Charles University

Faculty of Science

Departement of Physical and Macromolecular Chemistry

Modelling of Chemical Properties of Nano- and Biostructures

DEVELOPMENT AND APPLICATIONS OF MOLECULAR DYNAMICS FOR MOLECULAR SPECTROSCOPY

Vývoj a aplikace molekulové dynamiky pro molekulovou spektroskopii

Doctoral Thesis

Mgr. Jiří Kessler

Supervisor: prof. RNDr. Petr Bouř, DSc.

Prague 2017

Poděkování

V prvé řadě bych chtěl poděkovat prof. RNDr. Petru Bouřovi, DSc. za nezměrnou podporu a vedení mé disertační práce, dr. Josefu Kapitánovi za změření ROA spekter proteinů a Valerymu Andrushchenkovi za změření jejich VCD spekter. Dále bych chtěl poděkovat Shigeki Yamamotovi, že jsem se mohl podílet na zajímavé práci v oblasti ROA insulinových fibril, a Václavu Parchaňskému, díky němuž jsem se ocitl v naší skupině. V neposlední řadě bych chtěl poděkovat všem v naší skupině za vytváření příjemné pracovní atmosféry a své rodině za podporu během předchozích a tohoto studia. Za nezměrnou podporu bych chtěl velmi poděkovat své přítelkyni Heleně Houškové, že mi nepřetržitě dodávala sílu k sepsání této práce.

Souhrn

Tato práce se zabývá simulacemi chiroptických spekter pomocí kombinace molekulárně dynamických a kvantově chemických výpočtů. Molekulová dynamika je použita ke zkoumání konformačního chování studovaných systémů (převážně proteinů), kvantová chemie pro výpočet jejich spektrálních vlastností. Výpočetně náročné kvantově chemické metody jsou však omezeny pouze na relativně malé systémy. My jsme překonali tento problém zejména pomocí fragmentace systémů na sadu menších, výpočetně zvládnutelných částí. Tyto fragmenty jsou pak použity pro výpočet spektrálních vlastností, které jsou následně přeneseny zpět na původní molekulu.

Studována jsou spektra vibrační optické aktivity (VOA) proteinových systémů (fibrily poly-L-glutamové (PLGA) kyseliny, prefibrilární formy insulinu a globulární proteiny v nativním stavu). Simulovaná spektra většinou uspokojivě souhlasila s experimentem a byla použita k jeho interpretaci. V případě spekter Vibračně Cirkulárního Dichroismu (VCD) poly-L-glutamové kyseliny simulovaná jen spektra kvalitativně reprodukovala experiment. Ve výpočtech jsme byli např. schopni reprodukovat silný pás v oblasti amidu I a slabší negativní pás patřící karboxylové skupině postraního řetězce.

Podobná výpočetní procedura byla pak použita na soubor vybraných globulárních proteinů. Jejich spektra Ramanovy optické aktivity (ROA) poskytla uspokojivou přesnost a simulovaná spektra mohla být použita k interpretaci experimentálních výsledků. Byli jsme schopni reprodukovat experimentální rozdíly mezi převážně α-helikálním lidským sérovým albuminem a concanavalinem A obsahujícím převážně struktury β-skládaného listu, a nebo mezi velmi podobným lidským a slepičím lysozymem. V případě insulinových fibril jsme zjistili, že ROA technika je velmi citilivá ke konformačním změnám proteinů a ze spekter jsme byli schopni extrahovat informace o molekulové struktuře.

Dále byly provedeny dvě studie v oblasti spektroskopie elektronového cirkulárního dichroismu (ECD) a cirkulárně polarizované luminiscence (CPL). První byla zaměřena na to, jak kovové ionty ovlivňují ECD spektrum jejich komplexu s Monensinem. Druhá se zabývala CPL europiového komplexu indukovanou interakcí s aminokyselinou.

Součástí práce jsou také metodické projekty zahrnující implementaci helikálních periodických podmínek v molekulární dynamice a přenos frekvenčně závislé polarizovatelnosti pro výpočet UV-vis a ECD spekter velkých systémů.

Klíčová slova: molekulární dynamika, simulace spekter, kvantová chemie, chiralita, optická aktivita

Summary

This Thesis deals with simulations of chiroptical spectra using a combination of molecular dynamics and quantum chemistry. Molecular dynamics was used to explore conformational behaviour of studied systems (proteins), quantum chemistry for calculation of spectral properties. The Quantum chemical methods are limited to relatively small systems. We overcome this problem mostly by a fragmentation of studied systems, when smaller, computationally feasible, fragments are created and used for the quantum chemical calculations. Calculated properties were then transferred to the big molecule.

Vibrational Optical Activity (VOA) spectra of poly-L-glutamic acid fibrils (PLGA), insulin prefibrillar form and native globular proteins were studied. The simulated spectra provided satisfactory agreement with the experiment and were used for its interpretation. Experimental Vibrational Circular Dichroism (VCD) spectra of poly-L-glutamic acid fibrils were only qualitatively reproduced by the simulation. We could reproduce the major amide I band and a smaller negative band associated with the side chain carboxyl group.

Our simulation procedure was then extended to a set of globular proteins and their Raman Optical Activity (ROA) spectra. Here we achieved an exceptional precision. For example, we were able to reproduce the main experimental differences between α-helical human serum albumin and concanavalin A containing mainly β-sheets, or between very similar human and hen egg-white lysozymes. In case of insulin fibrils, we found how the ROA technigue is sensitive to protein coformational changes. By using our simulation, we were able to extract the information of molecular structure and flexibility from the spectra.

Chiral systems were also studied by Electronic Circular Dichroism (ECD) and Circularly Polarized Luminiscence (CPL) spectroscopies. Using ECD we studied how the metal ion influence ECD spectra of respective metal-monensin complexes. The second project dealt with CPL of racemic europium complex induced by interaction with aminoacids.

The Thesis also contains methodical projects dealing with implementation of helical periodic boundary conditions in molecular dynamics and transfer of frequency-dependent polarizabilities for calculation of UV-vis and ECD spectra of larger systems.

Keywords: Molecular Dynamic, Spectra Simulations, Quantum Chemistry, Chirality, Optical Activity

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne .....

.....

Jiří Kessler

Contents

1 Introduction

Molecular spectroscopy includes a wide range of techniques that can be helpful for structural studies in chemistry and biology. For example, vibrational spectroscopy is advantageous for protein studies even though it cannot provide detailed information about 3D structure. It can be used to study dynamic structural changes. The time scale of vibrational transitions is shorter than for conformational changes and the vibrational spectrum thus provide an average of all present conformations. As an example of applications, we can study enzyme-substrate interactions or protein folding.

As proteins are chiral molecules, they can be studied using technigues of Vibrational Optical Activity (VOA) including Vibrational Circular Dichroism (VCD) and Raman Optical Activity (ROA) spectroscopies. Their advantage is that they exhibits extreme sensitivity to protein structural variations [1, 2]. It is possible to clearly distinguish proteins with the same primary but different secondary structure, because the secondary structure elements such as α-helices or β-sheets have very characteristic spectral patterns.

VCD provides mainly information about protein backbone, and its experimental frequency range is relatively narrow (usually 900-2000 cm^{-1}). On the other hand, ROA is also sensitive to sidechain groups and the experimental range is wider than in case of VCD (usually 100-2400 cm^{-1}).

In general, VOA spectra are very complicated. For their interpretation, we use the simulations to assign signals to appropriate vibrations and to make a link between geometry and spectral pattern. The complexity calls for development of reliable simulation techniques. In this work, a multistep approach combining molecular dynamics and quantum chemistry procedures is applied. The molecular dynamic is used for conformational studies, while the quantum chemical part provides VOA properties.

Due to computational demands of the quantum chemical methods, direct VOA calculations are limited to relatively small systems. As proteins are large molecules, it was necessary to apply procedures making VOA calculation feasible. This was achieved by system fragmentation and transfer of molecular property tensors. We used an automatic fragment generation combined with Cartesian Coordinate Transfer (CCT) methods [3, 4].

This multi-level computational strategy was used in VOA simulations of poly-L-glutamic acid, where the specific chiral ordering of sidechain carboxyl groups in the studied fibrillar structure was well explained. The procedure was then extended to a series of globular proteins and insulin fibrils.

Other chiral systems were studied by Electronic Circular dichroism (ECD) and Circularly Polarized Luminiscence (CPL) spectroscopies. We focused on two projects. The first one

studied how the metal ion influences ECD spectra of metal-monensin complexes. The second one dealt with CPL of racemic europium complex induced by interaction with aminoacids.

2 Theoretical Part

2.1 Chiroptical Spectroscopy

Chiroptical spectroscopy explores interaction of polarized light with a chiral sample. Historically, the first observation of this phenomenon was rotation of the plane of linearly polarized light, referred to as optical rotation (OR). Later, the absorption difference of left- and righthanded circularly polarized light, circular dichroism (CD), was observed as well.

2.1.1 Chirality

Chirality of a molecule means that it cannot be superimposed on its mirror image. These mirror "isomers" are called enantiomers. They have many chemical properties the same, which often makes them indistinguishable. One of possible techniques to resolve them is interaction with the linearly polarized light, when an opposite rotation of the plane of polarization is observed. Observation of this phenomenon started the chirality research.

The first experiment was carried out by Arago in 1811, who observed that plane of linearly polarized light was rotated when passed through a quartz crystal [5]. Later, in 1824, it was discovered by Fresnel that this was due to circular birefringence, i.e. difference in the refraction indices for left- and righthanded circularly polarized lights. Also, he suggested that molecules of the optically active medium exist in right- and left helical forms [6]. In the revolutionary year 1848 Pasteur conducted another remarkable experiment. He found a relation between the geometry of the crystals and optical activity of their solutions. He achieved this via separation of racemic mixture of tartaric acid crystals by hand [7]. However, he did not suggest any relation to molecular geometry. Next breakthrough came with the concept of the asymmetric carbon [8, 9]. This structural motif has two mirror-image forms. In 1904, the term of chirality started to be widely used. Lord Kelvin defined it as a property of any geometrical figure or group of points, mirror-image of which cannot be superimposed on itself [10].

A simplest molecular element, which leads to chirality, thus is the chiral center (fig. 2.1). Usually, it is an atom with sp^3 orbital symmetry forming four different bonds in a tetrahedron. The configuration of the chiral center is specified by the letters R for rectus (right) and S for sinister (left) orientation according Cahn-Ingold-Prelog system [11, 12].

Another possibility is helicity of the molecule. These structures are specified via positive P-helicity or negative M-helicity. Examples of such compounds are helicens shown in fig. 2.2. Other chiral structures lack chiral stereogenic center but have axial (fig. 2.3) or planar (fig. 2.4) chirality. Axial chirality is most commonly observed in a biaryl compounds where the rotation about the aryl-aryl bond is restricted, e.g. for biphenyl or binaphthyls. The

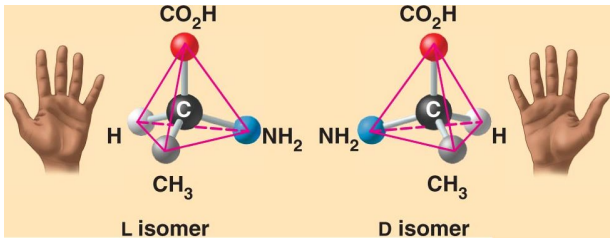


Figure 2.1: Example of asymmetric carbon center

compounds having planar chirality possess two non-coplanar rings. Example for ferocene molecule is shown in figure 2.4.

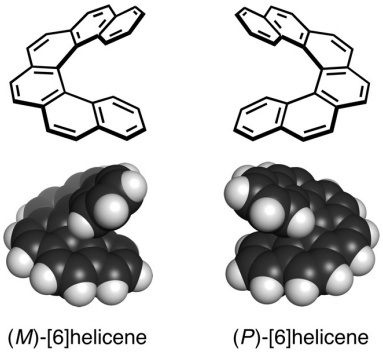


Figure 2.2: Helical chirality

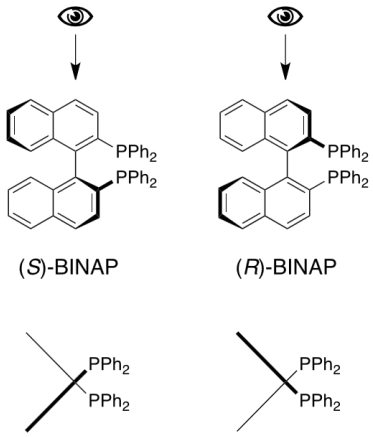


Figure 2.3: Axial chirality

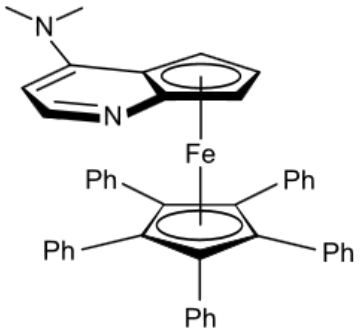


Figure 2.4: Planar chirality

2.1.2 Chiral Light

Electromagnetic radiation is oscillating wave of an electric and magnetic fields. In plane waves, the fields are orthogonal and change synchronously in time. Thus it is sufficient to consider the electric field (eq. 2.1).

$$\boldsymbol{E}(\boldsymbol{r},t) = E_0\boldsymbol{e}\exp[i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)] \tag{2.1}$$

where $\boldsymbol{E}(\boldsymbol{r},t)$ denotes the electric field vector, E_0 its magnitude, $\boldsymbol{e}$ the polarization vector, $\boldsymbol{r}$ the position vector, t time and ω angular frequency.

The polarization vector $\boldsymbol{e}$ describes the polarization state of the radiation. If we assume that the radiation is travelling in the z direction, we can simplify eq. 2.1 to 2.2:

$$\boldsymbol{E}(\boldsymbol{r},t) = (\boldsymbol{E}_X+\boldsymbol{E}_Y)\exp\Big[i(kz-\omega t)\Big] = E_0\boldsymbol{e}\exp\Big[i(kz-\omega t)\Big] \tag{2.2}$$

The tip of vector $\boldsymbol{E}$ follows a right or left handed helix in space, depending on the polarization. In the case the radiation moves towards an observer, he sees a circle. Thus

the polarization vector $\boldsymbol{e}$ rotates clockwise or anticlockwise (fig. 2.5). The vector $\boldsymbol{E}$ can be expressed as sum of two vectors $\boldsymbol{E}_X$ and $\boldsymbol{E}_Y$ for X and Y direction. These orientations can be written in the form of equations 2.3, where $\boldsymbol{e}_L$ means polarization vector for left circularly polarized and $\boldsymbol{e}_R$ for right circular polarized light.

$$\boldsymbol{e}_L = \frac{1}{\sqrt{2}}(\boldsymbol{e}_X + i\boldsymbol{e}_Y) \tag{2.3}$$

$$\boldsymbol{e}_R = \frac{1}{\sqrt{2}}(\boldsymbol{e}_X - i\boldsymbol{e}_Y) \tag{2.4}$$

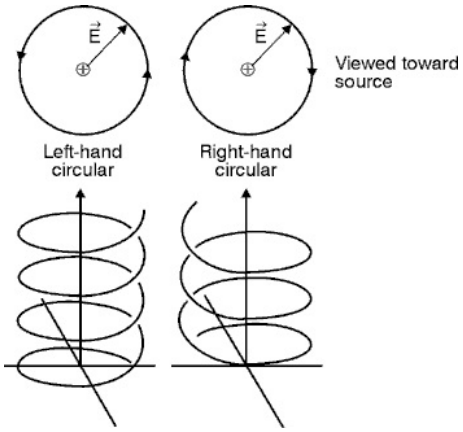


Figure 2.5: Left- and righthanded circularly polarized light

2.1.3 Electronic Circular Dichroism

Electronic Circular Dichroism (ECD) is defined as absorption difference of left and right-handed circularly polarized light in the UV-vis wavelength region ($\Delta A = A_L - A_R$). ECD is quick and undemanding chiroptical method for structural study of chiral molecules. It can be applied to studies of molecular complex formations, protein folding, etc. ECD is very suitable method for determination of the secondary structure content in proteins [13]. The amides groups are the most active chromophores of protein backbone. Their different arrangement leads to characteristic ECD spectra of different secondary structures (fig. 2.6).

2.1.4 Circularly Polarized Luminiscence

Circularly polarized luminescence (CPL) spectroscopy is differential spontaneous emission of left and right circularly polarized radiation ($\Delta I = I_L - I_R$). This phenomenon can be additionally quantified by luminiscence dissymmetry factor g_{lum}

$$g_{lum} = \frac{2(I_L - I_R)}{(I_L + I_R} = \frac{\Delta I}{I} \tag{2.5}$$

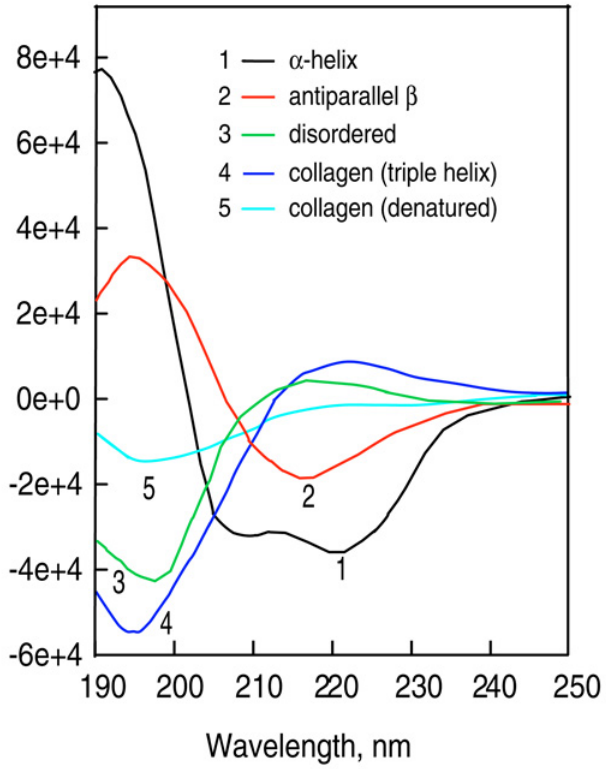


Figure 2.6: ECD of proteins of various secondary structure [13]

where I_L and I_R are left and right circularly polarized emission intensities. Situation when $g_{lum} = \pm 2$ means complete polarization of the emitted light; $g_{lum} = 0$ means totally unpolarized beam.

Typical value of g_{lum} for isolated organic molecules ranges from 10^{-3} to 10^{-2}. But for lanthanide complexes, values as high as one can be observed [14]; $g_{lum} = 1.38$ was the highest measured value so far [15]. The lanthanide complexes are thus excellent candidates to be studied by CPL. We can extract the information about the excited state chirality and observe transitions not easily observable in absorption spectra (e.g. $f \rightarrow f$ transitions in lanthanides) or we can select transitions using certain excitation wavelength [16].

2.1.5 Vibrational Optical Activity

Vibrational Optical Activity (VOA) is dealing with vibrational transitions. For its simulation, we need to obtain vibrational frequencies and intensities. We can achieve this with the aid of quantum chemical methods. In the Born-Oppenheimer approximation, we assume wavefunction of the system as a product of electronic and nuclear parts (eq. 2.6), where $\boldsymbol{R}_N$ means nuclear and $\boldsymbol{r}$ electron coordinates.

$$\psi(r_i, R_N) = \Phi^{el}_{R_n}(\boldsymbol{r})\Phi_{Ncl}(R_N) \tag{2.6}$$

Then we can solve vibrational problem separately. The nuclear wavefunction can be obtained from equation 2.7. It depends on $3N$ coordinates, but the vibrational potential of nonlinear molecule depends only on $3N-6$ ($3N-5$ for linear one) coordinates, where N is number of atoms. The remaining 6 (5) coordinates represent translational and rotational motions.

The Schrödinger equation for $\Phi_{ncl}(\boldsymbol{R}_N)$ is:

$$[\hat{T}_{ncl} + \epsilon(\boldsymbol{R}_N)]\Phi_{ncl}(\boldsymbol{R}_N) = E\Phi_{ncl}(\boldsymbol{R}_N) \tag{2.7}$$

The instant value of a coordinate R_j can be written as (eq. 2.8)

$$R_j = R_j^0 + \Delta R_j \qquad j = 1, 2 \ldots 3N. \tag{2.8}$$

The kinetic energy operator $\hat{T}_{ncl}^{vib}$ is:

$$\hat{T}_{ncl}^{vib} = -\sum_{j=1}^{3N} \frac{\hbar^2}{2M_j} \frac{\partial^2}{\partial \Delta R_j^2} = -\sum_{j=1}^{3N} \frac{\hbar^2}{2M_j} \frac{\partial^2}{\partial R_j^2} \tag{2.9}$$

and the potential $\epsilon(R)$ is:

$$\epsilon(\boldsymbol{R}) = \epsilon(\boldsymbol{R}^0) + \sum_{j=1}^{3N} \frac{\partial \epsilon}{\partial \Delta R_j} \Delta R_j + \frac{1}{2} \sum_{j,k=1}^{3N} \frac{\partial^2 \epsilon}{\partial \Delta R_j \partial \Delta R_k} \Delta R_j \Delta R_k + \ldots \tag{2.10}$$

The first term of the Taylor series we can be set to zero and the second term is also zero for equilibrium geometry. For the rigid molecules in the case of the harmonic approximation, we keep only the quadratic term. It is characterized by force field matrix ($\boldsymbol{f}$) with the elements:

$$f_{jk} = \frac{\partial^2 \epsilon}{\partial \Delta R_j \partial \Delta R_k} = \frac{\partial^2 \epsilon}{\partial R_j \partial R_k} \tag{2.11}$$

The equation 2.7 becomes:

$$\sum_{j,k=1}^{3N} \frac{1}{2} [f_{jk} \Delta R_j \Delta R_k - \frac{\hbar^2}{M_j} \frac{\partial^2}{\partial \Delta R_j^2}] \Phi_{Ncl}(\Delta R) = E^{vib} \Phi_{Ncl}(\Delta R) \tag{2.12}$$

This equation (eq. 2.12) can be solved using diagonalization of the force field and a linear transformation of Cartesian into normal modes coordinates Q_k (eq. 2.13-2.15).

$$Q_k = \sum_{j=1}^{3N} S_{kj}^{-1} \Delta R_j \tag{2.13}$$

$$\sum_{i,j=1}^{3N} f_{ij} \Delta R_i \Delta R_j = \sum_{k=1}^{3N} \omega_k^2 Q_k^2 \tag{2.14}$$

$$\sum_{j=1}^{3N} \frac{1}{M_j} \frac{\partial^2}{\partial \Delta R_j^2} = \sum_{k=1}^{3N} \frac{\partial^2}{\partial Q_k^2} \tag{2.15}$$

The vibrational Hamiltonian can be written as the sum of $3N-6$ harmonic oscillators and the vibrational wavefunction as a product of $3N-6$ individual functions (eq. 2.16) depending on one coordinate, which are obtained as a solution of the set of equations 2.17. The $\{\omega_j\}$ denote frequencies of individual modes, $\boldsymbol{S}$ is the transformation matrix, n_j is integer value and H_j vibrational hamiltonian for specific mode ω_j.

$$\Phi_{ncl} = \prod_{j=1}^{3N-6} \phi_j(Q_j) \tag{2.16}$$

$$\hat{H}_j \phi_j = (\frac{1}{2} + n_j) \hbar \omega_j \phi_j \tag{2.17}$$

2.1.5.1 Raman Optical Activity

For the description of Raman Optical Activity (ROA) we also use quantum mechanics. During the electromagnetic wave interaction with the molecule, electric charges start to oscillate and a secondary wave is scattered. This process is described by induced electric and magnetic multipole moments.

We can introduce the induced moments: electric dipole μ_α, magnetic dipole m_α and electric quadrupole $\Theta_{\alpha\beta}$ as:

$$\mu_\alpha = \alpha_{\alpha\beta} E_\beta + \frac{1}{\omega} G'_{\alpha\beta} \dot{B}_\beta + \frac{1}{3} A_{\alpha\beta\gamma} \nabla_\beta E_\gamma + \ldots \tag{2.18}$$

$$m_\beta = -\frac{1}{\omega} G'_{\alpha\beta} \dot{E}_\alpha + \ldots \tag{2.19}$$

$$\Theta_{\alpha\beta} = A_{\alpha\beta\gamma} E_\gamma \tag{2.20}$$

where $\alpha_{\alpha\beta}$ denotes electric dipole–electric dipole polarizability, $A_{\alpha\beta\gamma}$ electric dipole–electric quadrupole polarizability and $G'_{\alpha\beta}$ electric dipole–magnetic dipole polarizability. The quantum-mechanical expressions for these tensors can be obtained from time dependent perturbation approach. In non-resonant case, the relations are:

$$\alpha_{\alpha\beta} = \frac{2}{\hbar}\sum_{j\neq n}\frac{\omega_{jn}}{\omega_{jn}^2-\omega^2}\mathrm{Re}(\langle g|\mu_\alpha|j\rangle\langle j|\mu_\beta|f\rangle) \tag{2.21}$$

$$G'_{\alpha\beta} = -\frac{2}{\hbar}\sum_{j\neq n}\frac{\omega}{\omega_{jn}^2-\omega^2}\mathrm{Im}(\langle g|\mu_\alpha|j\rangle\langle j|m_\beta|f\rangle) \tag{2.22}$$

$$A_{\alpha\beta\gamma} = \frac{2}{\hbar}\sum_{j\neq n}\frac{\omega_{jn}}{\omega_{jn}^2-\omega^2}\mathrm{Re}(\langle m_v|\mu_\alpha|j\rangle\langle j|\Theta_{\beta\gamma}|f\rangle) \tag{2.23}$$

In these expressions, g, j and f denote initial, excited and terminal state of molecule; $\omega_{jn} = \omega_j - \omega_n$ denotes angular frequency difference. The initial and terminal state can be written as a product of vibrational and electronic part ($g = |el_0\rangle|m_v\rangle$ and $f = |el_0\rangle|n_v\rangle$). As for both cases the $|el_0\rangle$ is the same, we can take into account only the vibrational states m_v and n_v. The operators of the electric dipole, magnetic dipole and electric quadrupole moment are defined as:

$$\mu_\alpha = \sum_i e_i r_{i\alpha} \tag{2.24}$$

$$m_\alpha = \sum_i \frac{e_i}{2m_i}\epsilon_{\alpha\beta\gamma} r_{i\beta} p_{i\gamma} \tag{2.25}$$

$$\Theta_{\alpha\beta} = \frac{1}{2}\sum_i e_i(3r_{i\alpha}r_{i\beta} - r_i^2\delta_{\alpha\beta}) \tag{2.26}$$

where particle i has position vector $\boldsymbol{r}_i$, charge e_i, mass m_i and momentum $\boldsymbol{p}_i$.

