman, X. Cao, R. K. Dukor and L. A. Nafie, *Chirality*, 2003, **15**, 743–758.
- 2 K. H. Hopmann, J. Šebestík, J. Novotná, W. Stensen, M. Urbanová, J. Svenson, J. S. Svendsen, P. Bouř and K. Ruud, *J. Org. Chem.*, 2012, **77**, 858–869.
- 3 L. Nafie, *Vibrational optical activity: Principles and applications*, Wiley, Chichester, 2011.
- 4 T. A. Keiderling, in *Infrared and Raman spectroscopy of biological materials*, ed. B. Yan, H. U. Gremlich and M. Dekker, New York, 1999.
- 5 M. Gulotta, D. J. Goss and M. Diem, *Biopolymers*, 1998, **28**, 2047–2058.
- 6 V. Maharaj, A. Rauk, J. H. van de Sande and H. Wieser, *J. Mol. Struct.*, 1997, **408**, 315–318.
- 7 V. Andrushchenko, D. Tsankov and H. Wieser, *J. Mol. Struct.*, 2003, **661–662**, 541–560.
- 8 G. Holzwarth, E. C. Hsu, H. S. Mosher, T. R. Faulkner and A. Moscowitz, *J. Am. Chem. Soc.*, 1974, **96**, 251–252.
- 9 L. A. Nafie and T. B. Freedman, *J. Chem. Phys.*, 1983, **78**, 7108–7116.
- 10 P. J. Stephens, *J. Phys. Chem.*, 1985, **89**, 748–752.
- 11 P. J. Stephens, *J. Phys. Chem.*, 1987, **91**, 1712–1715.
- 12 J. R. Cheeseman, M. J. Frisch, F. J. Devlin and P. J. Stephens, *Chem. Phys. Lett.*, 1996, **252**, 211–220.
- 13 K. L. Bak, F. J. Devlin, C. S. Ashvar, P. R. Taylor, M. J. Frisch and P. J. Stephens, *J. Phys. Chem.*, 1995, **99**, 14918–14922.
- 14 K. V. J. Jose, D. Beckett and K. Raghavachari, *J. Chem. Theory Comput.*, 2015, **11**, 4238–4247.
- 15 P. Bouř, J. Sopková, L. Bednářová, P. Maloň and T. A. Keiderling, *J. Comput. Chem.*, 1997, **18**, 646–659.
- 16 S. Yamamoto, X. Li, K. Ruud and P. Bouř, *J. Chem. Theory Comput.*, 2012, **8**, 977–985.
- 17 N. S. Bieler, M. P. Haag, C. R. Jacob and M. Reiher, *J. Chem. Theory Comput.*, 2011, **7**, 1867–1881.
- 18 J. Kessler, J. Kapitán and P. Bouř, *J. Phys. Chem. Lett.*, 2015, **6**, 3314–3319.
- 19 A. Litwińczuk, S. R. Ryu, L. A. Nafie, J. W. Lee, H. I. Kim, Y. M. Jung and B. Czarnik-Matusewicz, *Biochim. Biophys. Acta*, 2014, **1844**, 593–606.
- 20 D. Kuroski, X. Lu, L. Popova, W. Wan, M. Shanmugasundaram, G. Stubbs, R. K. Dukor, I. K. Lednev and L. A. Nafie, *J. Am. Chem. Soc.*, 2014, **136**, 2302–2312.
- 21 V. Andrushchenko, H. Wieser and P. Bouř, *J. Phys. Chem. B*, 2002, **106**, 12623–12634.
- 22 H. Sato, T. Yajima and A. Yamagishi, *Phys. Chem. Chem. Phys.*, 2018, DOI: 10.1039/C7CP06264H.
- 23 N. N. Kal'nin and S. Y. Ven'yaminov, *J. Appl. Spectrosc.*, 1988, **49**, 1028–1032.
- 24 P. Bouř and T. A. Keiderling, *J. Chem. Phys.*, 2002, **117**, 4126–4132.
- 25 P. Bouř, *Collect. Czech. Chem. Commun.*, 2005, **70**, 1315–1340.
- 26 J. Hudecová, K. H. Hopmann and P. Bouř, *J. Phys. Chem. B*, 2012, **116**, 336–342.
- 27 J. P. Perdew, K. Burke and Y. Wang, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 16533–16539.
- 28 A. Klamt, V. Jonas, T. Burger and J. C. W. Lohrenz, *J. Phys. Chem. A*, 1998, **102**, 5074–5085.
- 29 B. Mennucci, C. Cappelli, R. Cammi and J. Tomasi, *Chirality*, 2011, **23**, 717–729.
- 30 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich,



- A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, Gaussian, Inc., Wallingford CT, USA, 2009.
- 31 S. Yamamoto, J. Kaminský and P. Bouř, *Anal. Chem.*, 2012, **84**, 2440–2451.
  - 32 F. J. Devlin, P. J. Stephens, J. R. Cheeseman and M. J. Frisch, *J. Phys. Chem.*, 1997, **101**, 9912–9924.
  - 33 L. D. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, UK, 2004.
  - 34 J. Kubelka, J. Kim, P. Bouř and T. A. Keiderling, *Vib. Spectrosc.*, 2006, **42**, 63–73.
  - 35 R. D. Amos, K. J. Jalkanen and P. J. Stephens, *J. Phys. Chem.*, 1988, **92**, 5571.
  - 36 P. L. Polavarapu, *J. Phys. Chem.*, 1990, **94**, 8106–8112.
  - 37 T. A. Keiderling, *Curr. Opin. Chem. Biol.*, 2002, **6**, 682–688.