For Raman and ROA tensors we obtain the transient tensors between initial (m_v) and terminal (n_v) vibrational states as:

$$\alpha_{\alpha\beta} \to \langle m_v|\alpha_{\alpha\beta}(Q)|n_v\rangle \tag{2.27}$$

$$G'_{\alpha\beta} \to \langle m_v|G'_{\alpha\beta}(Q)|n_v\rangle \tag{2.28}$$

$$A_{\alpha\beta\gamma} \to \langle m_v|A_{\alpha\beta\gamma}(Q)n_v\rangle \tag{2.29}$$

where the electronic polarizabilities $\alpha_{\alpha\beta}(Q)$, $G'_{\alpha\beta}(Q)$ and $A_{\alpha\beta\gamma}(Q)$) are dependent on the normal mode vibrational coordinates (Q). We can expand the polarizability into expression 2.30:

$$\langle m_v|\alpha_{\alpha\beta}(Q)|n_v\rangle = (\alpha_{\alpha\beta})_0\delta_{m_v n_v} + \sum_P \left(\frac{\partial\alpha_{\alpha\beta}}{\partial Q_P}\right)_0 \langle m_v|Q_P|n_v\rangle + \ldots \tag{2.30}$$

The index 0 denotes that the function is considered in equilibrium nuclear positions. The first term of eq. 2.30 describes Rayleigh, the second one the Raman scattering with

selection rule $m_v - n_v = \pm 1$ for harmonic oscillator case. Similar expansion can be done for G' and A.

As ROA is two photon phenomenon, its measurement can be done in several different ways. One of the varied parameters is the angle between scattered and incident light beam, which leads to the direct (0°), right-angle (90°) and back (180°) scattering. The other parameter is the polarization of incident or scattered light (fig. 2.7).

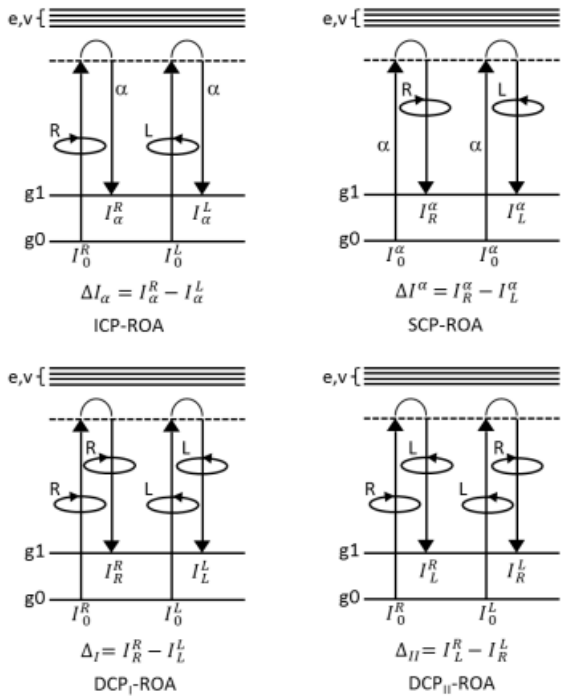


Figure 2.7: Diagram of energetic levels of a molecule and polarization states of incident and scattered light for the four experimental forms of ROA.

The simplest methods is ICP (Incident Circular Polarization), which is based on polarization modulation of the incident light. The detected scattered light is polarized or non-polarized. Another way is SCP (Scattered Circular Polarization) measurement, in which the sample is irradiated by linearly polarized light, and the left or right circular component of the scattered light is detected. Methods $\mathrm{DCP_I}$ and $\mathrm{DCP_{II}}$ (Dual Circular Polarization) use combinations of SCP and ICP; both the incident and scattered light are circularly polarized, either in phase ($\mathrm{DCP_I}$) or out of phase ($\mathrm{DCP_{II}}$).

For measurements in solutions, ROA tensors are averaged over all geometric orientations. The remaining tensor invariants are defined as:

$$\alpha = \frac{1}{3}\Big(\frac{\partial \alpha_{\alpha\alpha}}{\partial Q_P}\Big)_0 \tag{2.31}$$

$$G' = \frac{1}{3}(\frac{\partial G'_{\alpha\alpha}}{\partial Q_P}\Big)_0 \tag{2.32}$$

$$\beta(\alpha)^2 = \frac{1}{2}\Big[3(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_P}\Big)_0(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_P}\Big)_0 - (\frac{\partial \alpha_{\alpha\alpha}}{\partial Q_P}\Big)_0(\frac{\partial \alpha_{\beta\beta}}{\partial Q_P}\Big)_0\Big] \tag{2.33}$$

$$\beta(G')^2 = \frac{1}{2}\Big[3\Big(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_P}\Big)_0\Big(\frac{\partial G'_{\alpha\beta}}{\partial Q_P}\Big)_0 - (\frac{\partial \alpha_{\alpha\alpha}}{\partial Q_P}\Big)_0(\frac{\partial G'_{\beta\beta}}{\partial Q_P}\Big)_0\Big] \tag{2.34}$$

$$\beta(A)^2 = \frac{\omega_0}{2}\Big[\Big(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_P}\Big)_0\Big(\frac{\epsilon_{\alpha\gamma\delta}\partial A_{\gamma\delta\beta}}{\partial Q_P}\Big)_0\Big] \tag{2.35}$$

In the expressions, the Einstein summation convention is used. Two Greek subscripts in a symbol or a product means a summation over the Cartesian coordinates x,y and z.

Raman ($I_R + I_L$) and ROA ($I_R - I_L$) intensities can be then written as:

$$I_R + I_L = 4K(D_1\alpha^2 + D_2\beta(\alpha)^2) \tag{2.36}$$

$$I_R - I_L = \frac{8K}{c}(D_3\alpha G' + D_4\beta(G')^2 + D_5\beta(A)^2) \tag{2.37}$$

where c denotes the speed of light and K is a constant. Coefficients $D_1 - D_5$ are characteristic for specific ROA measurement setup and can be found in literature [17].

The spectrum at frequency ω is then obtained from the computed Raman intensities (I) of each mode with harmonic vibrational frequency ω_i using eq. 2.38. k denotes Boltzmann constant, T temperature, Δ full width at half maximum.

$$S(\omega) = \frac{1}{\omega_i}\frac{I}{(1-\exp[\frac{-\omega_i}{kT}])(4[\frac{\omega-\omega_i}{\Delta}]^2+1)} \tag{2.38}$$

2.1.5.2 Vibrational Circular Dichroism

Vibrational Circular Dichroism spectroscopy is a technique detecting difference of absorption of left and right circularly polarized radiation relevant to vibrational transitions (fig. 2.8). The contribution of particular transition $j \leftarrow n$ is determined by the rotational strength R_{jn}:

$$R_{jn} = Im\Big(\langle n|\boldsymbol{\mu}|j\rangle \cdot \langle j|\boldsymbol{m}|n\rangle\Big) \tag{2.39}$$

Dipole strength D_{jn} determines the absorption intensity:

$$D_{jn} = Re\Big(\langle n|\boldsymbol{\mu}|j\rangle \cdot \langle j|\boldsymbol{\mu}|n\rangle\Big) \tag{2.40}$$

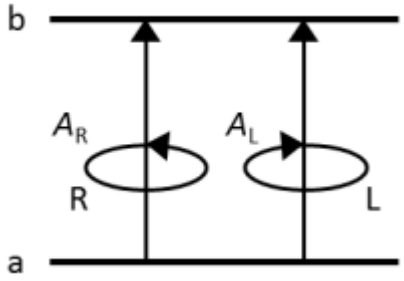


Figure 2.8: The diagram of vibrational transition measured in VCD

To characterize VCD intensity we can also use Kuhn's dimensionless dissymmetry factor (g) defined as:

$$g = \frac{\varepsilon_L - \varepsilon_R}{2(\varepsilon_L + \varepsilon_R)} = \frac{4R_{jn}}{cD_{jn}} \tag{2.41}$$

where ε_L and ε_R are extinction coefficients for left and right- circularly polarized light and c speed of light.

To compute VCD, we need to calculate the transition electric dipole moments. The electric dipole moment for the ground electronic state can be written as:

$$\mu_\beta = (\mu_\beta)_0 + \left(\frac{\partial \mu_\beta}{\partial R_\alpha^N}\right)_0 \Delta R_\alpha^N \tag{2.42}$$

where ΔR_α^N is displacement of nucleus N from its equilibrium position. The dipole derivates are called the Atomic Polar Tensor (APT):

$$P_{\alpha\beta}^N = \left(\frac{\partial \mu_\beta}{\partial R_\alpha^N}\right)_0 = \left[\frac{\partial}{\partial R_\alpha^N}\langle\phi_0(\boldsymbol{r};\boldsymbol{R})|\mu_\beta|\phi_0(\boldsymbol{r};\boldsymbol{R})\rangle\right]_0 \tag{2.43}$$

where $\phi_0(\boldsymbol{r};\boldsymbol{R})$ is the electronic wavefunction for fixed positions of nuclei. The partial derivative is taken in the equilibrium geometry.

The magnetic dipole $\boldsymbol{m}$ can be expressed as:

$$m_\beta = (m_\beta)_0 + \left(\sum_N \left(\frac{\partial m_\beta}{\partial \dot{R}_\alpha^N}\right)_0 \Delta \dot{R}_\alpha^N + \dots\right. \tag{2.44}$$

where $\dot{R}_\alpha^N$ are nuclear velocities and the set of partial derivatives appearing in the series is called Atomic Axial Tensor (AAT) $M_{\alpha\beta}^N$:

$$M_{\alpha\beta}^N = \left(\frac{\partial m_\beta}{\partial \dot{R}_\alpha^N}\right)_0 = \left[\frac{\partial}{\partial \dot{R}_\alpha^N}\langle\phi_0(\boldsymbol{r};\boldsymbol{R};\dot{\boldsymbol{R}})|m_\beta|\phi_0(\boldsymbol{r};\boldsymbol{R};\dot{\boldsymbol{R}})\rangle\right]_0 \tag{2.45}$$

The fundamental vibrational transition moments can be written as:

$$\langle 1|\mu_\beta|0\rangle = \sqrt{\frac{\hbar}{2\omega_i}} \sum_{N,\alpha} P^N_{\alpha\beta} s^N_{\alpha,i} \tag{2.46}$$

$$\langle 1|m_\beta|0\rangle = i\sqrt{\frac{\hbar\omega_i}{2}} \sum_{N,\alpha} M^N_{\alpha\beta} s^N_{\alpha,i} \tag{2.47}$$

2.1.5.3 Applications of Vibrational Optical Activity

High resolution techniques such as X-ray or NMR provide a detailed information about molecular structure, but also face many limitations. For the X-ray technique, it is the necessity to have crystal of studied molecule. The NMR spectroscopy, capable to study the systems in solutions, is limited by molecular size. Usually, molecular weight can not be larger than 100 kDa [18]. Both techniques suffer problems when they are used to study flexible molecules.

Therefore, it is convenient to use methods of vibrational optical activity. Although these methods cannot give us the high-resolution data at the atomic level, they can provide valuable information about the structure and system dynamical behaviour.

Proteins are very interesting and challenging molecules to study. Their conformational analysis is based on the evaluation of their secondary structure. Secondary structure elements such as α-helices,β-sheets, polyproline II, etc. provide characteristic spectral patterns [2, 19, 20, 21, 22]. The ROA and VCD spectroscopies have different selection rules and therefore provide different sensitivity to distinct parts of proteins.

From VCD spectra, the secondary structure can be derived from the patterns of amide I and amide II regions [2, 23, 24]. Characteristic patterns for α-helical structure include a negative/positive couplet at 1660 and 1640 cm^{-1} in amide I and negative band at 1515 cm^{-1} in the amide II region. The β-sheet structure has characteristic negative band in the amide I and the positive/negative couplet in the amide II region (fig. 2.9).

VCD is also sensitive to aggregated structures, such as protein fibrils [26]. The formation of fibrils sometimes causes an significant enhancement of VCD intensity (fig. 2.10) [27, 28].

Unlike VCD, Raman Optical Activity (ROA) spectroscopy provides wider range of vibrational frequencies (the VCD low-frequency limit is about 900 cm^{-1}, while the ROA limit is about 100 cm^{-1}). ROA spectra are also rich in structural information. Vibrations of protein backbone can be found in the three main regions: backbone skeletal stretch region (870-1150 cm^{-1}), extended amide III region (1230-1350 cm^{-1}) and amide I region (1600-1700 cm^{-1}). For ROA, the most informative is considered the extended amide III region, due to high sensitivity of the coupling between N-H and $^{\alpha}$CH deformations to protein geometry [29]

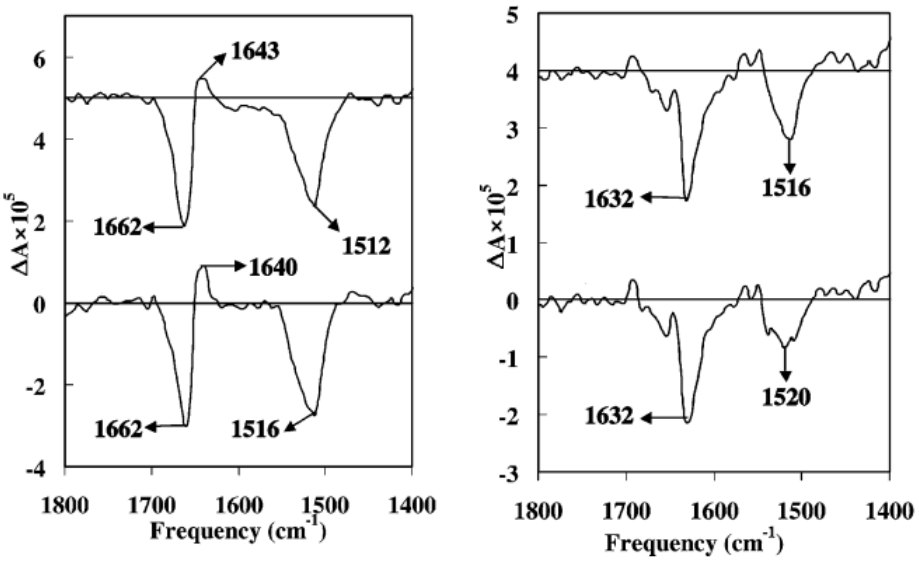


Figure 2.9: Example of VCD spectra of α-helical [bovine serum albumin (left top) and hemoglobin (left bottom)] and β-sheet [pepsin (right top) and chymotrypsin (right bottom)] proteins [25].

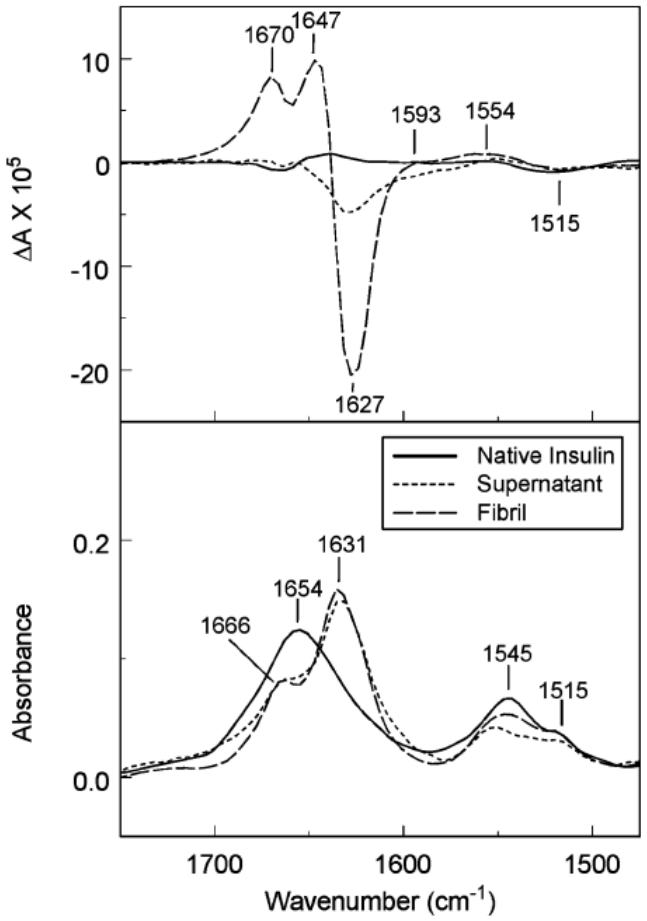


Figure 2.10: VCD and IR spectra of native insulin, centrifuged insulin supernatant, and centrifuged insulin fibril gel at pH=2 [26].

Thus bands can be assigned to particular secondary structures, such as α-helices (positive band at 1340-1345 cm^{-1} in extended amide III region, negative/positive couplet at 1640 and 1660 cm^{-1} in amide I region and positive signal in the range 870–950 cm^{-1}) or β-sheets (negative band at 1245 cm^{-1} in amide III region, negative/positive couplet at 1650 and 1680 cm^{-1} in amide I region) [30].

ROA can also provide information on the sidechains. For example, ROA signals of aromatic sidechains [31] can be found in several frequency regions, as 1545–1560 cm^{-1} and 1400–1480 cm^{-1} or 1600-1630 cm^{-1} [30, 19].

2.2 Methods for Simulations of Vibrational Spectra

2.2.1 Molecular Dynamics

Molecular dynamics (MD) describes evolvement of molecular systems in time. Depending on the theory used for potential determination, we can distinguish two MD types, i.e ab-initio [32] and classical [33, 34]. In both, the classical Newton equations of motion are used (2.49), where $\boldsymbol{f_i}$ is force acting on specific particle and $\ddot{\boldsymbol{R}}_i$ is the second time derivate of position, i.e. acceleration. In ab-initio MD the interaction potential U is calculated by DFT while in the classical MD) is empirical.

$$\boldsymbol{f_i} = -\frac{\partial U(\boldsymbol{R})}{\partial \boldsymbol{R}_i} \tag{2.48}$$

$$\ddot{\boldsymbol{R}}_i = \frac{\boldsymbol{f_i}}{m_i} \tag{2.49}$$

Due to its computational demand, ab-inito MD is limited to relatively small systems (100 atoms). Classical MD enables to investigate very extensive systems (10000 atoms). The Newton equations of motion can be solved by a variety of numerical methods. The most common integration algorithm is the Verlet [35] method.

Verlet method is a time reversible procedure for integration of Newton equations of motion. It can be derived from a numerical formula for the second derivatives of coordinates $\boldsymbol{R}_i$ (eq. 2.50), reorganized to equation 2.51. The term $O(h^2)$ is neglected. As the dynamic starts from time t_0, and coordinates in time $t_0 - h$ are also needed, which can be obtained from equation 2.52. The initial velocities $\dot{\boldsymbol{R}}_i(t_0)$ are determined randomly but with respect to the desired temperature.

$$\ddot{\boldsymbol{R}}_i = -\frac{\boldsymbol{R}_i(t-h) - 2\boldsymbol{R}_i(t) + \boldsymbol{R}_i(t+h)}{h^2} + O(h^2) \quad (2.50)$$

$$\boldsymbol{R}_i(t+h) = 2\boldsymbol{R}_i(t) - \boldsymbol{R}_i(t-h) + h^2\frac{\boldsymbol{f}_i(t)}{m_i} \quad (2.51)$$

$$\boldsymbol{R}_i(t_0-h) = 2\boldsymbol{R}_i(t) - h^2\dot{\boldsymbol{R}}_i(t_0) + \frac{h^2}{2}\frac{\boldsymbol{f}_i(t_0)}{m_i} \quad (2.52)$$

The Verlet method is often expressed in form of relations 2.53–2.54. The positions and velocities are calculated in different times, differing by half of the integration step. This method is called "leap-frog" [36].

$$\dot{\boldsymbol{R}}_i(t+\frac{h}{2}) = \dot{\boldsymbol{R}}_i(t-\frac{h}{2}) + h\frac{\boldsymbol{f}_i}{m_i} \quad (2.53)$$

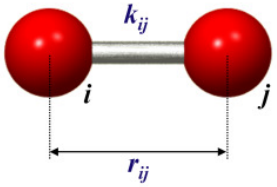
$$\boldsymbol{R}_i(t+h) = \boldsymbol{R}_i(t) + h\dot{\boldsymbol{R}}_i(t+\frac{h}{2}) \quad (2.54)$$

$$\dot{\boldsymbol{R}}_i(t+\frac{h}{2}) = \frac{\boldsymbol{R}_i(t+h) - \boldsymbol{R}_i(t)}{h} + O(h^2) \quad (2.55)$$

In classical MD the interaction potential is usually determined empirically. It contains many terms. Some of them describe covalent bonds. They can be described by harmonic functions 2.56 or by anharmonic Morse potential 2.57 (fig. 2.11) [37]; k_{ij} denotes harmonic bond force constant, r_{eq} equillibrium distance of nuclei, D_{eq} value of dissociation and a bond strength.

$$V_B = \sum_{bonds} k_{ij}\,(r_{ij} - r_{eq})^2 \quad (2.56)$$

$$V_B = \sum_{bonds} D_{eq}[1 - e^{-a(r_{ij}-r_{eq})^2}] \quad (2.57)$$



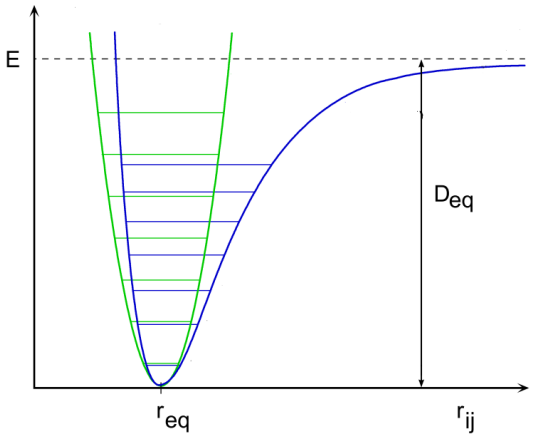


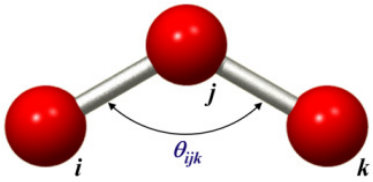
Figure 2.11: Morse potentiall

Other terms describe bond angle deformations (eq. 2.58), where K_θ denotes bond angle force constant and θ_{eq} equilibrium bond angle value. In some cases, Urey-Bradley potential (eq. 2.59) is used, introducing a virtual bond r_{ik}, securing an angle value by a distance r_0.

For torsion angles, we distinguish two types, proper (fig. 2.12, eq. 2.60) and improper (fig. 2.13, eq. 2.61). The proper ones describe torsion angle rotational motion, whereas

$$V_A = \sum_{angles} K_\theta(\theta - \theta_{eq})^2 \quad (2.58)$$

$$V_A = \sum_{angles} K_{ik}^{UB}(r_{ik} - r_0)^2 \quad (2.59)$$



improper torsion angles are used to control configuration of 4 atoms, where 3 of them are in plane. The V_D denotes proper and V_I improper torsion potential and the Φ is value of respective torsion angles.

$$V_D = \sum_{n=1}^{3} K_\phi[1 - \cos(n\phi - \delta)] \quad (2.60)$$

$$V_I = K_I(\phi_{ijkl} - \phi_0)^2 \quad (2.61)$$

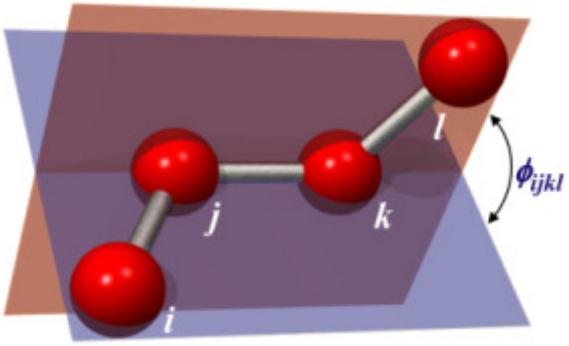


Figure 2.12: Torsion angle

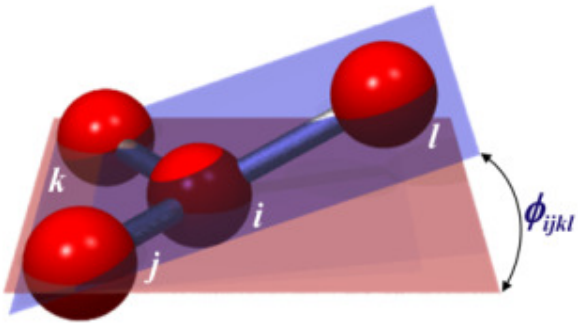


Figure 2.13: Improper torsion angle

Interaction potential also includes nonbonding terms. One of them is the dispersion interaction. The most common and popular are the Lennard-Jones [38] (eq. 2.62) and Buckhingham potentials (eq .2.63) [39].

$$V(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (2.62)$$

$$V(r_{ij}) = A_{ij}\exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^6} \quad (2.63)$$

where r_{ij} is distance between nuclei, parameters for Lennard-Jones potential ε_{ij} and σ_{ij} describe energy of the interaction and equilibrium distance, A_{ij}, B_{ij}, C_{ij} are constants for the Buckingham potential.

Another type of the non-bonding terms are the electrostatic interactions. The most significant one is charge-charge (q-q) interaction (eq. 2.64).

$$V = \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{2.64}$$

Polarizability is included as an advanced term. Induced dipole moment is linearly proportional to the intensity of the electric field via the polarizability [40, 41]. In some nonpolarizable force fields, the polarizability is compensated by adjusted partial charges [42].

To deal with border effects, the periodic boundary conditions are often applied. The simulated set of particles interacts with itself so that the box is replicated in all directions [43]. Atoms in the cell are replicated to form an infinite lattice. For atoms with positions $\boldsymbol{r}_i$, the Periodic Boundary Condition (PBC) produces their images defined as:

$$\boldsymbol{r}_i^{image} = \boldsymbol{r}_i + l\boldsymbol{a} + m\boldsymbol{b} + n\boldsymbol{c} \tag{2.65}$$

where $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ are vectors corresponding to the edges of the box and l, m, n any integers from $-\infty$ to ∞.

Each atom in the cell is interacting with other atoms in the central box (green, fig. 2.14) and also with the images in other boxes.

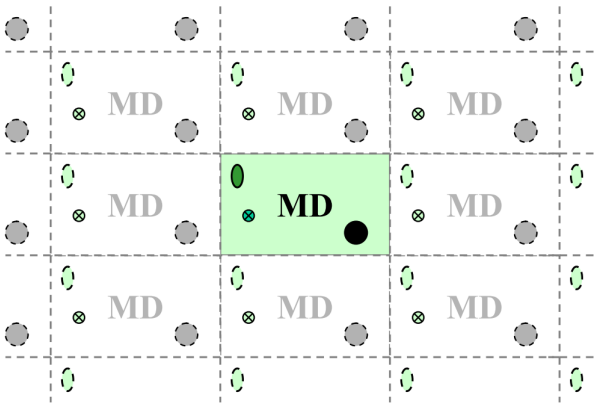


Figure 2.14: Infinite latice produced from primitive cell using application of the PBC

2.2.2 Methods of Configuration Space Sampling

For reliable description of systems by MD, it is necessary to reach every possible microstate. However, some configurations are energetically unfavourable and thus unreachable by classical MD simulation. For these cases, techniques of effective sampling of the configuration space were developed. One of the approaches is the calculation of potential of mean

force $W(\xi)$ (PMF) along coordinate ξ [44]. It is defined from average distribution function $\langle\rho(\xi)\rangle$ (eq. 2.66).

$$W(\xi) = W(\xi^*) - k_B T \ln\left[\frac{\langle\rho(\xi)\rangle}{\langle\rho(\xi^*)\rangle}\right] \tag{2.66}$$

ξ^* and $\langle\rho(\xi^*)\rangle$ are constants. Average distribution function is described in equation 2.67, $U(\boldsymbol{R})$ denotes energy of system as function of coordinates $\boldsymbol{R}$. Function $\xi^{'}(\boldsymbol{R})$ describes behaviour of other degrees of freedom, apart from observed ξ.

$$\langle\rho(\xi)\rangle = \frac{\int d\boldsymbol{R}\delta(\xi^{'}(\boldsymbol{R}) - \xi)\exp(-\frac{U(\boldsymbol{R})}{k_B T})}{\int d\boldsymbol{R}\exp(-\frac{U(\boldsymbol{R})}{k_B T})} \tag{2.67}$$

To search the whole space along coordinate ξ, a series of restrained simulations, biased by a harmonic potential $w_i(\xi)$, is performed [45].

One of such approaches is the WHAM method [46, 47] (Weighted Histogram Analysis Method). The unrestrained distribution function $\langle\rho(\xi)\rangle$ is expressed as weighted sum N_w of restrained distribution functions $\langle\rho(\xi)\rangle_{(i)}$(eq.2.68), n_i denotes number of points used for restrained distribution function, F_i values of free energies of simulation windows, which are defined by equation 2.69. These equations are solved self-consistently.

$$\langle\rho(\xi)\rangle = \sum_{i=1}^{N_w} n_i \langle\rho(\xi)\rangle_{(i)} \left[\sum_{j=1}^{N_w} n_j \exp\left(-\frac{w_j(\xi) - F_j}{k_B T}\right)\right]^{-1} \tag{2.68}$$

$$\exp[-\frac{F_i}{k_B T}] = \int d\xi \exp\left(-\frac{w_i(\xi)}{k_B T}\right)\langle\rho(\xi)\rangle \tag{2.69}$$

Similar technique is the W-PMF method [48], based on individual unrestricted PMF profiles $W_i(\xi)$, according to equation 2.70, C_i denotes constant defined by equation 2.71. The individual values $W_i(\xi)$ are obtained iteratively. The final potential $W(\xi)$ is obtained from equation 2.72.

$$W_i(\xi) = -k_B T \ln[\langle\rho(\xi)\rangle_i] - w_i(\xi) + C_i \tag{2.70}$$

$$C_i = k_B T ln[\langle\rho(\xi)\rangle_{(i)}] \tag{2.71}$$

$$W(\xi) = \sum_{i=1}^{N_w} W_i(\xi)\left[\frac{\langle\rho(\xi)\rangle_{(i)}}{\sum_{j=1}^{N_w}\langle\rho(\xi)\rangle_{(j)}}\right] \tag{2.72}$$

2.2.3 Special Techniques

2.2.3.1 Normal Mode Coordinate Optimization

These coordinates enable to separate high and low frequency modes. We can restrain, for example, low frequency modes connected with mutual orientation of molecular segments responsible for specific conformations. The high frequency modes, such as C-H stretching, can be relaxed to improve the structure and calculated spectral properties.

As follows from the equations 2.12 - 2.15, the displacements in Cartesian coordinates are connected with the normal mode coordinates through linear transformation:

$$\Delta R_i = \sum_j \frac{1}{\sqrt{m_i}} s_{ij} Q_j = \sum_j S_{ij} Q_j \tag{2.73}$$

which is referred to as S-matrix transformation.

The algorithm consists of the following steps. The first one is estimation of initial force field ($\boldsymbol{F^{(i)}}$) and obtaining the $\boldsymbol{S}$ matrix. In the second step, Cartesian gradient $\boldsymbol{g_c^{(i)}}$ is calculated. In the case the previous step is available, the Hessian is updated using formula [49, 50, 51, 52]:

$$\boldsymbol{F}^{(i+1)} = \boldsymbol{F}^{(i)} - \left(\frac{\Delta \boldsymbol{g}^{(i)^t} \Delta \boldsymbol{g}^{(i)}}{\boldsymbol{dR}^{(i)} \cdot \Delta \boldsymbol{g}^{(i)}} + \frac{\boldsymbol{F}^{(i)} \cdot \boldsymbol{dR}^{(i)^t} \boldsymbol{dR}^{(i)} \cdot \boldsymbol{F}^{(i)}}{\boldsymbol{dR}(i) \cdot \boldsymbol{F}^{(i)} \cdot \boldsymbol{dR}^{(i)}} \right) \tag{2.74}$$

where Cartesian displacements $\boldsymbol{dR}^{(i)} = \boldsymbol{R}^{(i)} - \boldsymbol{R}^{(i-1)}$ and the gradient differences $\Delta \boldsymbol{g}^{(i)} = \boldsymbol{g}_c^{(i)} - \boldsymbol{g}_c^{(i-1)}$. Afterwards, we obtain new $\boldsymbol{S}$ matrix. Then, the gradient is calculated in normal mode coordinates as $\boldsymbol{g}^{(i)} = \boldsymbol{S}^t \boldsymbol{g}_c^{(i)}$. In the case the gradient is below certain limit, the optimization is terminated. Otherwise a new step is produced, using quadratic dependence and its RFO (Rational Function Optimization) extension [53, 54, 55, 56, 57] (eq. 2.75). New Cartesian coordinates are defined by equation 2.76

$$\boldsymbol{dQ}^{(i+1)} = -\frac{2\boldsymbol{g}^{(i)}}{\boldsymbol{D}_{ii} + \sqrt{\boldsymbol{D}_{ii}^2 + 4(\boldsymbol{g}^{(i)})^2}} \tag{2.75}$$

$$\boldsymbol{R}^{(i+1)} = \boldsymbol{R}^{(i)} - \boldsymbol{S}\boldsymbol{dQ}^{(i+1)} \tag{2.76}$$

where $D_{ij} = \omega_i^2 \delta_{ij}$ is element of a matrix containing the second derivatives of energy at optimization point i.

After this step, the next Cartesian gradient is calculated and optimization procedure scheme is repeated until convergence.

2.2.3.2 Transfer of Molecular Property Tensors

Ab-initio vibrational property calculations are computationally demanding with respect to the system size. If we consider the simple GGA DFT functionals, computational time needed for the calculation of the energy is proportional to the 3^{rd} power of the system size. For hybrid or double-hybrid functionals, the dependence is even steeper. One way to overcome this obstacle is the transfer of molecular property tensors [3, 4]. The systems is divided into a set of smaller fragments, which are then used for ab-initio calculation of molecular property tensors, followed by a transfer to the original big molecule.

For a set of fragments containing atom pair $\lambda\mu$ (e. g. xfx), we search the best overlap of fragments f with the a segment F corresponding to big systems (e.g. XXFXX). Then sets of n atoms containing or having connection to atoms λ and μ are selected. The orientation of the f part of fragment xfx is adjusted to optimize overlap with the big molecule (XXFXX) using minimization function:

$$\delta(\boldsymbol{U}) = \sum_{i=1..n} (\boldsymbol{r}_i(F) - \boldsymbol{U} \cdot \boldsymbol{r}^i(f))^2 \tag{2.77}$$

where $\boldsymbol{r}_i(F)$ and $\boldsymbol{r}_i(f)$ denote coordinates of atoms in the big and small parts with the respect to geometric centers of the fragments. $\boldsymbol{U}$ denotes the unitary transformation matrix depending on the Euler angles. Cartesian force constants for atom pair $\lambda\mu$ are transferred by unitary transformation:

$$\frac{\partial^2 E}{\partial \boldsymbol{r}^\lambda_\alpha \partial \boldsymbol{r}^\mu_\beta}(F) = \sum_\gamma \sum_\delta \boldsymbol{U}_{\alpha\gamma} \boldsymbol{U}_{\beta\delta} \frac{\partial^2 E}{\partial \boldsymbol{r}^\lambda_\gamma \partial \boldsymbol{r}_\delta}(f) \tag{2.78}$$

Other molecular property tensors needed for evaluation of vibrational spectra intensities comprise several electric property tensors, respectively their derivatives, as mentioned in sections 2.1.5.1 and 2.1.5.2. The important issue is their origin dependence. The electric dipole $\boldsymbol{\mu}$ and electric dipole–electric dipole polarizabilities $\boldsymbol{\alpha}$ are origin independent. But the origin dependence of $\boldsymbol{m}$, $\boldsymbol{G'}$ and $\boldsymbol{A}$ tensors makes the necessity to use the distributed origin gauge [58, 59], when derivatives are expressed in systems of coordinates having origins on the moving atom λ (eq. 2.79–2.81).

$$\frac{i}{2\hbar}\frac{\partial m_\alpha(\lambda)}{\partial p^\lambda_\epsilon} = M^\lambda_{\epsilon,\alpha}(\lambda) = M^\lambda_{\epsilon,\alpha}(0) - \frac{i}{4\hbar c}\epsilon_{\alpha\gamma\delta} R^\lambda_\gamma P^\lambda_{\epsilon,\delta} \tag{2.79}$$

$$\frac{\partial G'_{\alpha\beta}(\lambda)}{\partial R^\lambda_\epsilon} = \frac{\partial G'_{\alpha\beta}(0)}{\partial R^\lambda_\epsilon} + \frac{\omega}{2}\epsilon_{\beta\gamma\delta} \times R^\lambda_\gamma \frac{\partial \alpha_{\alpha\beta}(0)}{\partial x^\lambda_\epsilon} \tag{2.80}$$

$$\frac{\partial A_{\alpha,\beta\gamma}(\lambda)}{\partial R^\lambda_\epsilon} = \frac{\partial A_{\alpha,\beta\gamma}(0)}{\partial R^\lambda_\epsilon} - \frac{3}{2}[R^\lambda_\beta \frac{\partial \alpha_{\alpha\gamma}(0)}{\partial R^\lambda_\epsilon}(0) + \; R^\lambda_\gamma \frac{\partial \alpha_{\alpha\beta}(0)}{\partial R^\lambda_\epsilon}] + R^\lambda_\delta \frac{\partial \alpha_{\alpha\delta}(0)}{\partial R^\lambda_\epsilon}\delta_{\beta\gamma} \tag{2.81}$$

These tensors are transferred using the same transformation matrix $\boldsymbol{U}$ as for the force field.

2.2.4 Quantum Chemical Methods

In this section, the quantum chemical methods used for calculation of vibrational spectra will be introduced. The simplest reasonable approach is the Hartree-Fock method. Another one is the Density Functional Theory (DFT), which I mainly used for simulation of Vibrational and Vibrational Optical Activity (VOA) and Electronic Circular Dichroism (ECD) spectra.

2.2.4.1 Hartree-Fock Method

Approximate wave function of many electron system is usually constructed form one electron functions known as molecular spinorbitals $\chi_i(\boldsymbol{x}_i)$. They are orthonormal functions defined as product of space $\psi_i(\boldsymbol{r}_i)$ and spin $s(\sigma_i)$ wavefunctions:

$$\chi(\boldsymbol{x}_i) = \psi_i(\boldsymbol{r}_i)s(\sigma_i) \tag{2.82}$$

where $\boldsymbol{x_i}$ is general coordinate of i^{th} atom consisting of space $\boldsymbol{r}_i$ and spin σ_i coordinate.

For the whole systems wavefunction of electrons, it is necessary to fulfil the Pauli principle, i.e. the antisymmetry of wavefunction with the respect to two particle exchange. The simplest wavefunction can be written in the form of Slater determinant.

$$\Phi(\boldsymbol{r}_1;\boldsymbol{r}_2;...\boldsymbol{r}_N) = \frac{1}{\sqrt{N!}}\begin{bmatrix} \chi_1(\boldsymbol{x}_1) & \dots & \chi_N(\boldsymbol{x}_1) \\ \vdots & \ddots & \vdots \\ \chi_1(\boldsymbol{x}_N) & \dots & \chi_N(\boldsymbol{x}_N) \end{bmatrix} \tag{2.83}$$

More accurate solution of Schroedinger equation can be obtained by an expansion into more Slater determinants (eq. 2.84). The Hartree-Fock method uses only one determinant.

$$|\psi\rangle = \sum_{n=1}^{\infty} C_n|\Phi_n\rangle \tag{2.84}$$

It is practical to separate the spin component and the spatial molecular orbitals $\psi_i(\boldsymbol{r})$. Atomic orbitals $\phi_\mu(\boldsymbol{r})$ are used to construct molecular orbitals as:

$$\psi_i = \sum_{\mu} c_{i,\mu}\phi_\mu(\boldsymbol{r}) \tag{2.85}$$

The Hartree-Fock equations can be written as:

$$f_1\chi_m(\mathbf{1}) = \varepsilon_m\chi_m(\mathbf{1}) \tag{2.86}$$

where f_1 is Fock operator, which is defined in the terms of one-electron hamiltonian h_1, Coulomb J_m and exchange K_m operators:

$$f_1 = h_1 + \sum_{m'}(2J_{m'}(\mathbf{1}) - K_{m'}(\mathbf{1})) \tag{2.87}$$

$$J_{m'}(\mathbf{1})\psi_m = \int \psi^*_{m'}(\mathbf{2})\frac{1}{r_{12}}\psi_m(\mathbf{1})\psi_{m'}(\mathbf{2})\mathrm{d}\tau_2 \tag{2.88}$$

$$K_{m'}(\mathbf{1})\psi_m = \int \psi^*_{m'}(\mathbf{2})\frac{1}{r_{12}}\psi_{m'}(\mathbf{1})\psi_m(\mathbf{2})\mathrm{d}\tau_2 \tag{2.89}$$

The sum in eq. 2.87 runs over occupied molecular orbitals. The Coulomb operator (eq. 2.88) takes into account the Coulombic electron-electron repulsion and the exchange operator (eq. 2.89) effects of electron correlation. The sum in eq. 2.87 represents the average potential energy of electron **1** in the presence of all other $N_e - 1$ electrons. Each molecular orbital is obtained as a solution of equation 2.86 by iteration; the potential is sometimes called self consistent field (SCF).

If we substitute the atomic orbitals into molecular ones, we obtain:

$$f_1\sum_{\mu=1}^{N_o} c_{\mu m}\phi_\mu(\mathbf{1}) = \varepsilon_m\sum_{\mu=1}^{N_o} c_{\mu m}\phi_\mu(\mathbf{1}) \tag{2.90}$$

The N_o denotes number of basis functions and $c_{\mu m}$ coefficients. By multiplication of both equation sides by $\phi^*_{\mu'}(\mathbf{1})$ and integration over all space, we get:

$$\sum_{\mu=1}^{N_o} c_{\mu m}\int \phi^*_{\mu'}(\mathbf{1})f_1\phi_\mu(\mathbf{1})\mathrm{d}\tau_1 = \varepsilon_m\sum_{\mu=1}^{N_o} c_{\mu m}\int \phi^*_{\mu'}(\mathbf{1})\phi_\mu(\mathbf{1})\mathrm{d}\tau_1 \tag{2.91}$$

The previous equation can be simplified, when we introduce overlap matrix $\boldsymbol{S}$ and Fock matrix $\boldsymbol{F}$ as defined in eq. 2.92 and 2.93:

$$S_{\mu'\mu} = \int \phi^*_{\mu'}(\mathbf{1})\phi_\mu(\mathbf{1})\mathrm{d}\tau_1 \tag{2.92}$$

$$F_{\mu'\mu} = \int \phi^*_{\mu'}(\mathbf{1})f_1\phi_\mu(\mathbf{1})\mathrm{d}\tau_1 \tag{2.93}$$

Expression 2.91 then becomes:

$$\sum_{\mu=1}^{N_o} c_{\mu m}F_{\mu'\mu} = \varepsilon_m\sum_{\mu=1}^{N_o} S_{\mu'\mu}c_{\mu m} \tag{2.94}$$

which is known as the Roothan equations. They can be rewritten as a single matrix equation:

$$\boldsymbol{F}\boldsymbol{c} = \boldsymbol{\varepsilon}\boldsymbol{S}\boldsymbol{c} \tag{2.95}$$

where $\boldsymbol{c}$ is an $N_o \times N_o$ matrix containing elements $c_{\mu m}$ and $\boldsymbol{\varepsilon}$ is an $N_o \times N_o$ diagonal matrix containing orbital energies ε_m.

2.2.4.2 Density Functional Theory Method

The principal idea of DFT is the replacement of many electron wavefunction (Ψ) by electronic density ρ. It can be defined as:

$$\rho(\boldsymbol{r}) = N \int \dots \int |\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N)|^2 \mathrm{d}\boldsymbol{r}_2 ... \mathrm{d}\boldsymbol{r}_N \tag{2.96}$$

and satisfies the conditions:

$$\int \rho(\boldsymbol{r})\mathrm{d}\boldsymbol{r} = N \text{ and } \rho(\boldsymbol{r}) \geq 0 \tag{2.97}$$

According to the Hohenberg-Kohn theorem [60], the ground state properties of a many-electron system are uniquely determined by the electron density $\rho(\boldsymbol{r})$ which can be determined by the variation of energy (eq. 2.98).

$$\frac{\delta E(\rho)}{\delta \rho} = 0 \tag{2.98}$$

Functional of energy $E(\rho)$ is expressed as:

$$E(\rho) = T(\rho) + V(\rho) + \int \rho(\boldsymbol{r})\nu(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \tag{2.99}$$

where $T(\rho)$ is kinetic energy operator, external potential $\nu(\boldsymbol{r})$ and $V(\rho)$ contains the coulomb interactions of electrons with nuclei.

In the Kohn-Sham DFT formulation, we need to find a set of one electron equations. To derive it, a hypothetical system consisting of N_e non-interacting electrons in the external potential $\nu_{ref}(\boldsymbol{r})$ is considered [61]. The potential is selected to satisfy that ρ_{ref} is identical to true system electron density ρ. Hamiltonian for the reference system is defined as:

$$h_{ref} = \sum_{i=1}^{N_e} h_i^{KS} \qquad h_i^{KS} = -\frac{\hbar^2}{2m_e}\nabla_i^2 + \nu_{ref}(\boldsymbol{r}_i) \tag{2.100}$$

where the one electron Kohn-Sham orbitals ψ_m^{KS} are eigenfunctions of the one-electron hamiltonians h_i^{KS}:

$$h_i^{KS}\psi_m^{KS}(i) = \varepsilon_m^{KS}\psi_m^{KS}(i) \tag{2.101}$$

The total energy functional is:

$$E(\rho) = T(\rho) + J(\rho) + \int \rho(\boldsymbol{r})\nu(\boldsymbol{r})\mathrm{d}\boldsymbol{r} + E_{XC}(\rho) \tag{2.102}$$

where E_{XC} is the exchange-correlation energy.

Functional derivative of E_{XC} with the respect to the electronic density ρ is called exchange-correlation potential $\nu_{XC}(\boldsymbol{r})$:

$$\nu_{XC}(\boldsymbol{r}) = \frac{\partial E_{XC}}{\partial \rho} \tag{2.103}$$

The energy can be obtained by variational procedure, when we get Kohn-Sham equations in the form:

$$\Big[-\frac{\hbar^2}{2m_e}\nabla_i^2 + \nu_{eff}(\boldsymbol{r}_i)\Big]\psi_m^{KS}(i) = \varepsilon_m^{KS}\psi_m^{KS}(i) \tag{2.104}$$

where the effective potential $\nu_{eff}(\boldsymbol{r})$ is defined as:

$$\nu_{eff}(\boldsymbol{r}) = \nu(\boldsymbol{r}) + \frac{1}{4\pi\varepsilon_0}\int \frac{\rho(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|}\mathrm{d}\boldsymbol{r'} + \nu_{XC}(\boldsymbol{r}) \tag{2.105}$$

In ideal case, if we had the correct form of $\nu_{XC}(\boldsymbol{r})$, we could obtain the exact solution of Schrödinger equation.

3 Methods

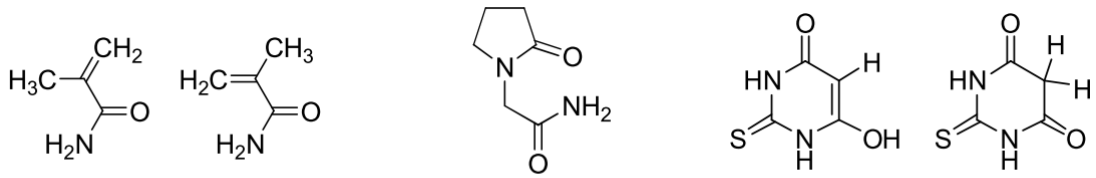
3.1 Studied Systems

3.1.1 Small Chiral Molecules

Raman Spectroscopy of Polymorphs

Depending on the crystallization conditions, many organic compounds form crystals of different structure and their proper characterization is important for pharmaceutical industry. Our goal was to investigate the usage of the Raman spectroscopy in combination with DFT calculations as a complementary method to the X-ray techniques. The potential to discriminate structural differences in polymorphic crystalline forms was tested on the set of three compounds.

The studied set consisted of methacrylamide (A), piracetam (B), and 2-thiobarbituric acid (C) as model molecules representing typical organic systems forming polymorphic crystals (fig. 3.1). Differences between polymorphs range from a small change in the crystal packing (piracetam) over conformational (methacrylamide) and tautomeric (2-thiobarbituric acid) variations. The geometries of the studied systems acid (fig. 3.1) were taken from the Cambridge Crystallographic Database [62].



A1 **A2** methacrylamide

B piracetam

C1 **C2** 2-thiobarbituric acid

Figure 3.1: Methacrylamide (A1 and A2, cis and trans conformers), piracetam (B), and 2-thiobarbituric acid (C1 and C2, enol and keto tautamers)

The Raman spectra were calculated using a cluster of molecules mimicking the crystal and Gaussian atomic orbitals. The crystal cell was propagated to a 3×3×3 packing geometry. Then clusters of a molecule and neighbouring molecules closer than 4 Å were created with our own software. The geometries of the clusters were optimized in normal mode vibrational coordinates [63] with $\omega_{max} = 300$ cm^{-1}.

Force field and polarizability derivatives of the clusters were calculated by the Gaussian 09 program [64] and transferred [3, 4] to the crystal cell. The BPW91 [65] functional with the 6-31G and 6-31G** basis sets and the B3PW91 [66] functional with the 6-311++G** and 6-31G basis sets for the central and neighbouring molecules were used.

For the cluster force fields, the crystal vibrational frequencies were calculated. Theoretical spectra were simulated with the Lorentzian bands, and the full width at half-maximum was set to 10 cm^{-1}, while applying the Boltzmann temperature correction (eq. 2.38 page 13).

Circular Dichroism Spectra of Monensin Complexes

Monensin is a natural antibiotic produced by Streptomyces cinnamonensis [67, 68, 69]. Monensin has the ability to form complexes with certain monovalent metal cations. The antibiotic acts as a monoanion through deprotonated carboxylic function, making an overall neutral charge of the complex formed and easily penetrate bacteria's cell membranes. Inside the cell dissociation processes occur and leads to disturbance of pH and metal ion equilibrium. The following changes activate a variety of further event causing cell death. Our aim was to investigate the behaviour of these complexes using ECD spectroscopy.

X-ray structures of monensic acid (MonH $\times$ H_2O - MONSNI) [70] and its monovalent metal complexes MonM (M = Li^+- MIPSIO [71], Na^+-DEYGAQ [72], K^+- FECROU10 [73], Rb^+- RITLIQ [74], Ag^+- MONSIN10 [75]) were used as starting geometries. The structures were fully optimized with the Gaussian 09 program [64] using the B3LYP [76] functional and the conductor-like polarizable continuum solvent model (CPCM) [77, 78] to include the methanol solvent environment. CAM-B3LYP, invented to improve B3LYP, B3PW91, LC-WPBE, and WB97XD functionals were also applied, but did not give better results than the standard B3LYP.