  - 38 A. Barth and C. Zscherp, *Q. Rev. Biophys.*, 2002, **35**, 369–430.
  - 39 T. A. Keiderling, B. Wang, M. Urbanová, P. Pančoška and R. K. Dukor, *Faraday Discuss.*, 1994, **99**, 263–286.
  - 40 J. Kubelka, R. Huang and T. A. Keiderling, *J. Phys. Chem. B*, 2005, **109**, 8231–8243.
  - 41 P. Bouř, D. Michalík and J. Kapitán, *J. Chem. Phys.*, 2005, **122**, 144501.
  - 42 K. K. Lee, S. Hahn, K. I. Oh, J. S. Choi, C. Joo, H. Lee, H. Han and M. Cho, *J. Phys. Chem. B*, 2006, **110**, 18834–18843.
  - 43 Z. Ganim, H. S. Chung, A. W. Smith, L. P. Deflores, K. C. Jones and A. Tokmakoff, *Acc. Chem. Res.*, 2008, **41**, 432–441.
  - 44 L. P. Deflores, Z. Ganim, S. F. Ackley, H. S. Chung and A. Tokmakoff, *J. Phys. Chem. B*, 2006, **110**, 18973–18980.
  - 45 V. Andrushchenko, P. Matějka, D. T. Anderson, J. Kaminský, J. Horníček, L. O. Paulson and P. Bouř, *J. Phys. Chem. A*, 2009, **113**, 9727–9736.
  - 46 R. Huang, J. Kubelka, W. Barber-Armstrong, R. A. G. D. Silva, S. M. Decatur and T. A. Keiderling, *J. Am. Chem. Soc.*, 2004, **126**, 2346–2354.
  - 47 V. Baumruk, P. Pančoška and T. A. Keiderling, *J. Mol. Biol.*, 1996, **259**, 774–791.
  - 48 B. I. Baello, P. Pančoška and T. A. Keiderling, *Anal. Biochem.*, 1997, **250**, 212–221.
  - 49 M. Urbanová, T. A. Keiderling and P. Pančoška, *Bioelectrochem. Bioenerg.*, 1996, **41**, 77–80.
  - 50 L. D. Barron, F. Zhu, L. Hecht, G. E. Tranter and N. W. Isaacs, *J. Mol. Struct.*, 2007, **834–836**, 7–16.
  - 51 S. Yamamoto, T. Furukawa, P. Bouř and Y. Ozaki, *J. Phys. Chem. A*, 2014, **118**, 3655–3662.
  - 52 B. A. Anderson, G. Kubelka and J. Kubelka, *Vib. Spectrosc.*, 2015, **77**, 40–50.
  - 53 B. A. Anderson, A. Literati, B. Ball and J. Kubelka, *Biopolymers*, 2014, **101**, 536–548.
  - 54 S. Ma, T. B. Freedman, R. Dukor and L. Nafie, *Appl. Spectrosc.*, 2010, **64**, 615–626.
  - 55 V. Baumruk and T. A. Keiderling, *J. Am. Chem. Soc.*, 1993, **115**, 6939–6942.
  - 56 K. Kwac, K. K. Lee, J. Han, K. I. Oh and M. Cho, *J. Chem. Phys.*, 2008, **128**, 105106.
  - 57 V. Andrushchenko, L. Benda, O. Páv, M. Dračinský and P. Bouř, *J. Phys. Chem. B*, 2015, **119**, 10682–10692.
  - 58 V. Andrushchenko, D. Tsankov, M. Krasteva, H. Wieser and P. Bouř, *J. Am. Chem. Soc.*, 2011, **133**, 15055–15064.
  - 59 K. H. Hopmann, K. Ruud, M. Pecul, A. Kudelski, M. Dračinský and P. Bouř, *J. Phys. Chem. B*, 2011, **115**, 4128–4137.
  - 60 T. A. Keiderling, P. Pančoška, V. Baumruk, M. Urbanová, V. P. Gupta, R. K. Dukor and D. Huo, in *Molecular Modeling*, ed. T. F. Kumosinski and M. N. Leibman, American Chemical Society, Washington, D.C., 1994, pp. 61–70.
  - 61 T. E. Creighton, *Proteins: Structures and Molecular Properties*, W. H. Freeman and Co., New York, 2nd edn, 1993.
  - 62 A. M. Polyanchko, V. Andrushchenko, P. Bouř and H. Wieser, in *Circular Dichroism: Theory and Spectroscopy*, ed. D. S. Rodgers, Nova, Hauppauge, NY, 2011.
  - 63 P. Bouř, J. McCann and H. Wieser, *J. Chem. Phys.*, 1998, **108**, 8782–8789.
  - 64 V. P. Nicu, *Phys. Chem. Chem. Phys.*, 2016, **18**, 21202–21212.
  - 65 T. B. Freedman and L. Nafie, *J. Phys. Chem.*, 1984, **88**, 496–500.
  - 66 I. Tinoco, *Radiat. Res.*, 1963, **20**, 133–139.
  - 67 G. Holzwarth and I. Chabay, *J. Chem. Phys.*, 1972, **57**, 1632–1635.
  - 68 P. Bouř and T. A. Keiderling, *J. Am. Chem. Soc.*, 1992, **114**, 9100–9105.
  - 69 J. Snir, R. A. Frankel and J. A. Schellman, *Biopolymers*, 1975, **14**, 173–196.
  - 70 W. H. Moore and S. Krimm, *Proc. Natl. Acad. Sci. U. S. A.*, 1975, **72**, 4933–4935.
  - 71 J. Průša and P. Bouř, *Chirality*, 2018, **30**, 55–64.
  - 72 I. H. McColl, E. W. Blanch, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 2004, **126**, 8181–8188.