Alternatively to the previous full optimization, X-ray geometries, partially optimized in the normal mode coordinates were used as well. Normal modes with frequencies $|\omega_i| <$ 300 cm^{-1} were fixed [63]. The partial optimization corrected bond lengths and angles of the hydrogen atoms, determined with a big error or completely missing in the crystal structures. The 6-311++G** basis set was used for the carboxyl group atoms, the MWB28 pseudopotential [79] and basis set were used for silver and rubidium atoms, and the 6-31G** basis set was used for the rest. For the optimized structures, UV and CD spectra were calculated at the TDDFT[80]/CPCM level. For each system, 100 electronic excited states were obtained to cover the experimentally observable spectral range.

Chiral Sensing of Aminoacids by Europium Complexes

Chiroptical spectroscopy of lanthanides sensitively reflects their environment. For example, they can be useful as probing agents for protein structure. In our study, we focused on Circularly Polarized Luminiscence of europium complex ($[Eu(DPA)_3]^{3-}$) induced by amino acids (fig. 3.2), monitored by ROA spectroscopy. Our aim was to correlate free energy profiles (obtained by MD) of differently charged alanine and histidine forms with obtained spectroscopic data.

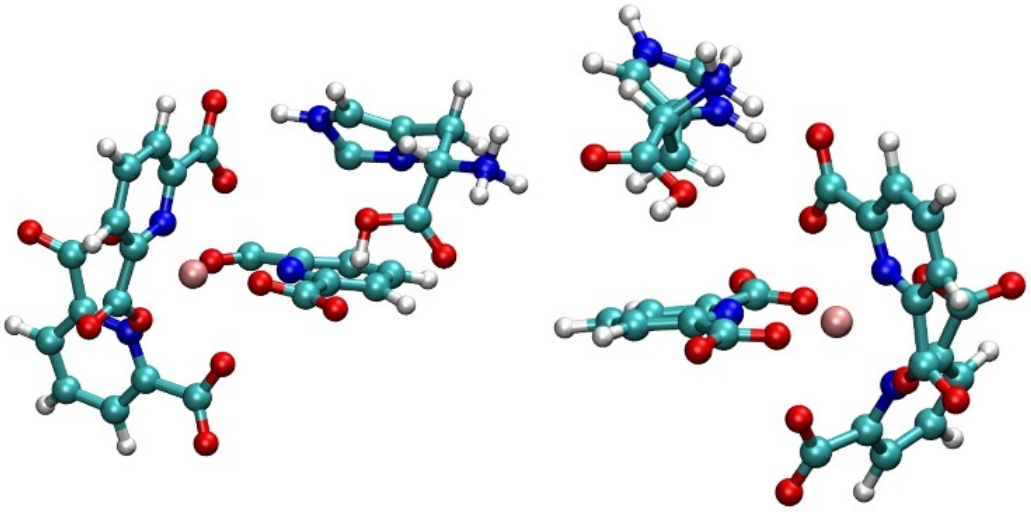


Figure 3.2: Geometries of L-histidine complexes with the Λ(left) and Δ (right) form of $[Eu(DPA)_3]^{3-}$.

Interaction of L -histidine with the Λ and Δ $[Eu(DPA)_3]^{3-}$ forms was studied using molecular dynamics and the Amber14 [81] software suite. Variously protonated histidine and alanine amino acids with the $[Eu(DPA)_3]^{3-}$ ion were placed inside a cubic box (30 Å a side) filled with water molecules. The energy of the system was minimized, and the geometry was equilibrated during a 1 ns NVT dynamics, using an integration step of 1 fs, temperature of 300 K, and the GAFF [82] (DPA ligands) and Amber14SB [83] (alanine and histidine) force fields. Europium force field parameters were taken from ref. [84]. Then, the distance between carbonyl oxygen of histidine/alanine and europium was decreasing (from 14 to 4 Å) in 1 Å steps using 1 ns long constrained dynamics runs and a harmonic potential force constant of 2 kcalÅ^{-2}mol^{-1}. From resultant histograms the potential of mean force was obtained using the weighted histogram analysis method (WHAM) [46, 47].

3.1.2 Proteins

Proteins are important chiral biomolecules and the chiroptical techniques are very sensitive to their secondary structure. In this work, a set of globular proteins were studied (table. 3.1). The set ranged from 783 atoms in case of bovine pancreas insulin to nearly 10000 for human serum albumin. They also differ in their secondary structure, ranging from mainly α-helical human serum albumin and equine myoglobin to concavalin A containing mainly β-sheet. The proteins were studied in their native forms. In the case of insulin, we studied also amyloid fibrills. Poly-L-glutamic acid aggregates were studied as well.

The simulation procedure consisted of multiple step. In the beginning, X-ray structures were taken from PDB structure database. The crude geometries were refined using the tleap

Table 3.1: Characterization of the globular proteins taken to experimental and computational study

protein name	PDB ID	Resolution (Å)	$N_{res.}$	N_{atoms}	α-helix	β-sheet
bovine pancreas insulin[85]	2A3G	2.25	51	783	66 %	9 %
bovine α-lactalbumin[86]	1HFZ	2.30	124	1960	43 %	9 %
hen egg-white lysozyme[87]	3WPJ	2.00	129	1960	41 %	10 %
human lysozyme[88]	1lZ1	1.50	130	2021	39 %	12 %
equine myoglobin [89]	3LR7	1.6	153	2387	73 %	0 %
jack bean concavalin-A[90]	1NLS	0.94	237	3566	3 %	46 %
human serum albumin[91]	1UOR	2.80	585	9161	46 %	0 %

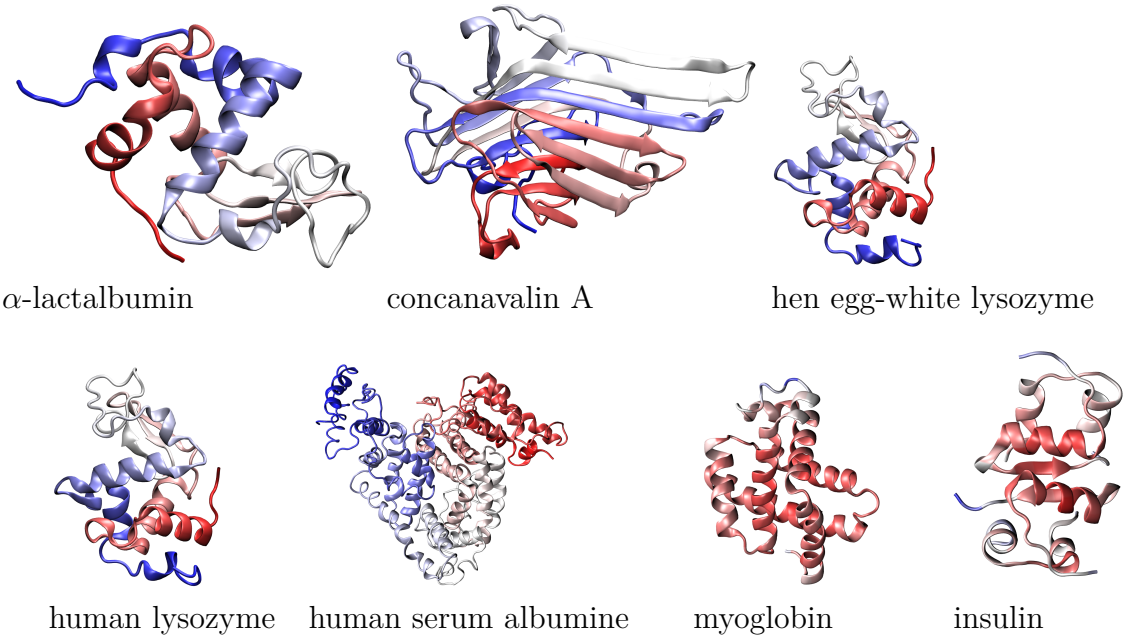


Figure 3.3: Structures of studied globular proteins

program of Amber software suite [92] when missing heavy and hydrogen atoms were added. The protein chains were divided into overlapping fragments, containing 4 amides, i.e. three aminoacids and $CH_3 - NHCO -$ and $- NHCO - CH_3$ terminal groups. Special fragments were created for the disulfide bridges.

In the case of α-lactalbumin, contact fragments were created for all atomic pairs closer than 2.4 Å, which were not covalently bonded. These two aminoacid fragments were caped as the previous types of fragments. However, the contacts fragment had only a minor effect on the spectra, therefore the usage of "contact" fragments was abandoned in other simulations. Our own software written in Fortran (fragmentdna.f) was used for the automatic fragmentation.

The structure of each fragment was partially optimized at the B3PW91/6-31++G** level [66]. The water solvent environment was simulated using COSMO (CPCM) [77, 78] as implemented in Gaussian09, revison D.01 program [64]. The partial optimization was

carried out in normal mode coordinates in order to restrain modes bellow 100-300 cm^{-1}. The harmonic force field, atomic axial, atomic polar, Raman and ROA tensors [93, 58] were transferred from small fragments to the big protein using the Cartesian Coordinate Transfer (CCT) procedure [3, 4].

Vibrational frequencies, and the Raman and ROA intensities were generated using usual procedure [58, 94, 95]. The intensities were convoluted with Lorentzian functions of 10 cm^{-1} bandwidth with temperature correction (T=298 K).

Another geometry model relied on molecular dynamic (MD). The simulation was carried out in the Amber program suite using Amber 2014 force field [83]. The studied proteins were soaked into cubic periodic water box with lengths 64 Å for α-lactalbumin, 58 Å for human and hen egg-white lysozyme, 68 Å for concanavalin A and 90 Å for human serum albumin. Counterions were added to neutralize the charged protein.

The MD was performed using NVT ensemble, temperature 298 K and 1 fs integration step. Molecular mechanic minimization was followed by 1 ns equilibration dynamic, which provided starting structure for 10 ns production run. Snapshots were taken every 10 ps. An average nuclear density was calculated and the most closely matching snapshot was chosen using the PVS (Parallel Variable Selection) algorithm [96]. The best geometry was then used in the same way as the X-ray one. To use also the other MD geometries, the VOA tensors and force fields were transferred to 1000 snapshots and the obtained spectra were averaged.

Amyloid fibrils are more complex than globular proteins. Their structure is often unknown; therefore we explored several possible geometries for fibrillar systems. The first tested system was the aggregates of poly-L-glutamic acid (PLGA). Assumed set of geometries was based on simple basic primitive cell (fig. 3.4), which was consisted of two antiparallel 15-amide protonated polyglutamic acid (PLGA) strands $[Ac - Glu_{14} - NH_2]_2$ with acetyl and NH_2 groups on N- and C-terminus. The system geometry was placed in a monoclinic periodic box (dimensions 66.00 Å, 9.77 Å, 8.07 Å, $\alpha = 105°$). Edges of the box were in the x-dimension filled with 15 water molecules. This starting geometry was base on previous X-ray studies [97, 98].

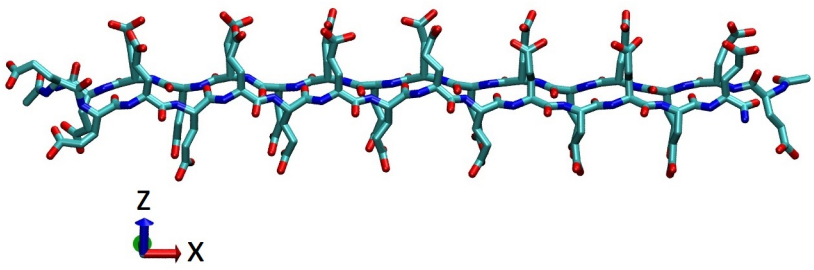


Figure 3.4: Primitive cell of poly-L-glutamic acid (PLGA) used for MD simulation

MD simulation was run in Gromacs program [99] using the Amber03 force field [100]. At first, simulated annealing was performed, starting at 1000 K and ending at 300 K, as an NVT ensemble. 19 independent annealing runs were performed with simulation time form 0.7 to 1.6 ns followed by 10 ns equilibration runs.

Obtained geometries were used for larger aggregate structure generation. Four aggregate types were created. The primitive cell was expanded in sheet stacking direction (z-axis), forming stacks of 2 or 4 two-stranded antiparallel β-sheets (systems A and B, fig. 3.5), or in the interstrand H-bonded direction (y-direction) to consider larger, 4 stranded sheets (systems C and D, fig. 3.5). The largest system D contained three 4-stranded antiparallel sheets and was used for final spectra generation as the biggest and most relevant model.

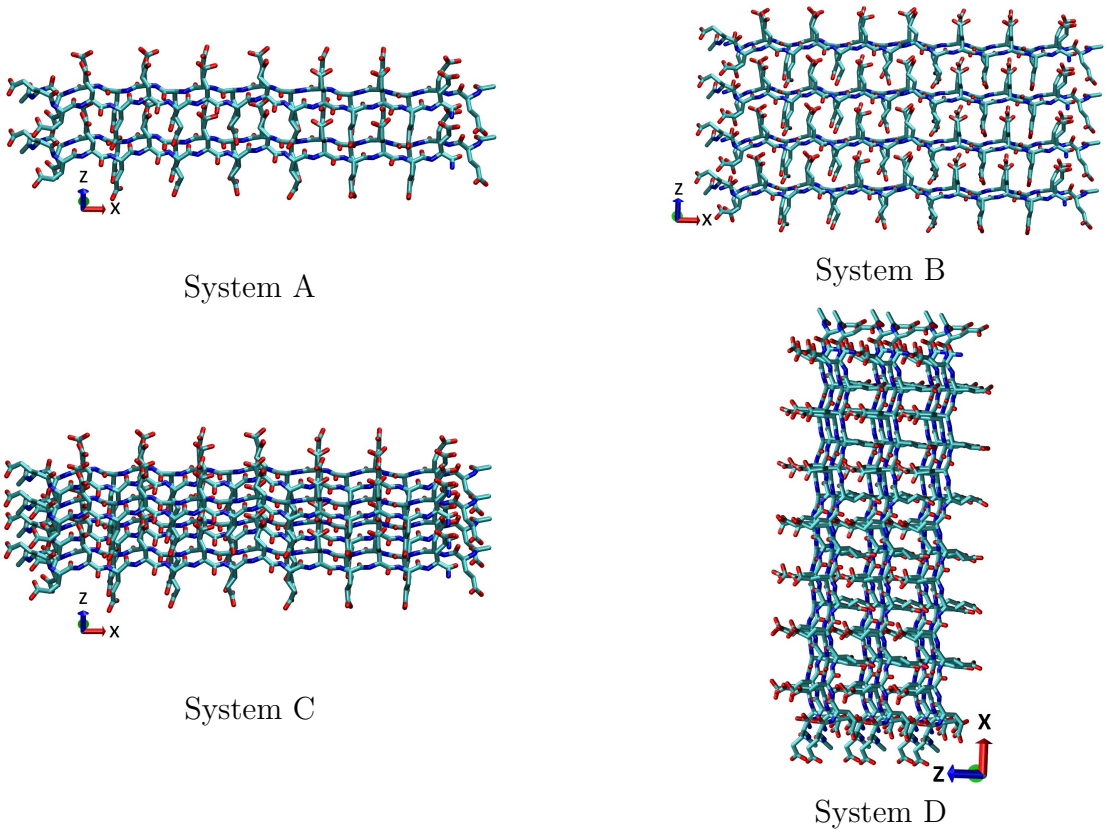
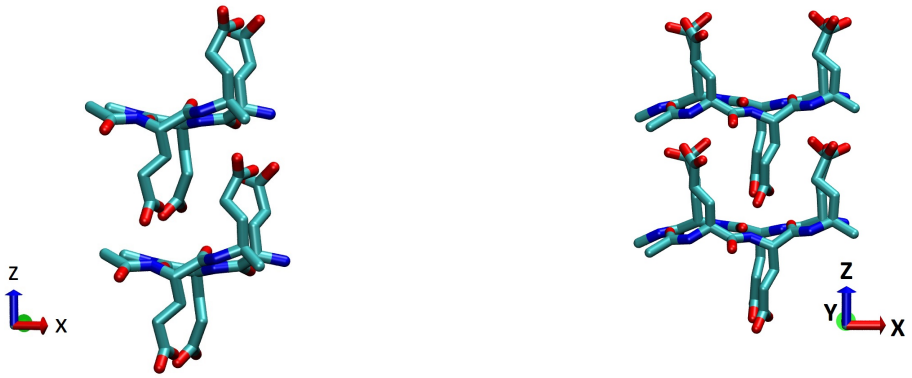


Figure 3.5: Structures of aggregate systems A and B created by a propagation of primitive cell in the z axis and aggregate systems C and D created via propagation of primitive cell in the y and z axes

The IR and VCD spectra of the model systems were obtained from the harmonic force field, atomic polar and atomic axial tensor calculated for smaller fragments (**F8** and **F12** in fig. 3.6), containing 8- or 12-amide structures. To include different sidechain conformations, fragments were chosen from 4 MD snapshots. The fragments were partially optimized using normal mode coordinates (300 cm^{-1} constrain limit). The Gaussian 09, revision D.01 program [64] and the BPW91/6-31G** method [65] were used. The solvent environment was simulated using CPCM dielectric solvent model [77, 78]. The VOA tensors were then

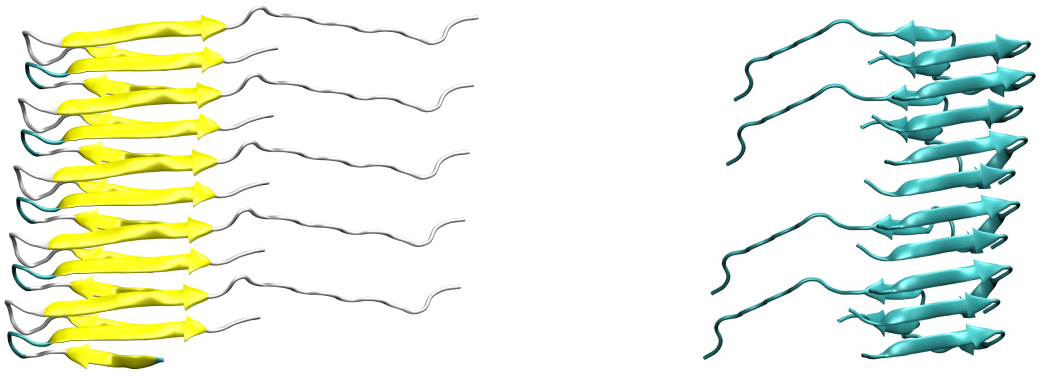
transferred [3, 4] to the sets of aggregates molecules. Final spectra were obtained as an average of all snapshots. Deuteration of peptides were simulated, because the experiment was carried out in deuterated water (D_2O).



F8 **F12**

Figure 3.6: Fragments **F8** and **F12** used for molecular properties tensor transfer

The second studied fibril structure was insulin. We proposed two basic geometries, so called β-roll and β-helix (fig. 3.7). The β-roll was based on a β-roll protein (PDB ID 1VH4), i.e. insulin torsion angles were set to those in residues A256-A306 in the protein. The polar groups were protonated to correspond to the experimental pH of 2.5-3.1. An analogous procedure was used for the β-helix, where the torsion angles followed residues A113 to A163 in a β-helix protein (PDB ID 1DAB). The terminal part of insulin B chain (B22-B30) was kept in the native β-sheet conformation.



β-roll β-helix

Figure 3.7: The β-roll and β-helix structures used to model the insulin fibril

Raman and ROA spectra of the fibrous insulin forms were generated using the same methods as in the case of globular proteins. To include protein flexibility and temperature fluctuations, molecular dynamics (MD) was performed within the Amber10 [92] software suite. A trimer was made from the X-ray derived insulin geometries and soaked in rectangular box (80 Å×30 Å×50 Å) filled with water molecules. Only the middle molecule in the trimer

could move during 1 ns equilibration MD run, performed with the Amber03 [100] force field, as NVT thermodynamic ensemble, at 300 K, using 1 fs integration step and periodic boundary conditions. During a production run (8.65 ns) 865 snapshot geometries were taken (each 10 ps). For an MD insulin geometry best matching the average nuclear density generated during the MD run spectral parameters were calculated as for the X-ray structures; then the tensors were transferred to remaining snapshots and the resultant spectra averaged.

Another MD run was performed with modified Tinker program [101], where the X-ray based insulin geometry (β-roll) was placed in rectangular box (12 Å $\times$ 110 Å $\times$110 Å) filled with water molecules and the helical periodic boundary conditions [102] were applied to allow for a minor twist between neighbouring peptide chains (Fig. 3.8). The Amber99 [103] force field was used in an NVT run with 1 fs integration time. During 0.1 nm 100 snapshots were selected. For an average geometry vibrational spectral parameters were calculated and transferred on all the other snapshots. As in the previous case, resultant spectra were obtained as a plain average.

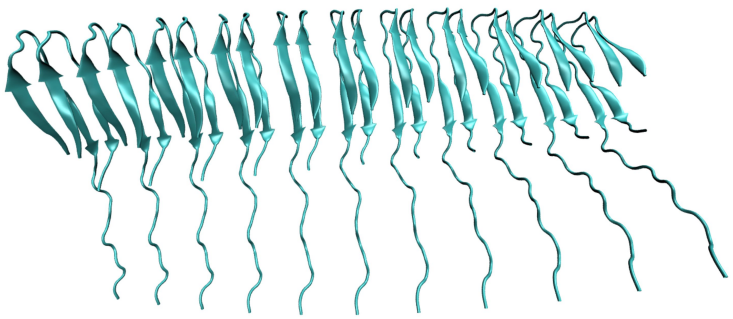


Figure 3.8: The twisted structure of insulin fibril

3.2 Methodical Projects

3.2.1 Helical Periodic Boundary Conditions

Helical symmetry is often appearing in nature and thus it is desirable to use it also in molecular dynamics simulations. An infinite helical periodicity can save a significant amount of computer time. But the standard simulations with the usual periodic boundary conditions (PBC) can not be used for simulation of such systems. Therefore we decided to investigate a new algorithm containing infinitely propagated helicity and implemented it into the commonly used Tinker [101] molecular dynamics software.

The helical periodicity in the vertical direction z is introduced for a central part (zone 1) of the MD simulation box, in a cylinder of radius r_1 (fig. 3.9). This part contains the molecule of our interest and part of the solvent. During the replication, the cylinder is translated, i.e. shifted by the box dimension in a direction of the z-axis (b_z), and rotated by φ_0. A transition zone (zone 2) is defined as a space between the above mentioned cylinder (radius r_1) and another cylinder of radius r_2. In this zone, an intermediate transformation takes place, as defined via $\boldsymbol{U}(\varphi)$, where $\varphi = f\varphi_0$, and f denotes a transition function.

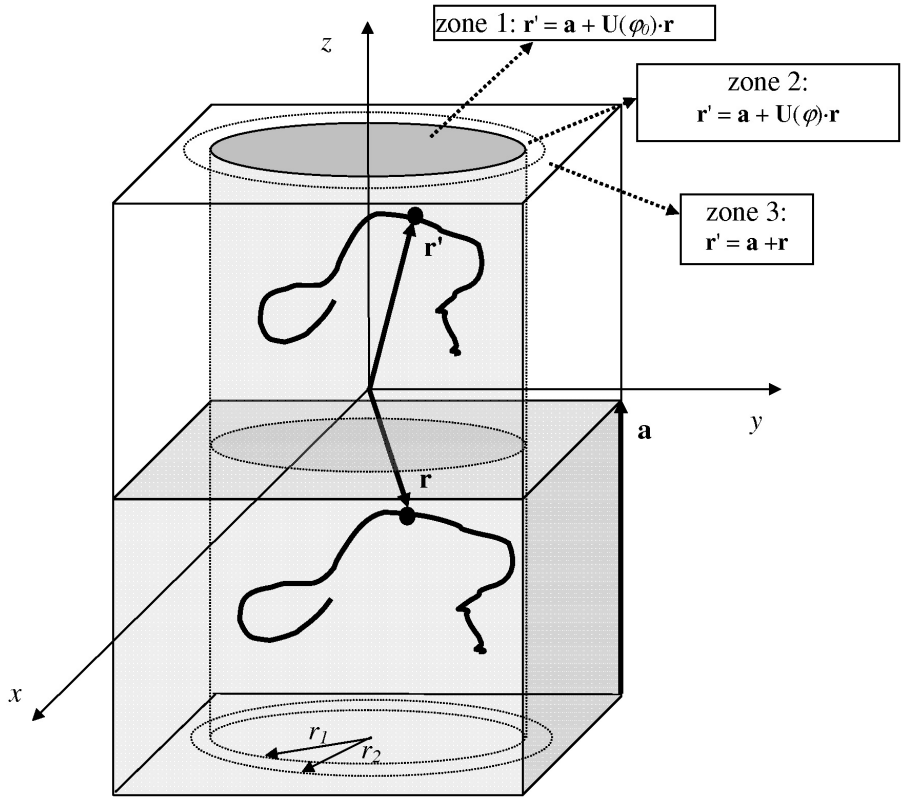


Figure 3.9: The definition of the helical periodic boundary conditions.

The transition function causes a damping of helicity, i.e. $f = 1$ at the wall of the inner and 0 at outer cylinder. As for similar potential-adjusting functions, e.g. the van der Waals or Coulomb interaction cutoffs [104, 105], it is necessary to have f simple and smooth, which restricts its choice to a sigmoidal dependence, e.g. defined as:

$$f(p) = \frac{1}{2}\left(1 - (2p-1)\sqrt{\frac{d+1}{d+(2p-1)}}\right) \tag{3.1}$$

where $p = \frac{r_{xy}-r_1}{r_2-r_1}$, r_{xy} denotes the distance from the z-axis, and d = 0.1 is a steepness parameter (fig. 3.10).

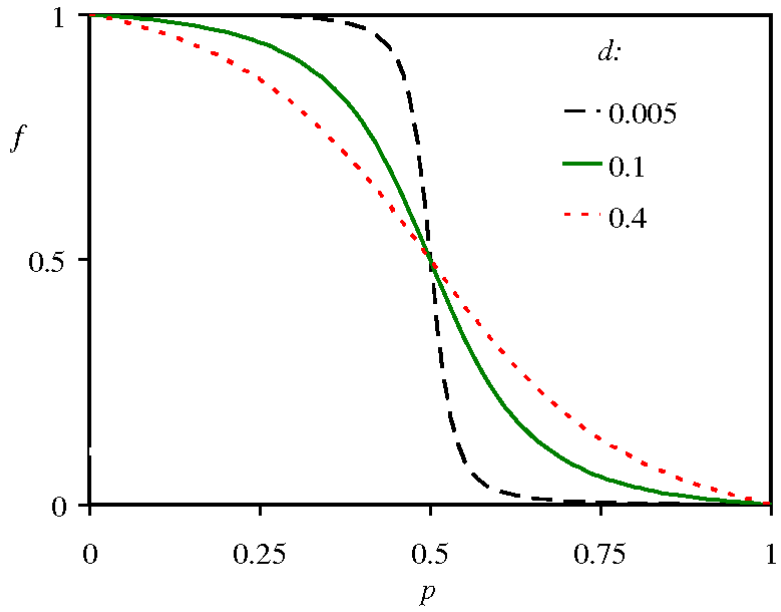


Figure 3.10: The shape of damping function for different parameters d

The helical periodic boundary conditions were implemented in the Tinker 6.2 [101] program. The Amber99 [103] force field was used. All simulations were performed in NVT conditions and default Tinker parameters, using the default 9 Å cut off distance both for van der Waals and Coulomb interactions. The cutoff was combined with potential switching and shifting, using 5^{th}-degree multiplicative and $7^{th}-$degree additive (Coulomb only) switching functions as implemented in the Tinker 6.2 program [101]. The Beeman [106] propagation scheme, 1 fs integration time step, temperature of 298 K, and kept fixed during the dynamics.

To test the algorithm various systems were chosen from the simple polypeptide chains to insulin amyloid fibril. The first test system was $[\mathrm{Ala12}]_\infty$, i.e. infinite polyalanine chain in the α-helical conformation. To allow for a realistic motion and fluctuations, 12 alanine residues and 813 water molecules were kept in the 37.2 Å × 37.2 Å × 18.56 Å (helical axis) box, with

$r_1 = 8$ Å and $r_2 = 17$ Å. Initial peptide backbone torsion angles ($\varphi = -60°$, $\psi = -45°$, $\omega = 180°$) corresponded to a standard α-helical geometry [107] After a minimization and an equilibrium dynamics, production run lasting 2 ns was performed.

The second test system was $[\mathrm{Pro}_{10}]_\infty$, i.e. the polyproline chain in the polyproline II conformation. The system was created as for the α-helix. Ten proline residues and 856 water molecules were put to a box 30 Å × 30 Å × 29.9 Å. The initial twist angle φ_0 was changed within 100°-120°, and initial backbone torsion angles [107] were set to $\varphi = -75°, \psi = 150°, \omega = 180°$. Values of $r_1 = 9$ Å and $r_2 = 13.5$ Å were used. Total time of each simulation was 2 ns.

The largest system involved protonated insulin molecule (785 atoms), 6 Cl^- ions and 4593 water molecules in 110 Å × 110 Å × 12 Å box, with $\varphi_0 = 0°, 3°,\ 6°$, $r_1 = 43$ Å and $r_2 = 53$ Å. An initial geometry was based on the β-roll protein (PDB ID 1VH4). After energy minimization and subsequent equilibration the system was let to develop for 300 ps using NVT ensemble at 298 K with 1 fs integration step.

3.2.2 Transfer of Frequency-Dependent Polarizabilities

Transfer of spectroscopic properties from small fragments to bigger system is an excellent tool to simulate vibrational spectra of big molecules [3, 4]. However, the situation is quite different for the case of electronic spectra. Changes of electronic states are associated with a massive rearrangement of the electronic cloud, which cannot be easily associated to individual atoms. In this study, we explore the Transferring of the Frequency-Dependent Polarizabilities (TFDP) as a way to extend the accurate ab initio (e.g., TDDFT) approaches to larger systems. In this method, the polarizability is associated with an atom, bond, or a fragment (chromophore), and it can be transferred to a similar unit in the modelled system.

The classical transition dipole coupling (TDC) method was used to compare with TFDP. In TDC method, the transition energies (e_i) and electric transition dipole moments ($\boldsymbol{\mu}_i$) obtained by TDDFT for a monomer were transferred to each chromophore in a dimer or larger system. Energies, dipole ($D = \boldsymbol{\mu}_{gk} \cdot \boldsymbol{\mu}_{kg}$) and rotational ($R = \mathrm{Im}(\boldsymbol{\mu}_{gk} \cdot \boldsymbol{m}_{kg})$) strengths for a transition $g \to k$ in the whole system are obtained after the diagonalization of the Hamiltonian.

Diagonal Hamiltonian elements are equal to the transition energies, $H_{ii} = e_i$, and off diagonal elements correspond to the dipole–dipole interaction, $V_{ij} = \frac{1}{4\pi\varepsilon_0} \frac{r_{ij}^2 \boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j - 3\boldsymbol{\mu}_i \cdot \boldsymbol{r}_{ij} \boldsymbol{\mu}_j \cdot \boldsymbol{r}_{ij}}{r_{ij}^5}$. Then $\boldsymbol{\mu}_{gk} = \sum_j c_j^k \boldsymbol{\mu}_j$ and $\boldsymbol{m}_{gk} = \frac{i}{2} \sum_j c_j^k \omega_j \boldsymbol{r}_j \times \boldsymbol{\mu}_j$, where c_{jk} are elements of the eigenvectors, ω_j transition frequencies and $\boldsymbol{r}_j$ are electric dipole moment positions. In the case of our computations, we positioned the dipoles in the geometrical center of the HNCO groups (N-methylacetamide) or in the mass center (porphyrin).

In the case of TFDP, we obtained the frequency-dependent polarizabilities $\boldsymbol{\alpha}$ and $\boldsymbol{G'}$ by TDDFT for every monomer instead of dipoles μ_i. Total polarizabilities of the system were acquired as:

$$\begin{aligned}
\alpha_{\alpha\beta} &= \sum_i \alpha_{i,\alpha\beta} + \sum_{j\neq i}[\alpha_{i,\alpha\chi} T_{ij,\chi\delta}\alpha_{j\delta\beta} + \\
&+ \frac{1}{3}(A_{i\alpha,\chi\gamma} t_{ij,\chi\gamma\delta}\alpha_{j,\delta\beta} - \alpha_{i,\alpha\chi} t_{ij,\chi\delta\gamma} A_{j\beta,\delta\gamma}) + \frac{1}{c^2} G'_{i\alpha\chi} T_{ij,\chi\gamma} G'_{j\beta,\gamma}] \qquad (3.2)\\
G'_{\alpha\beta} &= \sum_i \left(G'_{i\alpha,\beta} - \frac{\omega}{2}\epsilon_{\beta\gamma\delta} r_{i,\gamma}\alpha_{i,\delta\alpha}\right) - \frac{\omega}{2}\sum_{i\neq j}[\epsilon_{\beta\gamma\delta} r_{i,\gamma}\alpha_{i,\delta a} T_{ij,ab}\alpha_{j,b\alpha} + \\
&+ \frac{1}{3}\epsilon_{\beta\epsilon\delta} r_{i,\epsilon}(\alpha_{j,\alpha a} t_{ij,\lambda\gamma a} A_{i\delta,\lambda\gamma} - \alpha_{i,\delta a} t_{ij,a\lambda\gamma} A_{j\alpha,\lambda\gamma}) - \frac{2}{\omega} G'_{i\gamma,\beta} T_{ij,\gamma\delta}\alpha_{j,\delta\alpha}] \qquad (3.3)
\end{aligned}$$

$\boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$ denotes vector connecting the monomers i and j and the distance tensors are as $T_{ij,\alpha\beta} = \frac{3r_{ij,\alpha}r_{ij,\beta} - \delta_{\alpha\beta}r_{ij}^2}{4\pi\epsilon_0 r_{ij}^5}$ and $t_{ij,\alpha\beta\gamma} = \nabla_{i\gamma} T_{ij,\alpha\beta}$. The total polarizabilities are not only the sums of the individual components, but also contain terms responsible for mutual monomer polarization.

As an extension of the TFDP approach, monomer polarizabilities were redistributed to to individual atoms. Thus, the summation in equations 3.2 and 3.3 proceeded over the atomic indices. This method is called $\mathrm{TFDP_d}$.

Another approach is the case, when we calculated monomer polarizabilities in the presence of atomic partial charges to include the influence of the rest of the simulated system. We refer such method as $\mathrm{TFDP_e}$. Combination with previous one, referred as $\mathrm{TFDP_{de}}$ is also possible.

Model systems to test the performance of TFDP included N-methylacetamide (NMA), porphyrin dimers, and a porphyrin cluster derived from the photosystem reaction center. The geometry of one NMA molecule was optimized by energy-minimization using the Gaussian09, rev. D01 program suite [64]. The B3LYP [76]/6-311++G** approximation level was chosen as a default. Other methods (CIS [108], CAM-B3LYP [109], LC-wPBE [110] and TDHF [108]) with the same basis set were performed for control calculations. Electrostatic charges of NMA for the TFDP method were obtained by the Mulliken (MU) population analysis [111] and by the Merz-Singh-Kollman (MS) [112], CHelp (CHelp) [113], and Hu, Lu, and Yang (HLY) [114] electrostatic field fitting schemes at the B3LYP/ 6-311++G** level. The dimers were formed by rotation of one NMA molecule by 20° about an axis perpendicular to the NMA plane and shifting it by 4.5, 6.5, and 8.5 Å in a direction perpendicular to that plane (fig. 3.11).

For the dimers, the frequency-dependent polarizabilities were calculated from the transition moments and excitation energies obtained from the Gaussian output by the SOS ap-

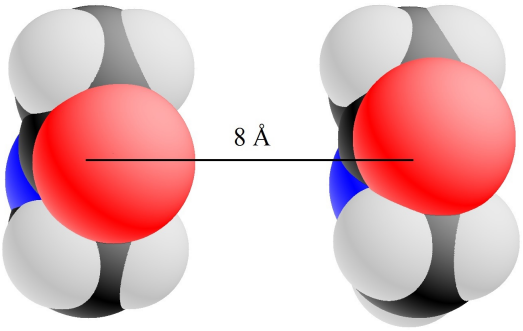
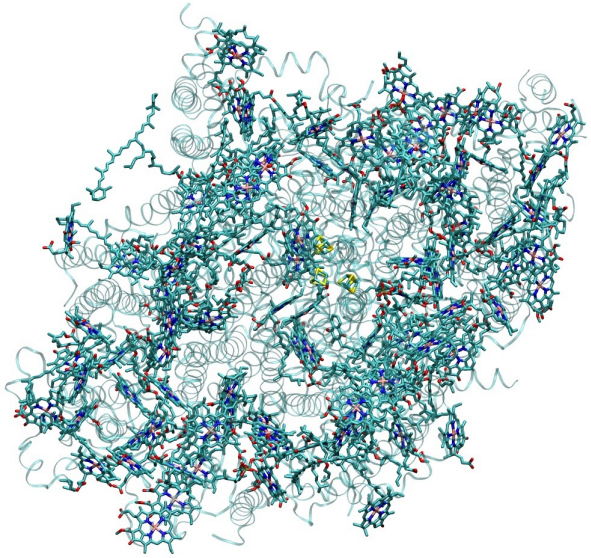
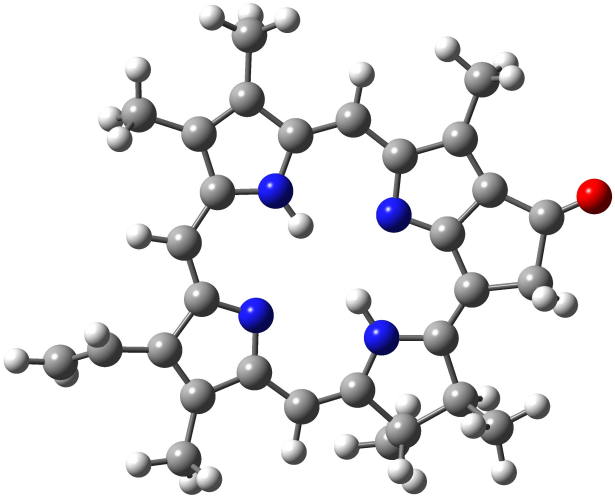


Figure 3.11: Dimer of N-methylacetamide, distance 8 Å



Photosystem



Simplified Porphyrin monomer

Figure 3.12: Photosystem and simplified porphyrin monomer model

proach [115]. The frequency-dependent polarizability components were saved for frequencies covering the entire excitation range incremented in 1 nm steps while applying a frequency uncertainty parameter (Γ) of 10 nm to match the usual experimental band broadening. For dimers, absorption and electronic circular dichroism spectra were calculated using the time-dependent density functional theory [80].

The cctn program originally developed for vibrational spectroscopic parameters was adapted to enable transfer of the frequency-dependent tensors. The transfer is based on the best (least-squares distance method) overlap between the source and target chemical entities and a unitary (rotation) transformation of Cartesian tensor indices [3, 4]. As a more advanced model, we investigated the cyanobacterial photosystem I (fig. 3.12) [116] for which X-ray coordinates containing 96 porphyrin chromophores were available (PDB ID 1JB0).

A monomer porphyrin molecule (fig. 3.12) was generated using the 1JB0 coordinates. Relaxation of the geometry was enabled within the constrained (ω_{max} = 300 cm^{-1}) normal mode optimization (NMO) [63] at the B3LYP/6-311++G** level. Transition dipole moments

and frequency-dependent polarizabilities were estimated as for the NMA. Additionally, two porphyrin dimers were constructed, for which the TFDP and TDC results could be compared together with the TDDFT benchmark. The geometries were based on the 1JB0 positions of porphyrins comprising magnesium atom numbers 23357 and 23402, 23.3 Å apart (dimer 1), and magnesia numbers 17407 and 20451, 6.34 Å apart (dimer 2).

The simplified monomer units, without the central Mg, ion were used in the computations. To estimate the computational time, we created larger and smaller porphyrin oligomers by adding or deleting other molecules in the 96-mer. For all systems, the spectra were generated by the TFDP method [117] using Γ 10 nm bandwidth. Likewise, the TDDFT and TDC spectra were generated from the calculated dipole and rotational strengths using Lorentzian bands (full width at half-maximum of 10 nm).

4 Results

4.1 Small Biomolecular Systems

4.1.1 Raman Spectroscopy of Polymorphs

The obtained results are more deeply explained in the publication (appendix A). We showed how the polymorphic differences were manifested in the Raman spectra. The computational models explained the observations. We concluded that the Raman spectroscopy combined with computational modelling can serve for structural analysis of organic compounds in polymorphic forms.

4.1.2 Circular Dichroism Spectra of Monensin Complexes

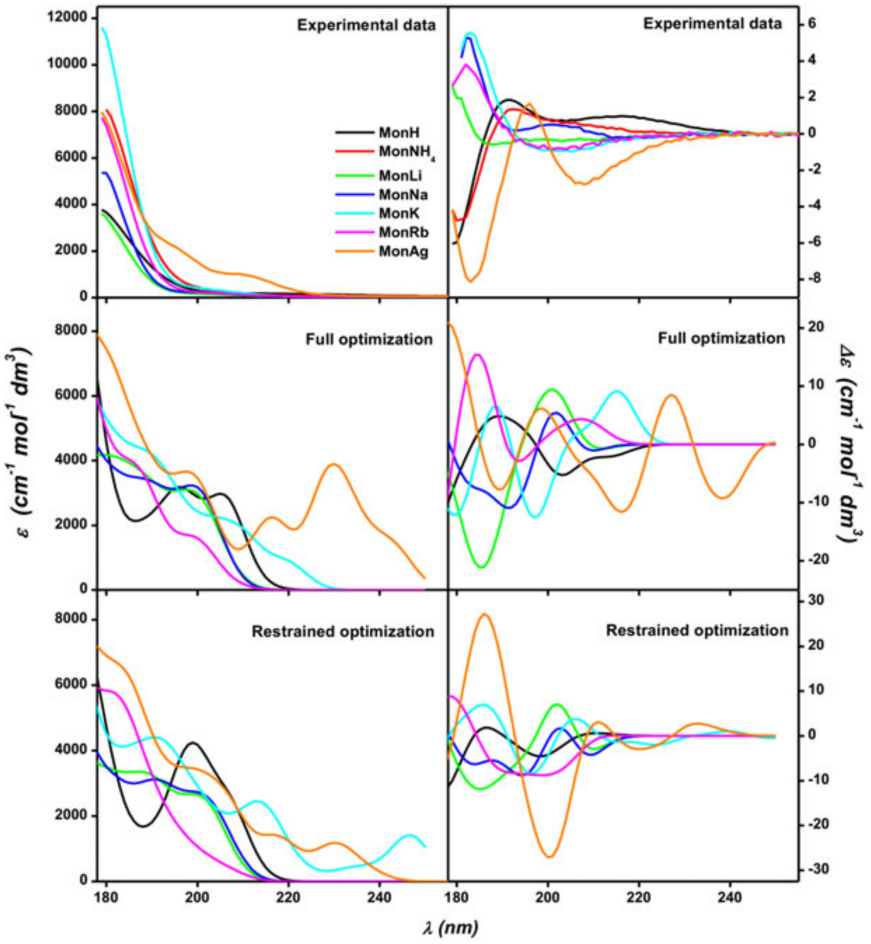


Figure 4.1: Experimental and calculated absorption (left) and CD (right) spectra of MonH and MonM (M =$Li^+, Na^+, K^+, Rb^+, Ag^+$).

The results are more deeply explained in appendix B. The calculated absorption and CD spectra (fig. 4.1) provided a better understanding of the behaviour monesin metal complexes,

although calculations do not reproduce the experiment quantitatively. We confirmed that the metal ions can induce specific CD shapes under a minimal change of conformation. The limited accuracy was explained by the complexity of the system and accumulation of computational error originating from the DFT and TDDFT approximations, solvent model, and lack of dynamics in the modelling. Several interesting trends were observed, such as the difference in the behaviour of the Ag^+ ion if compared to the others.

4.1.3 Chiral Sensing of Aminoacids by Europium Complexes

All results are discussed appendix C. Most interestingly, measured CPL spectra of $[Eu(DPA)_3]^{3-}$ complex with aminoacids (fig. 4.2) were correlated to complexation energy profiles obtained by WHAM method, both for all alanine and histidine forms (fig. 4.3).

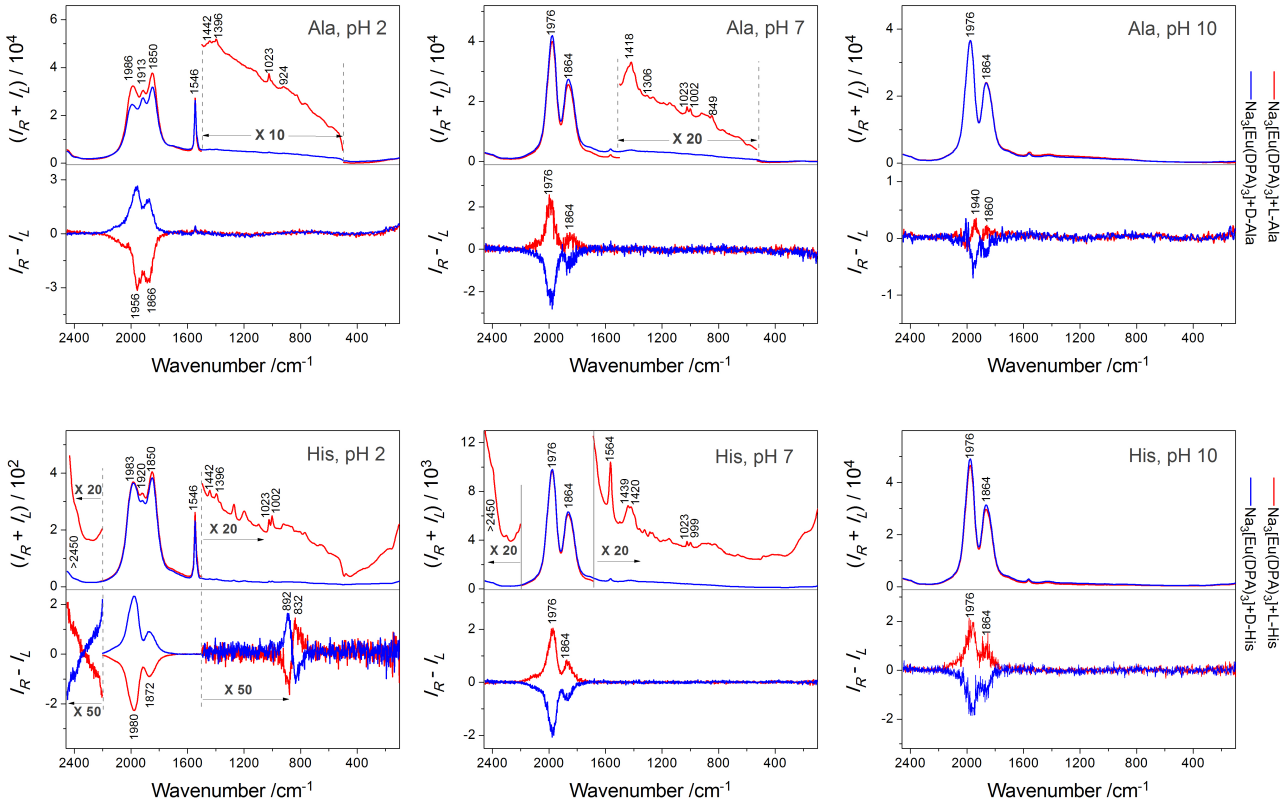


Figure 4.2: $[Eu(DPA)_3]^{3-}$ Raman and ROA spectra under presence of L- and D-alanine and L- and D-histidine at pH 2, 7 and 10

We hypothesised that complex formation shifted in favour the enantiomeric equilibrium and therefore we obtained the CPL signal. The spectra exhibit two main trends, the signal intensity decreases with increasing value of pH, and the intensity is higher for histidine complexes if compared with alanine.

The changes were correlated with obtained WHAM profiles. They exhibited larger differences for the enantiomers (Δ and Λ) in the case of histidine than for alanine (fig. 4.3). The energy profile differences also decreased with increasing pH value.

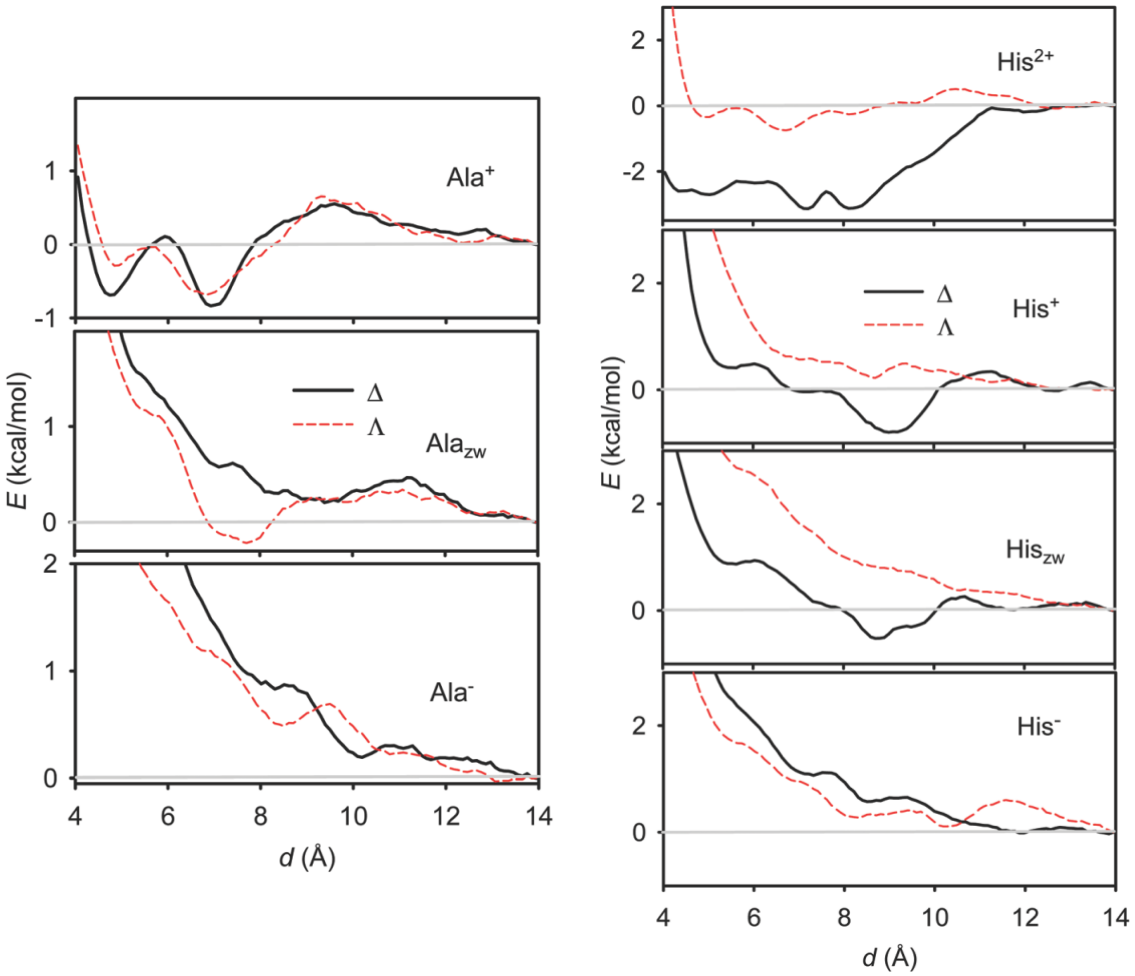


Figure 4.3: Calculated free energy profiles for different L-alanine (left) and L-histidine (right) forms interacting with Δ and Λ enantiomers of $[Eu(DPA)_3]^{3-}$.

4.2 Vibrational Optical Activity of Proteins

4.2.1 Globular Proteins

The experimental and simulated Raman and ROA spectra for the set of 5 globular proteins (bovine α-lactalbumin, hen egg-white and human lysozymes, jack bean concanavalin A, human serum albumin) are discussed in appendix E. All simulated spectra reasonably well reproduced the experimental ones. We reproduced the differences between secondary structures in the concanavalin A and human serum albumin (fig. 4.4).

For two types of lysozymes (4.5) we also reproduced a difference in peak ratios in the amide III extended region. The hen egg-white lysozyme exhibits much sharper negative ROA

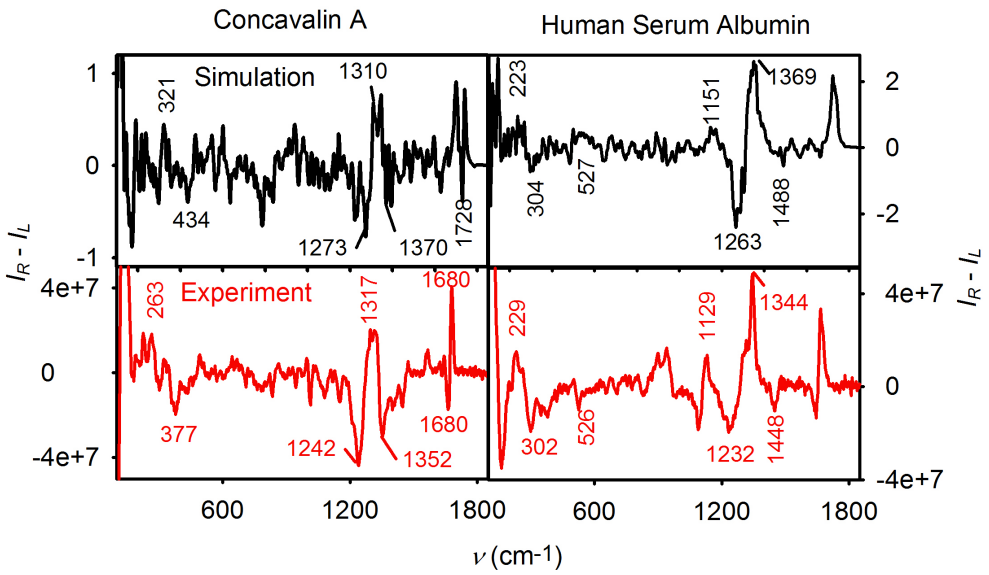


Figure 4.4: Simulated and experimental ROA spectra of concanavalin A and human serum albumine

band at 1250 cm^{-1} and a positive one at 1304 cm^{-1} than the human one. A visualization of the vibrational normal modes reveals that the intensity changes can be attributed to the different amino acid content. The human lysozyme has twice as many tyrosine residues as hen egg-white lysozyme. Their vibrational modes are coupled to the main chain $^{\alpha}C - H$ bending, which makes the band at 1341 cm^{-1} stronger.

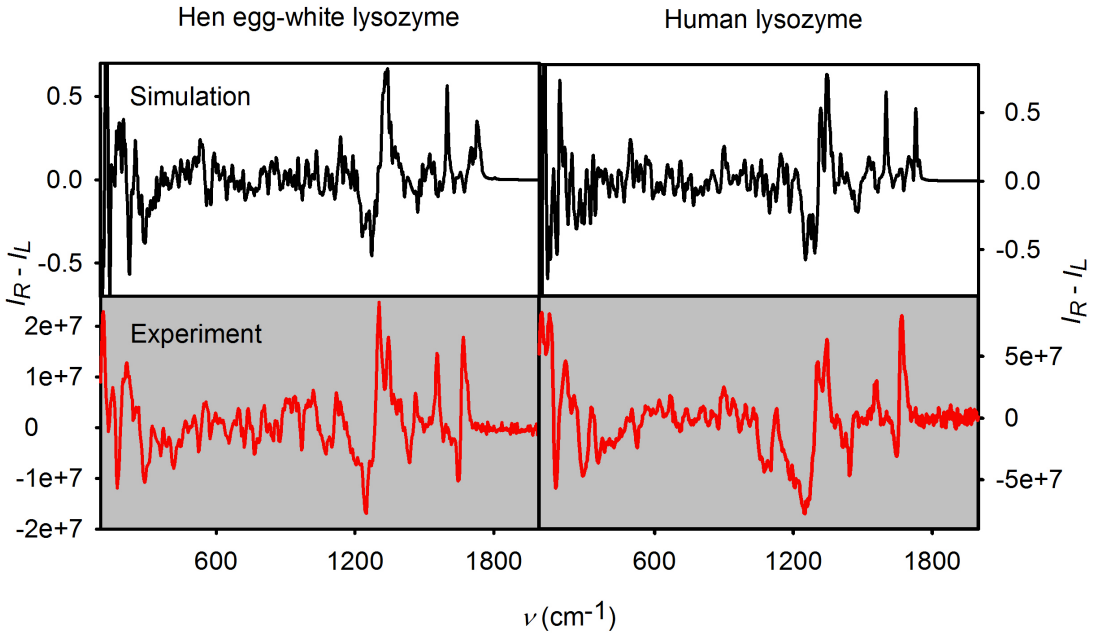


Figure 4.5: Simulated and experimental ROA spectra of hen egg-white and human lysozyme

We also simulated IR and VCD spectra (publication is being prepared). The VCD spectra are less complicated than for ROA. In VCD, we can observe dominant signals of the

amide I and amide II vibrations. As an example of such simulation, IR and VCD spectra of equine myoglobine are shown (fig. 4.6).

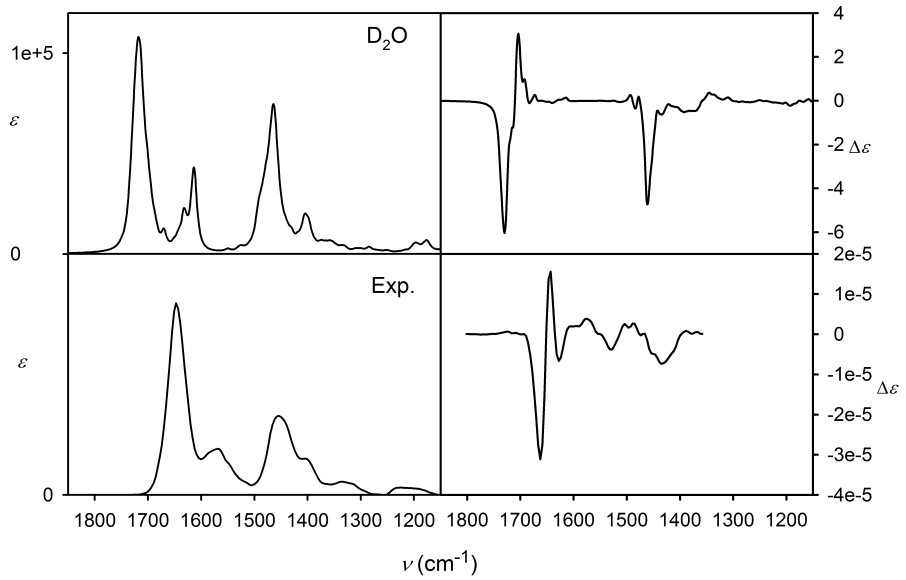


Figure 4.6: The IR and VCD spectra for equine myoglobin measured and calculated for the deuterated water solvent

In the amide I region (both in simulation and experiment), we can observe a negative/positive couplet (exp. 1662/1642 cm^{-1}), characteristic for α-helical structure. The second characteristic band is in the amide II region (exp. negative band at 1434 cm^{-1}).

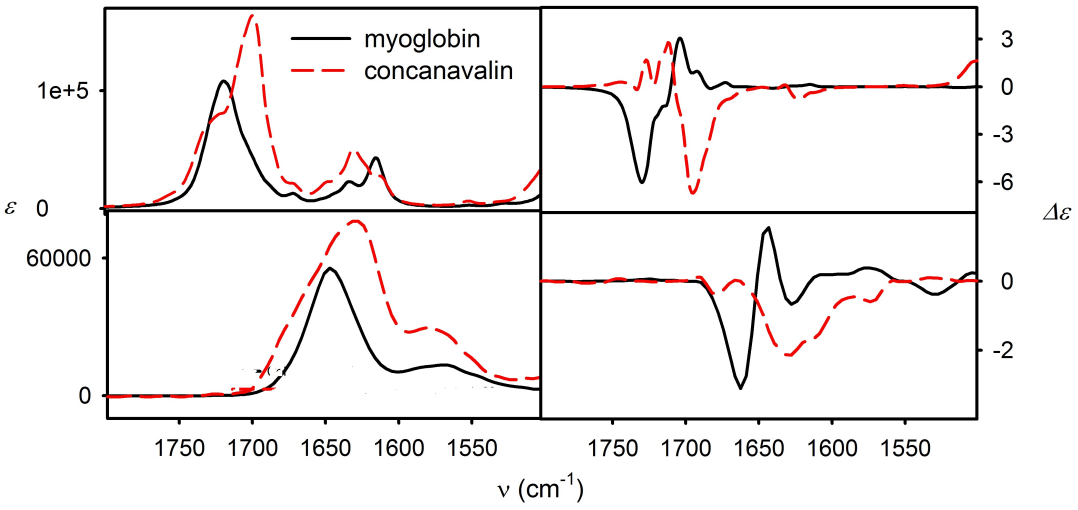


Figure 4.7: The difference between IR and VCD spectra of equine myoglobin and jack-bean concanavalin A

The differences in experimental IR and VCD spectra of the α-helical equine myoglobin and mainly β-sheet containing concanavalin A are satisfactory reproduced in simulations

(fig. 4.7). The differences in their IR spectra are small, only a slight shift (1647 to 1629 cm^{-1}) in amide I region is observed, whereas for the VCD spectra the differences are more markant. For myoglobin, we can observe, for α-helical proteins characteristic, negative/positive couplet (exp. 1662/1642 cm^{-1}), whereas for the concanavalin A the spectral pattern is opposite.

4.2.2 Fibrillar Systems

Poly-L-Glutamic acid

The results for poly-L-glutamic acid fibril IR and VCD spectra are summarized in appendix D. The most promising system seems to be the most complicated system D (fig. 3.5). The simulated IR and VCD spectra are plotted in fig. 4.8 and compared with the experiment.

We achieved satisfactory agreement between the observed and simulated IR pattern. The absorption peak calculated at 1735 cm^{-1} and observed at 1729 cm^{-1} corresponds to the C=O stretching vibration of the COOD group. In the experimental spectrum, this peak is slightly split. The splitting is not so apparent in the averaged simulated spectrum with the 10 cm^{-1} band width, only a shoulder is seen. A detailed analysis of the simulations suggests that two types of modes contribute to this band, correlated to the cis and trans orientations of the OD group with respect to the carbonyl group. The cis orientation provides slightly lower frequencies for this vibration (closer to the amide I region).

The absorption maxima calculated at 1681 cm^{-1}, 1636 cm^{-1} (with a weak feature at 1670 cm^{-1}) and 1610 cm^{-1} are primarily composed of C=O stretching of the amide groups (amide I bands) and may be associated with the experimental maxima at 1642 and 1600 cm^{-1}. The computed amide I bands are all higher in frequency than the corresponding experimental values and have a greater dispersion. The -COOD vibrations show closer agreement with experiment. The predicted CO stretching absorption maximum intensity is larger for the -COOD than for the amide, whereas the opposite relative dipole strength is observed experimentally, which is probably due to the dispersion distributing intensity differently in the experiment (dominant amide I band, split -COOD) and simulation (split amide I, both intense and single -COOD), but also may be influenced by the limited size of fibril, as the experimental fibrils are very long

The most intense experimental amide I peak in the IR is clearly correlated to a -/+ VCD couplet, and each of the intense computed amide I features is associated with a couplet contribution of the same sign. As in the IR absorption, the computed VCD positive signal at 1609 cm^{-1} does not have an experimental counterpart, possibly due to an imperfect averaging or edge effects in the simulation. Despite the split amide I band, the overall couplet pattern in simulated VCD are qualitatively in agreement with the experimental results.

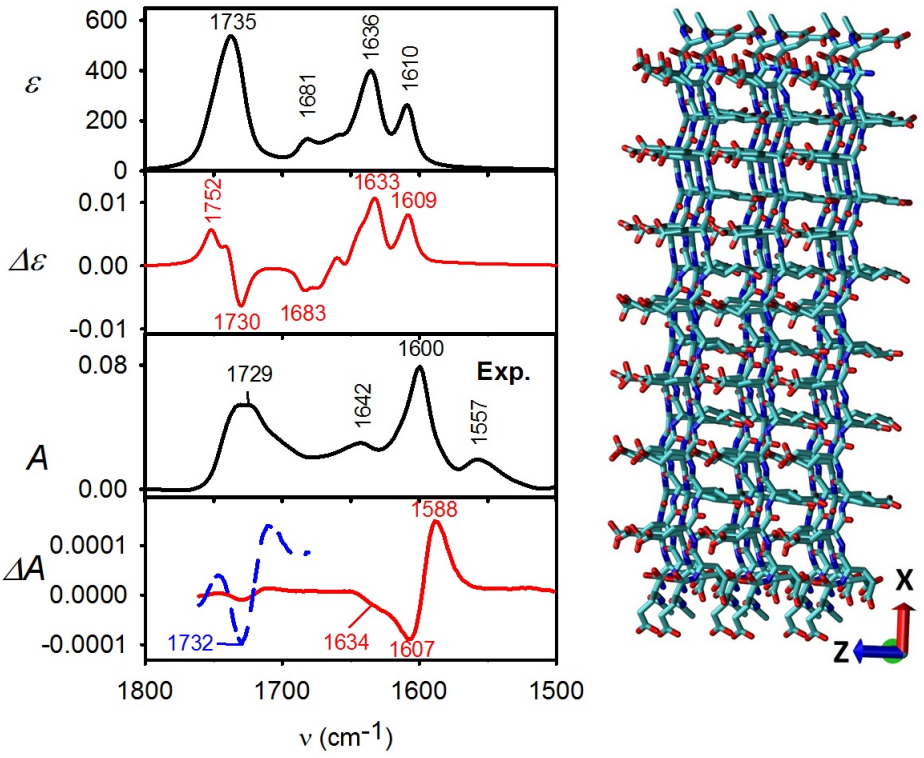


Figure 4.8: Simulated and experimental IR and VCD spectra of PLGA fibrils

Insulin fibrils

Also for insulin fibril part, publication does not exist yet. The simulated and experimental spectra of the native and fibrillar insulin are compared in fig. 4.9. In the experimental Raman spectra, formation of the fibrils is associated with minor spectral changes, such as the shift of the amide I band from 1659 to 1674 cm^{-1}, and small intensity variations in the other regions. The experimental ROA spectrum in the case of the fibril form has a more compact $-/+$ amide I couplet at 1660/1674 cm^{-1} instead of a broader one at 1640/1668 cm^{-1} for the native insulin. The native insulin $-/+$ 1250/1313 cm^{-1} band intensities become smaller and a positive 1271 cm^{-1} signal appears in this region for fibril. The ROA signal around 1313 cm^{-1} is an important band for α-helix, which forms approximately 43 % of native insulin, and the 1271 cm^{-1} band for β-turn. Negative bands at 282, 993 and 1447 cm^{-1} in ROA spectrum of native insulin disappear for the fibril.

The simulation reproduces many of experimental observations. The region 1580-1800cm^{-1}, mostly containing the C=O stretching frequencies of amide I and carboxyl is plotted in a different frequency-scale in order to provide more detailed view. The C=O frequencies are calculated too high, which is usual for the modelling at this level of theory [118, 119]. In the Raman spectra, the observed shift in amide I frequency (1659 $\rightarrow$ 1674 cm^{-1} for native to fibril transition) is reproduced satisfactory, as simulated 1730 $\rightarrow$ 1742 cm^{-1}. In ROA,

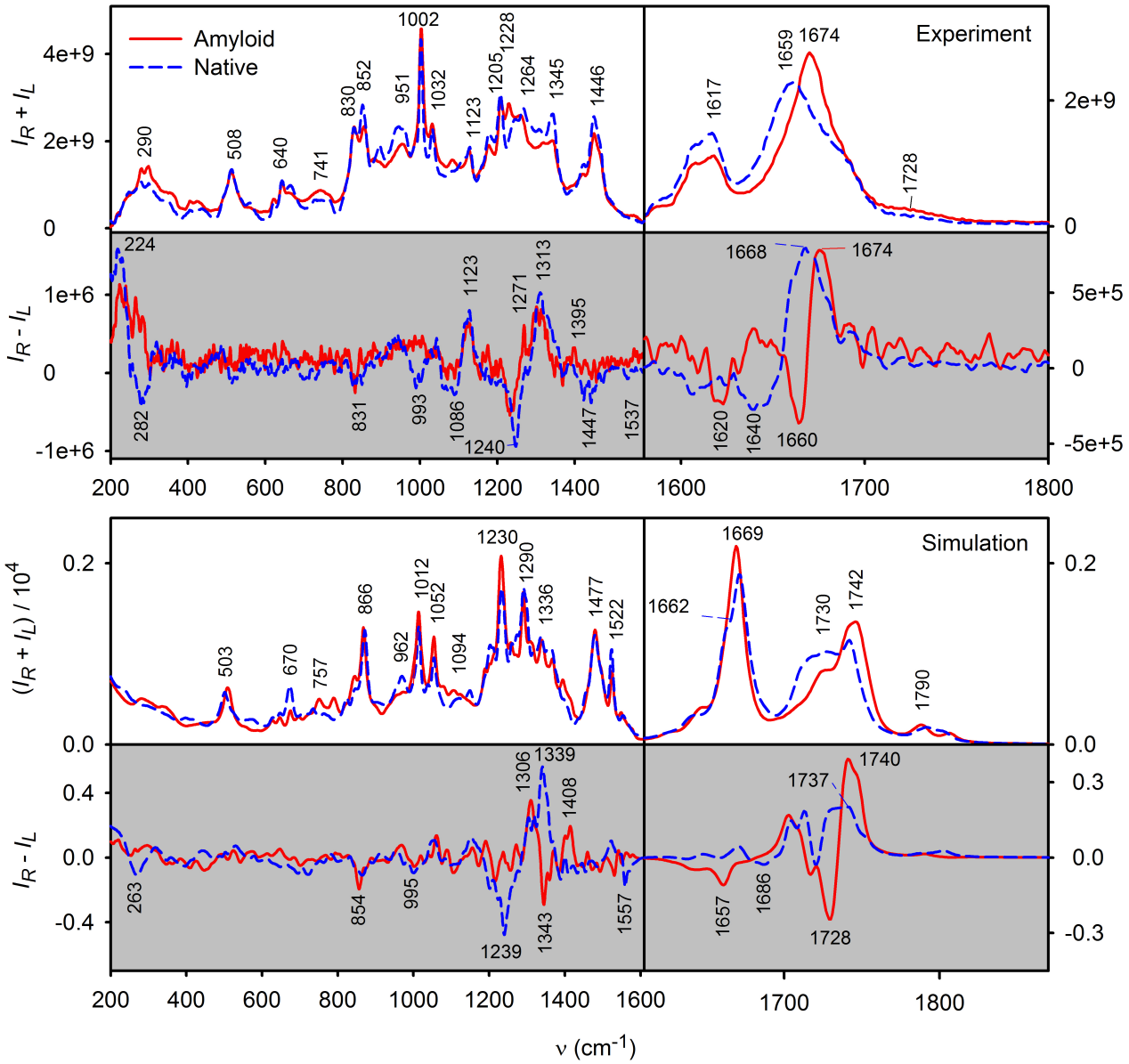


Figure 4.9: Experimental (top) and simulated (bottom) spectra of native and fibrous insulin

the fibril form has a sharp 1660(-)/1674(+) couplet in the experiment, which is reproduced by the simulation at 1728/1740 cm^{-1}. This seems to be characteristic for β-sheet structures observed also in globular proteins with a high β-sheet content [120]. In the native form, the amide I ROA signal is broader. The experimental Raman 1617 cm^{-1} band (calculated at 1669 cm^{-1}, ascribed to C=C stretching vibrations in aromatic tyrosine residues and is accompanied by a close band (experimentally 1608 cm^{-1}) of phenylalanine vibrations. These vibrations exhibit a strong negative ROA signal at 1620 cm^{-1} for the fibril, reproduced by the simulation at 1657 cm^{-1}.

The amide II band (C-N stretching and NH bending, around 1540 cm^{-1}) is weak, which is usual in non-resonance Raman peptide spectroscopy. In the native insulin, there is a weak negative ROA signal (1537 cm^{-1}, simulation 1557 cm^{-1}), disappearing for the fibril.

The histidine C-H bending vibrations (calculated 1522 cm^{-1}, experimentally 1446 cm^{-1}) give strong Raman bands but generates very weak ROA signal. In the extended amide III region (1200-1400 cm^{-1}), the simulation confirms that ROA is relatively strong and changes with the conformation. The 1313/1339 cm^{-1} (exp./calc.) ROA positive band of native insulin looses intensity and a new 1271/1306 cm^{-1} positive band appears for the fibril. The simulated spectrum changes even more, and predicts at 1343 cm^{-1} negative ROA band for the fibril, not observed in experiment. Vibrational modes in this region belongs to the main peptide chain coupled with ${}^{\alpha}C-H$ bending and side chain vibrations, which gives them sensitivity to geometry changes.

The negative ROA band of native insulin (exp. 1240 cm^{-1}) slightly shifts to 1233 cm^{-1} for the fibril. The same trend is observed in the simulations (1239 $\rightarrow$ 1213 cm^{-1}), but also here the observed intensity changes less than in theory. This can be explained by an incomplete conversion of the native state into the fibrillar one. The negative 993 cm^{-1} experimental ROA band of native insulin is not well predicted theoretically. A remarkable is the intense ROA signal in the lowest-wavenumber region (200-300 cm^{-1}) comparable with the strongest bands of the extended amide III region. The +/- native insulin bands at 224/282 cm^{-1} change into a positive signal for the fibril, which are qualitatively reproduced by the theory. Visualization of vibrational normal modes showed that the negative (282 cm^{-1}) ROA bands comes from α-helical segments of insulin. Similar +/- pattern was observed at 229/302 cm^{-1} for highly-helical human serum albumin [120].

Insulin aromatic residues comprise histidine, 3 phenylalanine and 4 tyrosine groups. The aromatic rings are not chiral, but a large ROA signal can be induced by coupling of their vibrations with neighbouring covalent bonds and other peptide parts [120]. Therefore they can sample locally the protein conformation. According to simulation, phenylalanine and tyrosine conformer ratios (the χ_1 and χ_2 side chain torsion angles) significantly change when insulin adopts the fibril form (table 4.1).

Table 4.1: Populations and χ_1, χ_2 torsion angles of the Phe and Tyr side chain conformers in native and fibrillar insulin obtained by molecular dynamics.

Conformer	$\chi_1/^\circ$	$\chi_2/^\circ$	p, native/ %	p, fibril/ %
A	-63	102	49	24
B	-173	72	31	68
C	57	92	20	8

To understand how the Phe and Tyr residues contribute to insulin ROA spectra, we simulated Raman and ROA spectra of $NH_2-Phe-COH$ and $NH_2-Tyr-COH$ model

molecules. The simulations exhibit significant changes in spectral shapes of aromatic bands at 1005, 1330 and 1650 cm^{-1} due to the conformation of the aromatic side chain. For $NH_2 - Phe - COH$, the conformers were averaged using MD populations. Raman and ROA spectra were generated in the native and fibrillar insulin and plotted in fig. 4.10. For the fibril, we can see that the C=C stretching band (number III in fig. 4.10) generates a relatively strong negative ROA signal corresponding to the observed band at 1620 cm^{-1} (fig. 4.9). A similar intensity change occurs for the aromatic hydrogen bending (1350cm^{-1}, number II) and out of plane motion (1000 cm^{-1}, number I), where correspondence to the experiment is not so obvious due to the overlap with other vibrations.

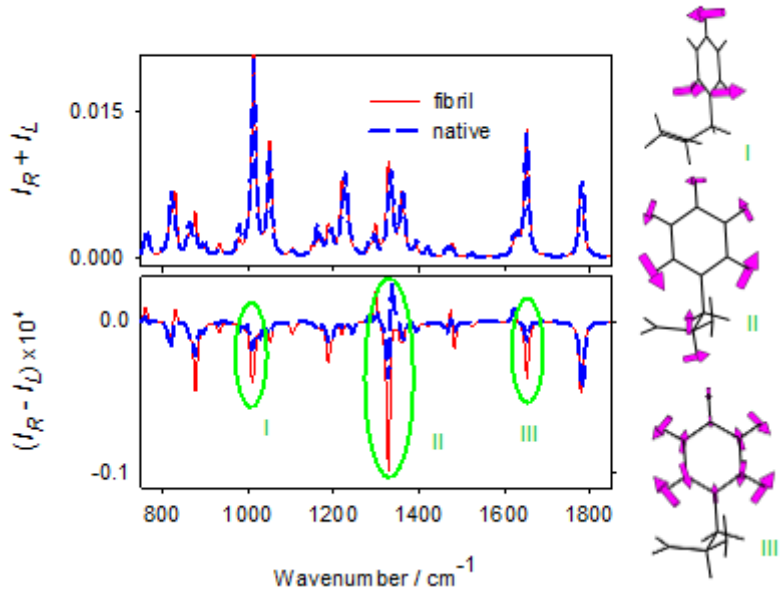


Figure 4.10: Simulated Raman and ROA spectra of a model $NH_2 - Phe - COH$ molecule mimicking Phe and Tyr average conformation in fibril and native insulin. Characteristic bands of the aromatic ring vibrations (I-III) are marked.

The ROA techniques reflect the twist of fibril threads because the twist causes local changes in insulin conformation. The spectra (fig. 4.11) are simulated for three values of the twist 0°, −3° and −6°. The twist variation approximately corresponds to that found in the model X-ray protein structures (fig. 3.7).

As shown in figure 4.11, the Raman spectra are practically insensitive to the twist, whereas ROA shapes are quite sensitive. The ROA intensity variations are less pronounced above 1500 cm^{-1}, e.g., the amide "W" shape is not much eroded, but the extended amide III region undergoes rather dramatic changes. All intensity variations are not monotonic. For the twist of −3° a positive ROA signal appears around 1275 cm^{-1}, while the intensity is smaller for 0° and −6°. In the lower-frequency region (below 1000 cm^{-1}), the spectra are less dependent on the twist. At the current level of experimental noise and computational

precision we can not determine the twist of insulin fibril from comparison of the theoretical and experimental spectra. However, the computational experiment well documents the potential of the ROA spectroscopy for future fibril studies.

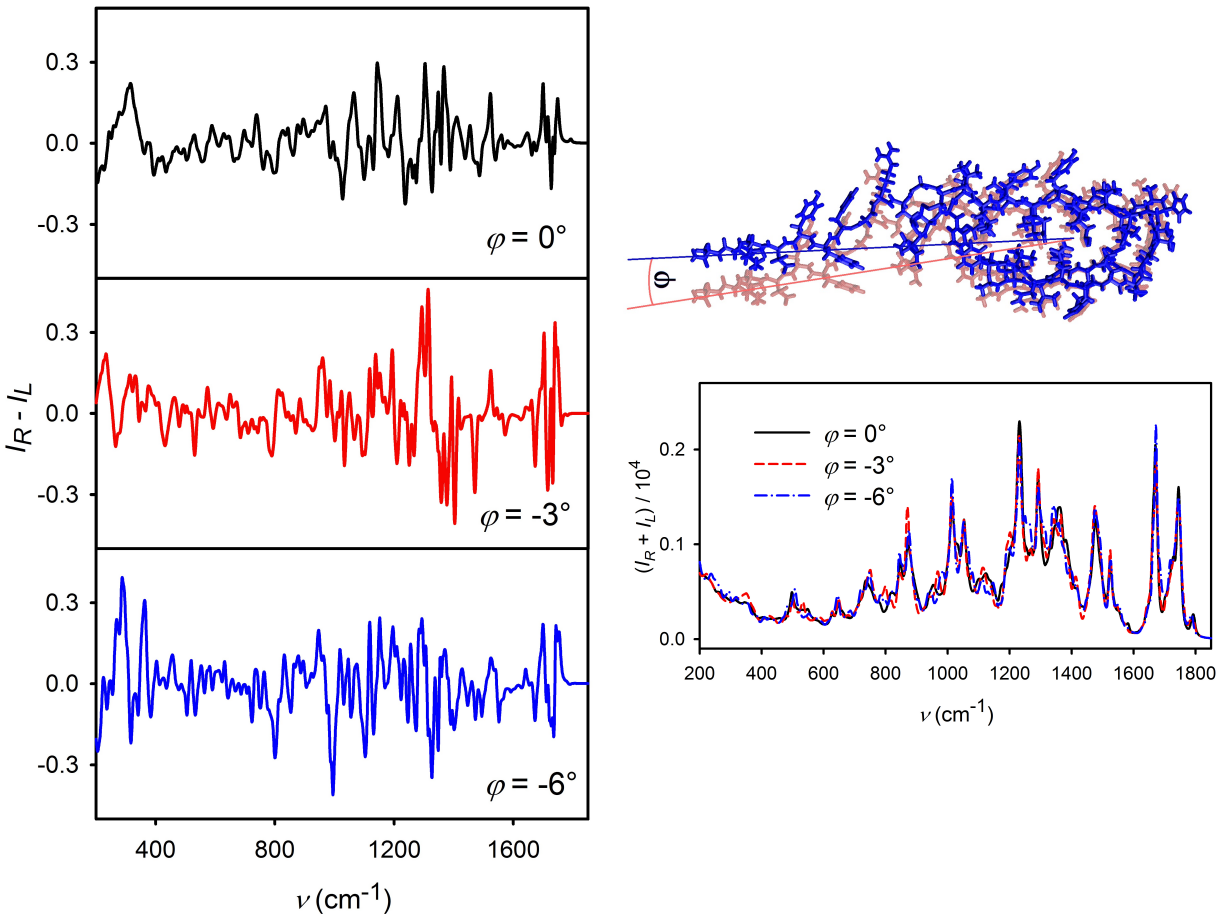


Figure 4.11: ROA and Raman spectra of insulin fibril simulated for three values of the twist between neighbouring insulin units (100 MD snapshots avaraged).

Because the spectra for different fibril twist were generated from a limited number of MD snapshots, we needed to be sure that the error of the averaging is smaller than the effect of the twist shown in Figure 4.11. In the figure 4.12, Raman and ROA spectra for the twist of $-6°$ are plotted for 50 and 100 MD snapshots. Small differences in Raman and ROA intensities for the 50 and 100 snapshot simulations indicate that the twist-induced changes in the spectra are reliable. They are mainly caused by structural changes, such as variation of the distribution of torsional angles. Examples of such distributions are given in figure 4.13, for the main chain φ and ψ torsion angles, and histidine and tyrosine χ_1 side chain angles. Many (φ, ψ) values are near $(-145°, 150°)$, corresponding to the standard β-sheet conformation, and that some twist-induced changes are not monotonic in the row $0° \rightarrow -3° \rightarrow -6°$, similarly as for the ROA spectra. It is also obvious that the individual side chain angles χ_1 are more sensitive to the twist than the average φ and ψ distributions.

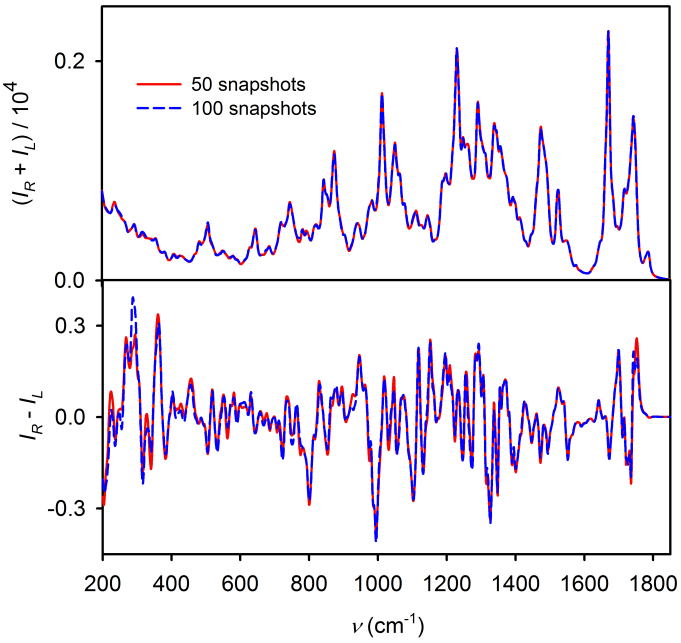


Figure 4.12: Convergence for twisted fibril simulations: Raman and ROA simulated for 50 and 100 snapshot averages

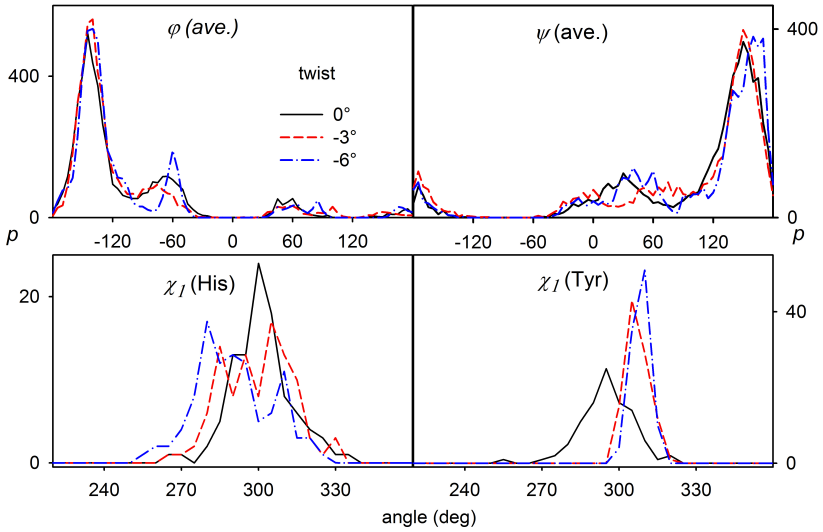


Figure 4.13: Distributions of torsional angles for 3 twisted fibril structures

4.3 Methodical Projects

4.3.1 Helical Periodic Boundary Conditions

These results discussed in publication in appendix F. For example on an insulin fibril model, we examined if its geometry was compatible with the β-roll protein. During the simulation, the insulin molecules were satisfactory stable and the β-roll structure was maintained.

The simulation time (300 ps) was relatively short but the simulations provide converged potential energy profiles. For the zero twist, the average geometry was very close to that obtained by the longer (8 ns) periodic boundary condition (PBC) computation (Fig. 4.14). The largest deviations are outside the insulin loop for the flexible part of the peptide chain. For PBC, the Amber program was used. The insulin molecule was stabilized by sandwiching it between two other molecules.

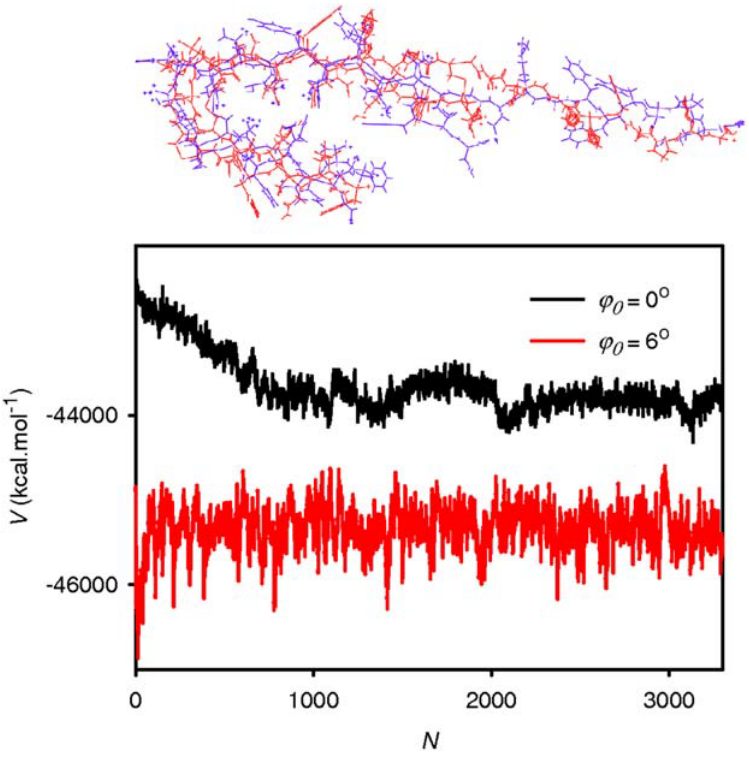


Figure 4.14: On the top: overlap of the average structures of the insulin fibril model obtained by HPBC (Tinker,$\varphi_0 = 0°$, 0.3 ns) and PBC (Amber10, 8 ns) dynamics, and on the bottom: the potential energy from the HPBC simulations for the two twist angles, N is number of the snapshot taken in 100 fs intervals.

For all cases, the conformation required for the fibril formation, was supported by the Amber99 force field. Also, for $\varphi = 0°$, the structure was more flexible than for $\varphi = 6°$, which indicates that the twisted fibrillar geometry is preferred by the Amber99 force field,

in agreement with the experimental observations of the β-roll protein in other fibril studies [121, 122, 123, 124].

4.3.2 Transfer of Frequency-Dependent Polarizabilities

All results are summarized in appendix G. We comment on the behaviour of the method for the case of N-methylacetamide (NMA) dimer. If we look on fig. 4.15, we can see an absorption and ECD spectra for 3 NMA dimers types with separation distances of 4.5, 6.5 and 8.5 Å. The TFDP and TDDFT computation are compared. For 4.5 Å where excited electronic states involving charge-transfer between the NMA molecules significantly contribute to spectral intensities, ECD spectra obtained by the TFDP and TDDFT methods are very different. Only some spectral features are reproduced, such as the mostly negative signal within 180-200 nm, a positive one within 145-180 nm, and a negative lobe at 135 nm. Much smaller differences can be found between the TFDP and TDDFT absorption spectra. The small differences correspond to the different mechanisms providing the absorption and ECD signals. While the monomer ECD is zero, the absorption dimer signal is very well approximated by the sum of the monomer spectra. It means that ECD is more sensitive to proper modelling of the NMA-NMA interaction in the dimer.

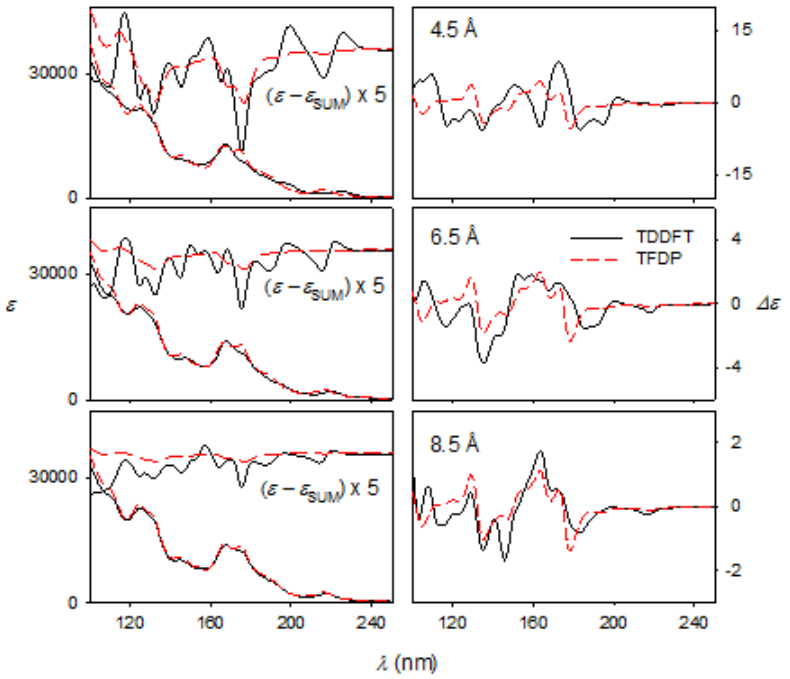


Figure 4.15: Absorption (left) and ECD (right) spectra of NMA dimer, with monomers separated by 4.5, 6.5 and 8.5 Å. TDDFT and TFDP methods are compared.

For 6.5 Å, the absorption spectra obtained by the two tested methods are nearly identical. The differences between the TDDFT and TFDP spectra are smaller than for 4.5 Å, both for

absorption and CD. The TFDP curve reasonably well mimics several TDDFT CD features, such as the positive/negative signal around 130 nm, positive signal within 150-170 nm, and a negative peak at 217 nm. Around 180 nm, both approaches provide a negative lobe, but exact positions of the minimum differs (179 nm for TFDP vs. 186 nm for TDDFT).

For the distance of 8.5 Å a direct interaction of the NMA electronic clouds is almost excluded, and the influence of intermolecular charge-transfer transitions is limited. The TFDP and TDDFT spectra are even more similar, although minor differences still occur. For example, the negative TFDP CD signal at 146 nm is predicted by TDDFT as a shoulder only. However, the TFDP seems to be suitable to simulate spectra for distant chromophores.

5 Conclusion

This Thesis was focused on Optical Activity of biologically-relevant systems. The largest molecules included poly-L-glutamic acid fibrils, a set of globular proteins and insulin fibrils.

For poly-L-glutamic acid (PLGA), we used a crystal-like geometry model. With several annealing cycles and molecular dynamic equilibration we verified that the structure was stable. Using the Cartesian coordinate-based transfer of the force field and intensity tensors calculated for smaller fragments, IR and VCD spectra of sheet segments were simulated. They exhibited a qualitative agreement with experimental data. The averaged spectral patterns converged relatively quickly with the number of MD snapshots and annealing cycles. Analysis of the side chain VCD indicated that the spectrum reflects the chiral -COOH group orientation. We can thus conclude that VCD spectroscopy combined with the simulations is able to recognize specific geometry features of fibrillar structures.

For the ROA of globular proteins, the computational techniques provided more satisfactory precision in frequencies and spectral intensities. They could be used to interpret experimental Raman and ROA spectra more thoroughly. We were able to reproduce the differences between mainly α-helical human serum albumin and concanavalin A containing mainly β-sheets, or the differences between very similar hen egg-white and human lysozymes. The ROA spectroscopic method is also a useful tool to monitor protein amino acid composition. The possibility to monitor biomolecules, in this case globular proteins, in their natural aqueous environment makes the ROA suitable for applications in biology and medical imaging.

We used a combination of molecular dynamics and density functional theory to simulate Raman and ROA spectra of the complicated structure of the insulin fibril. Based on the modelling we could understand most of the spectral changes observed experimentally, i.e. the dependence of ROA peak signs and intensities on the structure. Insulin geometries derived from the β-roll and β-helix proteins provided similar spectra. The most spectral changes could be interpreted as a α-helix $\rightarrow$ β-sheet secondary structure transition. These changes were observed in the amide I, extended amide III and lowest-frequency (bellow 300 cm^{-1}) spectral regions.

Computational modelling with simplified systems also revealed that the aromatic residues in insulin had not only strong Raman, but also ROA signal that can be potentially used as a local probe of the structure. A surprising fact was the sensitive response of the amide III ROA signal to the changes of the conformation simulated for different twists. We showed that this does not reflect a direct non-covalent interaction of peptide chains, but is caused by changes in local protein conformation. The results showed that the ROA technique is sensitive to protein conformational changes during the fibrillation, and that the complex sim-

ulations are needed to extract the information about the structure and molecular behaviour from the spectra.

Apart from the proteins, we studied also smaller molecules. We confirmed that the Raman spectroscopy was suitable to distinguish polymorphic structures of various drugs. The UV and ECD spectra of monensin metal complexes were calculated; they were quite sensitive to metal ion, which can be used in analytical chemistry. Interesting results were also obtained for lanthanide complexes. The Circularly Polarized Luminiscence (CPL) was induced by interaction with aminoacids. The experimental data were correlated with complex formation energies acquired using WHAM method. The induced CPL can thus be used to monitor the biomolecules indirectly in the future.

Methodical projects included the implementation and testing of helical periodic boundary conditions into the Tinker program. The implementation appeared to be stable and usable for systems possessing helical symmetry, such as the fibrils.

The second methodical project dealt with transfer of molecular properties needed for UV-vis and Electronic Circular Dichrosim (ECD) spectroscopies. Good results were achieved by using the frequency-dependent polarizabilities (TFDP). This method provided results at least comparable with an older similar semiempirical model (TDC). The transfer parameters could additionally be varied to account for the chromophore environment and polarizability localization. The favourable computational time allows to study very large systems.

We believe that improved molecular dynamics methodology can provide even more detailed information about the world of biomolecules in the future.

6 List of Abbreviations

CD Circular Dichroism.

CPL Circularly Polarized Luminiscence.

DFT Density Functional Theory.

MD Molecular Dynamic.

NMA N-methylacetamide.

NMR Nuclear Magnetic Resonance.

OR Optical Rotation.

PBC Periodic Boundary Conditions.

PMF Potential of Mean Force.

RFO Rational Function Optimization.

ROA Raman Optical Activity.

SOS Sum Over States.

TDC Transition Dipole Coupling.

TDDFT Time Dependent Density Functional Theory.

TFDP Transfer of Frequency-Dependent Polarizabilities.

UV-vis Ultraviolet and visible wavelength region.

VCD Vibrational Circular Dichroism.

VOA Vibrational Optical Activity.

WHAM Weighted Histogram Analysis Method.

7 References

[1] Zhu, F.; Kapitán, J.; Tranter, G.; Pudney, P. D. A.; Isaacs, N. W.; Hecht, L.; Barron, L. D. *Proteins* **2008,** *70,* 823.

[2] Keiderling, T. A. *Curr. Opin. Chem. Biol.* **2002,** *6,* 682.

[3] Bouř, P.; Sopková, J.; Bednarová, L.; Maloň, P.; Keiderling, T. A. *J. Comput. Chem.* **1997,** *18,* 646.

[4] Yamamoto, S.; Li, X.; Ruud, K.; Bouř, P. *J. Chem. Theory Comput.* **2012,** *8,* 977.

[5] Arago, D. F. *Mem. de L' Inst.* **1811,** *12,* 93.

[6] Fresnel, A. *Bull. Soc. Philomath.* **1824,** 147.

[7] Pasteur, L. *Ann. Chim.* **1848,** *24,* 457.

[8] Bel, J. L. *Bull. Soc. Chim.* **1874,** *22,* 337.

[9] van't Hoff, J. H. *Chemische Constitutie van Organische Verbindungen;* Utrecht, 1874.

[10] Kelvin, L. *The Baltimore Lecture;* C.J. Clay and Sons: London, 1904.

[11] Cahn, R. *J. Chem. Educ.* **1964,** *41,* 116.

[12] Cahn, R.; Ingold, C.; Prelog, V. *Angewandte Chemie, Intl. Ed.* **1966,** *5,* 385.

[13] Greenfield, N. J. *Nature protocols* **2006,** *1,* 2876.

[14] Muller, G. *Dalton Trans.* **2009,** 9692.

[15] Lunkley, J. L.; Shirotani, D.; Yamanari, K.; Kaizaki, S.; Muller, G. *J. Am. Chem. Soc.* **2008,** *130,* 13814.

[16] Dekkers, H. P. J. M. Circular dichroism: principles and applications.. In ; Berova, N.; Nakanishi, K.; Woody, R. W., Eds.; Wiley- VCH: New York, 2000; Chapter Circularly polarized luminesence: a probe for chirality in the excited state, pages 185–215.

[17] Kapitán, J. "Využití Ramanovy optické aktivity při studiu konformace biomolekul ve vodném prostředí", Master's thesis, MFF UK, 2001.

[18] Burmann, B. M.; Hiller, S. *Prog. Nucl. Magn. Reson. Spectrosc.* **2015,** *86-87,* 41.

[19] Barron, L. D.; Hecht, L.; Blanch, E. W.; Bell, A. F. *Prog. Biophys. Mol. Biol.* **2000,** *73,* 1.

[20] Krimm, S.; Mark, J. E. *Proc. Natl. Acad. Sci. U. S. A.* **1968,** *60,* 1122.

[21] Dukor, R. K.; Keiderling, T. A. *Biopolymers* **1991,** *31,* 1747.

[22] Tiffany, M. L.; Krimm, S. *Biopolymers* **1972,** *11,* 2309.

[23] Ma, S.; Freedman, T. B.; Dukor, R. K.; Nafie, L. A. *Appl. Spectrosc.* **2010,** *64,* 615.

[24] Baumruk, V.; Pancoška, P.; Keiderling, T. A. *J. Mol. Biol.* **1996,** *259,* 774.

[25] Shanmugam, G.; Polavarapu, P. L. *J. Am. Chem. Soc.* **2004,** *126,* 10292.

[26] Ma, S.; Cao, X.; Mak, M.; Sadik, A.; Walkner, C.; Freedman, T.; Lednev, I. K.; Dukor, R. K.; Nafie, L. A. *J. Am. Chem. Soc.* **2007,** *129,* 12364.

[27] Kurouski, D.; Lombardi, R.; Dukor, R.; Lednev, I. K.; Nafie, L. A. *Chem. Commun. (Camb).* **2010,** *46,* 7154.

[28] Kurouski, D.; Lu, X.; Popova, L.; Wan, W.; Shanmugasundaram, M.; Stubbs, G.; Dukor, R. K.; Lednev, I. K.; Nafie, L. A. *J. Am. Chem. Soc.* **2014,** *136,* 2302.

[29] Barron, L. D.; Buckingham, A. D. *Chem. Phys. Lett.* **2010,** *492,* 199.

[30] Zhu, F. J.; Issacs, N. W.; Hecht, L.; Barron, L. D. *Structure* **2005,** 1409.

[31] Barron, L. D.; Zhu, F.; Hecht, L.; Tranter, G. E.; Isaacs, N. W. *J. Mol. Struct.* **2007,** *834-836,* 7.

[32] Car, R.; Parrinello, M. *Phys. Rev.* **1985,** *55,* 2471.

[33] Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959,** *31,* 459-466.

[34] Rahman, A. *Phys. Rev.* **1964,** *136,* A405–A411.

[35] Verlet, L. *Phys.Rev.* **1967,** *159,* 98–103.

[36] Hockney, R. W.; Goel, S. P.; Eastwood, J. *J. Comp. Phys.* **1974,** *14,* 148–158.

[37] Morse, P. M. . *Phys. Rev.* **1929,** *34,* 57–64.

[38] Lennard-Jones, J. E. *Proc. R. Soc. Lond. A* **1924,** *106,* 463–477.

[39] Buckingham, R. A. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **1938,** *168,* 264–283.

[40] Ponder, J. W.; Case, D. A. *Adv. Prot. Chem.* **2003,** *66,* 27–85.

[41] Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. *Biochim. Biophys. Acta* **2006,** *1764,* 1647-1676.

[42] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987,** *91,* 6269–6271.

[43] Adams, D.; Adams, E. ands Hills, G. *Mol. Phys.* **1979,** *38,* 387–400.

[44] Kirkwood, J. G. *J. Chem. Phys.* **1935,** *3,* 300-313.

[45] Torrie, G.; Valleau, J. *Chem. Phys. Lett.* **1974,** *28,* 578-581.

[46] Roux, B. *Comput. Phys. Commun.* **1995,** *91,* 275–282.

[47] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992,** *13,* 1011–1021.

[48] Woolf, T. B.; Roux, B. *J. Am. Chem. Soc.* **1994,** *116,* 5916–5926.

[49] Broyden, C. G. *J. Inst. Math. Appl.* **1970,** *6,* 76.

[50] Fletcher, R. *Comput. J.(UK)* **1970,** *13,* 317.

[51] Goldfarb, D. *Math. Comput.* **1970,** *24,* 23.

[52] Shanno, D. F. *Math. Comput.* **1970,** *24,* 647.

[53] Banerjee, A.; Adams, N.; Simons, J.; Shephard, R. *J. Phys. Chem.* **1985,** *89,* 52.

[54] Simons, J.; Nichols, J. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1990,** *24,* 263.

[55] Schlegel, H. B. *Modern Electronic Structure Theory;* World Scientific: Singapore, 1995.

[56] J. Baker, W. J. H. *J. Comput. Chem.* **1991,** *12,*.

[57] Baker, J. *J. Comput. Chem.* **1993,** *14,* 1085.

[58] Barron, L. D. *Molecular Light Scattering and Optical Activity;* Cambridge University Press: Cambridge, 2004.

[59] Stephens, P. J. *J. Phys. Chem.* **1987,** *91,* 1712.

[60] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1694,** *136,* B864.

[61] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965,** A1133.

[62] Allen, F. H. *Acta Crystallogr., Sect. B* **2002,** *58,* 380.

[63] Bouř, P.; Keiderling, T. A. *J. Chem. Phys.* **2002,** *117,* 4126.

[64] Frisch, M. J. *et al.* "Gaussian 09, Revision B.01", 2009.

[65] Becke, A. D. *Phys. Rev. A* **1988,** *38,* 3098–3100.

[66] Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996,** *54,* 16533.

[67] Haney, M. E.; Hoeln, M. M. *Antimicrob Agents Chemother (Bethesda)* **1967,** *7,* 349.

[68] Agtarap, A.; Chamberlin, J. W. *Antimicrob Agents Chemother (Bethesda)* **1967,** *7,* 359.

[69] Agtarap, A.; Chamberlin, J. W.; Pinkerton, M.; Steinrauf, L. K. *J. Am. Chem. Soc.* **1967,** *89,* 5737.

[70] Lutz, W. K.; Winkler, F. K.; Dunitz, J. *Helv Chim Acta* **1971,** *54,* 1103.

[71] Huczynski, A.; Katrusiak, M. R.-S. A.; Brzezinski, B. *J. Mol. Struct.* **2007,** *871,* 92.

[72] Huczynski, A.; Katrusiak, M. R.-S. A.; Brzezinski, B. *J. Mol. Struct.* **2007,** *832,* 84.

[73] Pangborn, W.; Duax, W. L.; Langs, D. *Journal of Americal Chemical Society* **1987,** *109,* 2163.

[74] Yildirim, S. O.; McKee, V.; Khardli, F. Z.; Mimouni, M.; Hadda, T. B. *Acta Cryst.* **2008,** *64,* 154.

[75] Pinkerton, M.; Steinrauf, L. K. *J. Mol. Biol.* **1970,** *49,* 533.

[76] Becke, A. D. *J. Chem. Phys.* **1993,** *98,* 1372.

[77] A. Klamt, A.; Jonas, V.; Burger, T.; Lohrentz, J. C. W. *J. Phys. Chem. A* **1998,** *102,* 5074.

[78] Mennucci, B.; Cappelli, C.; Cammi, R.; Tomasi, J. *Chirality* **2011,** *23,* 717.

[79] Dolg, M. *Relativistic effective core potentials. In: Schwerdtfeger P, editor. Relativistic electronic structure theory, Part I: Fundamentals.;* Elsevier: Amsterdam, 2002.

[80] Furche, F.; Ahlrichs, R.; Weber, C. W. E.; Voegtle, A. S. F.; Grimme, S. *J. Am. Chem. Soc.* **2000,** *122,* 1717.

[81] Case, D. A. *et al.* "AMBER 14", 2014.

[82] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004,** *25,* 1157–1174.

[83] Lindorff-Larsen, K.; Piana, S.; Palmo, P.; Maragakis, P.; Kleipis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010,** *78,* 1950.

[84] Li, P.; Song, L. F.; Merz, K. M. *J. Phys. Chem. B* **2015,** *119,* 883.

[85] Frankaer, C. G.; Mossin, S.; Ståhl, K.; Harris, P. *Acta Crystallographica Section D* **2014,** *70,* 110–122.

[86] Pike, A.; Brew, K.; Acharya, K. R. *Structure* **1996,** *4,* 691.

[87] Mizutani, R.; Shimizu, Y.; Saiga, R.; Ueno, G.; Nakamura, Y.; Takeuchi, A.; Uesugi, K.; Suzuki, Y. *Sci. Rep.* **2014,** *4,* 5731.

[88] Artymiuk, P. J.; Blake, C. C. *J. Mol. Biol.* **1981,** *152,* 737.

[89] Goodman, J. S.; Chao, S.-H.; Pogorelov, T. V.; Gruebele, M. *The Journal of Physical Chemistry B* **2014,** *118,* 6511.

[90] Deacon, A.; Gleischmann, T.; Kalb, A. J.; Price, H.; Raftery, J.; Bradbrook, G.; Yarivc, J.; Helliwella, J. R. *J. Chem. Soc., Faraday Trans.* **1997,** *93,* 4305.

[91] Min, X.; Carter, D. C. *Nature* **1992,** *358,* 209.

[92] Case, D. A. *et al.* "AMBER 10", 2008.

[93] Devlin, F. J.; Stephens, P. J.; Cheeseman, J. R.; Frisch, M. J. *J. Phys. Chem.* **1997,** *101,* 9912.

[94] Polavarapu, P. L. *J. Phys. Chem.* **1990,** *94,* 8106.

[95] Nafie, L. *Vibrational Optical Activity: Principles and Applications;* Wiley: Chichester, 2011.

[96] Kessler, J.; Dračínský, M.; Bouř, P. *J. Comput. Chem.* **2013,** *34,* 366.

[97] Keith, H. D.; Giannoni, G.; Padden, F. J. *Biopolymers 1969* **1969,** *7,* 775.

[98] Itoh, K.; Foxman, B. M.; Fasman, G. D. *Biopolymers* **1976,** *15,* 419.

[99] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008,** *4,* 435–447.

[100] N. Kamiya, N.; Watanabe, Y. S.; Ono, S.; J. Higo 2005, 401, . *Chem. Phys. Lett.* **2005,** *401,* 312.

[101] Ponder, J. W. *Tinker 6.2;* Washington University School of Medicine:Saint Louis: 200.

[102] Kessler, J.; Bouř, P. *J. Comput. Chem.* **2014,** *35,* 1552.

[103] Kamiya, N.; Watanabe, Y. S.; Ono, S.; Higo, J. *Chem. Phys. Lett.* **2005,** *401,* 312.

[104] Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids; , 1987.;* Oxford University Press: New York: 1987.

[105] Schlick, T. *Molecular Modeling and Simulation;* Springer: Berlin: 2002.

[106] Beeman, D. *J. Comput. Phys.* **1976,** *20,* 130.

[107] Creighton, T. E. *Proteins: Structures and Molecular Properties;* W. H. Freeman and Co.: New York, 1993.

[108] Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *J. Phys. Chem.* **1992,** *96,* 135.

[109] Yanai, T.; Tew, D.; Handy, N. *Chem. Phys. Lett.* **2004,** *393,* 51.

[110] Vydrov, O. A.; Scuseria, G. E.; Perdew, J. P. *J. Chem. Phys.* **2007,** *126,* 154109.

[111] Muliken, R. S. *J. Chem. Phys.* **1955,** *23,* 2343.

[112] Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984,** *5,* 129.

[113] Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987,** *8,* 894.

[114] Hu, H.; Lu, Z.; Yang, W. *J. Chem. Theory Comput.* **2007,** 1004.

[115] Štěpánek, P.; Bouř, P. *J. Comput. Chem.* **2013,** *34,* 1531.

[116] Jordan, P.; Fromme, P.; Witt, T. H.; Klukas, O.; Saenger, W.; Kraus, N. *Nature* **2001,** *411,* 909.

[117] Kessler, J.; Bouř, P. *J. Chem. Theory Comput.* **2015,** *11,* 2210.

[118] Kubelka, J.; Keiderling, T. A. *J. Phys. Chem. A* **2001,** *105,* 10922.

[119] Bouř, P.; Michalík, D.; Kapitán, J. *J. Chem. Phys.* **2005,** *122,* 144501.

[120] Kessler, J.; Kapitán, J.; Bouř, P. *J. Phys. Chem. Lett.* **2015,** *6,* 3314.

[121] Chi, H.; Welch, W. R. W.; Kubelka, J.; Keiderling, T. A. *Biomacromolecules* **2013,** *14,* 3880.

[122] Wojcik, S.; Babenko, V.; Dzwolak, W. *Langmuir* **2010,** *26,* 18303.

[123] Dzwolak, W.; Surmacz-Chwedoruk, W.; Babenko, V. *Langmuir* **2013,** *29,* 365.

[124] Kurouski, D.; Dukor, R. K.; Lu, X.; Nafie, L. A.; Lednev, I. K. *Chem. Commun.* **2012,** *48,* 2837.

8 Declaration of Co-Authorship

Prohlášení autorů upřesňující podil Jiřího Kesslera na publikacích přiložených k disertaci:

Declaration of co-authors declaring the contribution of Jiří Kessler to publications included in the Thesis:

Publications Discussed in the Thesis

1. Dračínský, M.; Procházková, E.; Kessler, J.; Šebestík, J.; Matějka, P.; Bouř, P. *J. Phys. Chem. B* **2013**, 117, 7297. (contribution 50 %)
2. Nezhib, A.; Kessler, J.; Bouř, P.; Gyurcsik, B.; Pantcheva, I. *Chirality* **2016**, 28, 420. (contribution 80 %)
3. Wu, T.; Kessler, J.; Bouř, P.*PCCP* **2016**, 18, 23803. (contribution 30 %)
4. Kessler, J.; Keiderling, T. A.; Bouř, P. *J. Phys. Chem. B* **2013**, 118, 6937. (contribution 80 %)
5. Kessler, J.; Kapitán, J.; Bouř, P. *J. Phys. Chem. Lett.* **2015**, 6, 3314. (contribution 60 %, 100 % of theoretical calculations)
6. Kessler, J.; Bouř, P. *J. Comput. Chem.* **2014**, 35, 1552. (contribution 80 %)
7. Kessler, J.; Bouř, P. *J. Chem. Theory Comput.* **2015**, 11, 2210. (contribution 80 %)

Other Publications

8. Melcrová, A.; Kessler, J.; Bouř, P.; Kaminský, J. *PCCP* **2016**, 18, 2130. (contribution 30 %)
9. Wu, T.; Průša, J.; Kessler, J.; Dračínský, M.; Valenta, J.; Bouř, P. *Anal. Chem.* **2016**, 88, 8878. (contribution 20 %)

Mgr. Jiří Kessler carried out the work in all publications with a high degree independence. He concentrated on the theoretical aspects of the work, approximate contributions to the publications are given in the parentheses. They are relatively low in the other publications and in publication three based largely on experimental data.

Mgr. Jiří Kessler provedl práci na všech publikacích s vysokou mírou nezávislosti. Zaměřoval se na teoretické aspekty práce, přibližné podíly na publikacích jsou uvedeny v závorkách. Podíly jsou relativně nízké v ostatních publikacích a v publikaci číslo tři obsahující převážně experimentální data.

Praha, květen 2017

prof. RNDr. Petr Bouř, DSc.

9 Publications Discussed in the PhD. Thesis

9.1 Appendix A – *J. Phys. Chem. B* 2013, 117, 7297.

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Resolution of Organic Polymorphic Crystals by Raman Spectroscopy

Martin Dračínský,[†,‡] Eliška Procházková,[†] Jiří Kessler,[†] Jaroslav Šebestík,[†] Pavel Matějka,[§] and Petr Bouř*[,†]

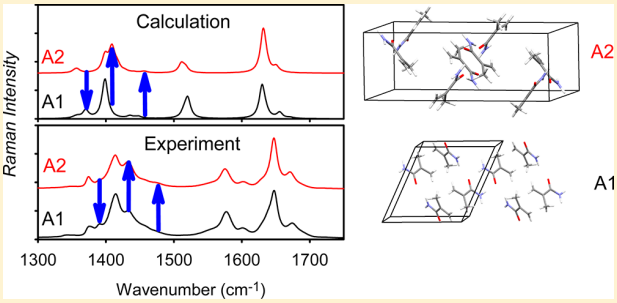
[†]Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Flemingovo nám. 2, 166 10 Prague, Czech Republic

[‡]Department of Chemistry, Durham University, South Road, Durham DH1 3LE, United Kingdom

[§]Department of Physical Chemistry, Institute of Chemical Technology, Technická 5, 166 28 Prague, Czech Republic

Supporting Information

ABSTRACT: Depending on crystallization conditions, many organic compounds can form crystals of different structure. Their proper characterization is important, for example, in the pharmaceutical industry. While the X-ray diffractometry established as a standard method, alternative techniques are desirable for broader application flexibility and economic reasons. In the present study, Raman spectroscopy combined with the density functional calculations is suggested as a complementary method to the X-ray and other higher resolution techniques. The potential to discriminate structural differences in polymorphic crystalline forms is documented on three model compounds of industrial importance. Methacrylamide, piracetam, and 2-thiobarbituric acid were crystallized under various conditions, and their Raman spectra were recorded using 532 and 1064 nm laser excitations. X-ray diffractometry and nuclear magnetic resonance spectroscopy were used as complementary techniques to verify sample composition and structure. To interpret the observed differences in Raman frequencies and intensities, three computational strategies were explored based on single molecule, a cluster model, and a plane-wave periodic boundary conditions calculation. The single-molecule modeling was found inadequate, whereas the plane-wave approach provides the most realistic spectra. For all compounds, the differences in the Raman spectra of polymorphic forms could be unambiguously assigned to the simulations. The modeling revealed that the spectral differences were caused by the molecular structure itself as well as by crystal packing. The relative importance of these factors significantly varied across the investigated samples. Owing to its simplicity, Raman spectroscopy appears to be a promising technique capable of reliable discriminating between organic crystal polymorphic states.



INTRODUCTION

Many elements, ionic compounds, or molecules can form crystals of more than one structure. This behavior, polymorphism, is also encountered in single-component organic crystals.[1] Individual crystal forms can then exhibit different physical or even chemical properties.[2] The identification of polymorphic forms is therefore of crucial importance, for example, in the pharmaceutical industry. Various pharmaceutical processes yield different polymorphs, hydrates, and solvates of drugs.[3−5] In particular, the crystalline state of a given compound influences its bioavailability, or just the production cost, as it may be easier to make drugs from one polymorph than from another. Drug regulatory authorities such as FDA in USA demand information about polymorphism before granting licenses for product distribution. Even patents have been made on the basis of the discovery of new polymorphs.[6−8]

Apart from polymorphs that have the same molecular composition, organic substances can form various solvates or hydrates, which is sometimes referred to as pseudopolymorphism.[9] Inclusion of more than one molecule in the asymmetric part of the unit cell ($Z' > 1$) represents yet another interesting crystal variation.[10,11] Complementary to the standard diffraction methods, the ^{13}C cross-polarization (CP) magic-angle spinning (MAS) NMR is also an efficient way of determining Z' by means of comparing the number of observed resonances with the number of nonequivalent carbon atoms present in the molecule.[5,12] Polymorphic crystals were also successfully studied by neutron diffraction.[13]

In the present study, we explore the potential of Raman spectroscopy to reliably distinguish three model polymorphic crystals. The low-resolution spectroscopic methods are often more practical than the X-ray diffraction or NMR, e.g., samples are easier to prepare and/or a smaller amount is required, and the spectra can be collected faster. Raman scattering is very sensitive to fine structural details in crystals and is thus increasingly used in analytical chemistry.[14−19] It measures the relative intensity of scattered light as dependent on its shift from the laser excitation frequency, usually caused by

Received: May 3, 2013
Revised: May 25, 2013
Published: May 30, 2013

vibrational motions in molecules. In terms of sample preparation, the Raman technique resembles X-ray powder diffraction, where polycrystalline samples are used as well.

Lately, the vibrational Raman spectroscopy in the solid state has been boosted by the possibility to reliably simulate spectra of infinite periodic systems by accurate quantum-chemical methods. In particular, the plane-wave density functional theory (DFT) methodology is convenient as it allows for the crystalline translational symmetry.[20] However, we are not aware of any systematic accuracy test with respect to the discrimination of the polymorphic crystal vibrational spectra. Especially for organic molecules bound by weak crystal forces, high-precision computations are required to detect the finer crystal structure differences.

To assess the potential for structural studies for the Raman experiment combined with the computations, we chose methacrylamide (**A**), piracetam (**B**), and 2-thiobarbituric acid (**C**) as model molecules representing typical organic systems forming polymorphic crystals (Figure 1). The differences

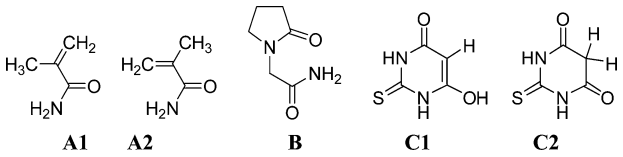


Figure 1. Methacrylamide (**A1** and **A2**, *s-cis* and *s-trans* conformers), piracetam (**B**), and 2-thiobarbituric acid (**C1** and **C2**, enol and keto forms) as model molecules examined in the present study.

between their polymorphs range from a subtle change in the crystal packing (piracetam) over conformational (methacrylamide) and tautomeric (2-thiobarbituric acid) variations. Within DFT, we model Raman spectra of isolated molecules as well as take into account the crystal environment of model systems. A cluster crystal model and a periodic-boundary condition plane-wave computation are employed.

The industrially important compound methacrylamide is, for example, a key intermediate in the acetonecyanohydrin process employed in manufacturing methyl methacrylate.[3] Of the two known polymorphs, the monoclinic form I (denoted here as **A1**) contains only the *s-cis* molecules, whereas the orthorhombic form II (**A2**) is exclusively formed by the *s-trans* conformer.[3]

Similarly, piracetam (2-oxo-pyrrolidineacetamide) is a nootropic agent, currently marketed by UCB Pharma as Nootropil. It is used to treat the age-related mental decline and similar disorders of the nervous system. Five polymorphs have been identified and their structures determined.[21] Among them, stable forms II (**B1**, triclinic, of *P*-1 symmetry) and III (**B2**, monoclinic, *P*21/*n*) can be prepared by recrystallization from various solvents under ambient conditions. For example, propan-2-ol and methanol provide forms II and III, respectively. Crystal structures of both forms have also been reported.[22,23] Unlike in **A**, however, molecular conformations **B1** and **B2** are almost identical. Two approximately planar parts, the pyrrolidine ring and the acetamide group (Figure 1), are almost perpendicular. The crystals are formed from dimers linked by two N−H···O hydrogen bonds. The crystal packing of the two forms looks quite similar if projected on the *ac* or *bc* crystallographic planes and differs most if observed along the *c*-axis (projection to *ab* plane).

Above 400 K both forms transform into form I (triclinic, *P*-1),[21,24] which at ambient temperature can be converted to phase II. High-pressure recrystallization of aqueous and methanolic solutions of piracetam resulted in the formation of yet another high-pressure polymorph IV, structurally characterized by single-crystal X-ray diffractometry.[21] A compression of form II by 0.45−0.70 GPa causes a reversible single-crystal to single-crystal transition to the last form V.[21]

As for 2-thiobarbituric acid, a substance valued since long time for its pharmacological properties,[25] six polymorphs have been isolated and characterized.[26,27] We have chosen forms II (**C1**) and III (**C2**) that can be prepared by recrystallization from absolute ethanol and dry acetonitrile, respectively. In **C1**, the molecules are present in the enol form, whereas only the keto isomer is present in **C2**. The other 2-thiobarbituric polymorphs and a hydrate also contain these two isomers; details can be found elsewhere.[26]

METHODS

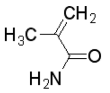
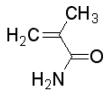
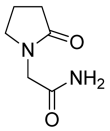
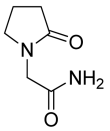
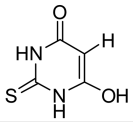
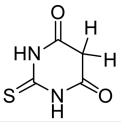
Crystal Preparation, X-ray, and NMR Characterization. Methacrylamide, piracetam, and 2-thiobarbituric acid were obtained from Sigma-Aldrich, and samples of different polymorphs were prepared as described in the literature listed above. Briefly, crystals of **A1** were obtained by recrystallization from a hot chloroform solution, and **A2** was prepared by slow cooling of warm water solution of methacrylamide. **B1** and **B2** were obtained by recrystallization of piracetam from propan-2-ol and methanol solutions, respectively. Polymorphs **C1** and **C2** were obtained by recrystallization of 2-thiobarbituric acid from dry ethanol and dry acetonitrile, respectively. Sample **C2** contained traces of **C1** (∼5%), and its contribution was subtracted from each resultant Raman spectrum of **C2**. Solid-state NMR spectra were recorded with Bruker Avance II 500 spectrometer operating at 499.8 MHz in the ^{1}H and at 125.7 MHz in the ^{13}C experiments, using rotors 3.2 mm in diameter. The chemical shifts were referenced to crystalline α-glycine as a secondary reference (δ_{st} = 176 ppm for carbonyl group carbon). The ramped amplitude shape pulse was used during the cross-polarization. The contact time in the basic ^{13}C experiments was 2 ms, the relaxation delay 4 s, and the spinning rate 12 kHz. The identity of the methacrylamide sample was also verified by X-ray measurement on KappaCCD four-circle diffractometer with a CCD area detector and Mo Kα radiation. The structures were solved by direct methods[28] and refined by full-matrix least-squares based on F2[29] with absorption neglected.

Raman Experiments. Solid-state Raman spectra within the 100−2000 cm^{-1} range were collected with BioTools ChiralRAMAN-2X instrument based on a design of Hug[30] (532 nm excitation wavelength, 32−128 scans, 1.03 s illumination time, laser power at sample 50−300 mW) and processed using in-house software.[31] Calibration was carried out with a neon lamp and verified with crystalline α-glycine. First, one polymorph was recorded and then the second one; this was repeated three times. As another control, the mixture of both polymorphic materials was measured and compared with individual polymorphs. Polycrystalline samples were measured in fused silica sample cell.

Alternatively, a near-infrared (NIR) Fourier spectrometer Equinox 55/S fitted with a Raman module FRA 106/S (Bruker, Germany) was employed to collect spectra excited by a Nd:YAG laser (1064 nm excitation, 1024 or 2048 scans, 2 cm^{-1} resolution, at least four spectra were averaged per sample).

dx.doi.org/10.1021/jp404382f | J. Phys. Chem. B 2013, 117, 7297−7307

Table 1. Overview of Studied Compounds and Their Crystal Structure

	Compound	Form **1**	Crystal[a] $a\ b\ c$ (Å) $\alpha\ \beta\ \gamma$ (deg.)	Form **2**	Crystal[a] $a\ b\ c$ (Å) $\alpha\ \beta\ \gamma$ (deg.)
A	methacrylamide Ref. [3]		WANSAG02 9.365 6.086 9.743 90.00 115.30 90.00		WANSAG01 5.934 10.242 16.436 90.00 90.00 90.00
B	Piracetam Ref. [22]		BISMEV 6.403 6.618 8.556 79.85 102.39 91.09		BISMEV01 6.525 6.440 16.463 90.00 92.19 90.00
C	2-thiobarbituric acid Ref. [26]		PABNAJ 9.874 8.776 7.066 90.00 109.30 90.00		THBARB01 9.523 7.703 8.002 90.00 105.68 90.00

[a]Cambridge Structural Database ID and unit cell parameters.

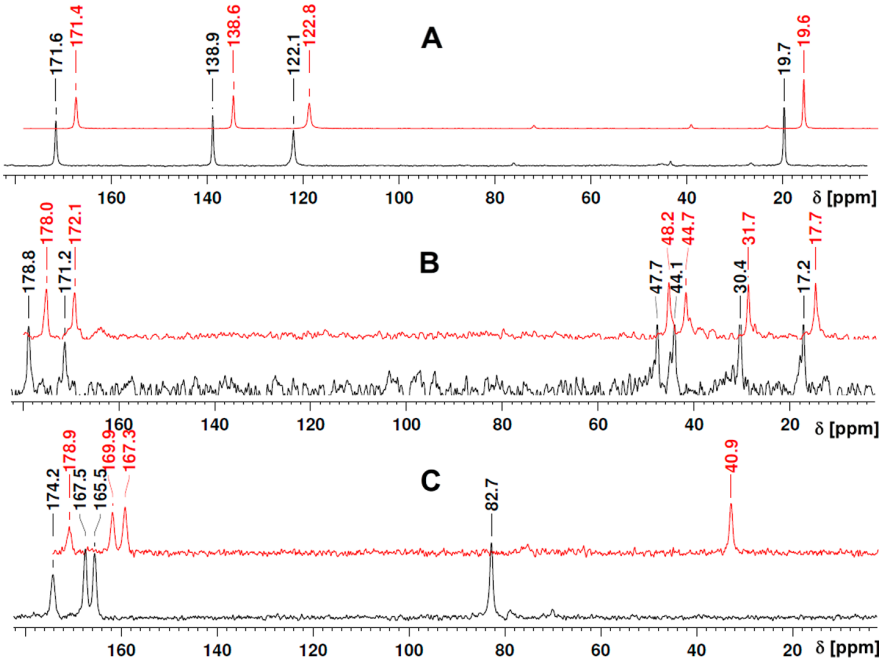


Figure 2. ^{13}C solid-state NMR spectra of polymorphic crystals. From top to bottom: **A1** vs **A2**, **B1** vs **B2**, and **C1** vs **C2**. For clarity, the forms **A1**, **B1**, and **C1** are plotted in red and offset by ~4 ppm; in **A**, the three low-intensity signals in the 20−80 ppm region correspond to spinning side bands of the sp^2 carbon atoms.

Powdered sulfur was used as a reference sample to check the wavenumber scale. A ChiralRAMAN-2X spectrometer provided a stronger signal, whereas the NIR-excited measurement reduced the fluorescence and provided a wider range of frequencies (−300 to 4500 cm^{-1}). The NIR Raman spectra are shown by default.

Raman and NMR Spectra Calculations. The atomic coordinates of the polymorphs of methacrylamide, piracetam, and 2-thiobarbituric acid were derived from the Cambridge Crystallographic Database[32] (see Table 1 for the reference codes). The shielding values of the infinite crystals, harmonic force fields, and Raman intensities were calculated by using the CASTEP program.[33] The calculations were based on X-ray structures with positions of all atoms optimized by energy minimization; the experimental lattice parameters were kept fixed. The generalized gradient approximation (GGA) functional of Perdew, Burke, and Ernzerhof (PBE)[34] and norm-conserving pseudopotentials were employed. Harmonic force field and Raman intensities within the periodic boundary conditions[35] were calculated at the same level as for the optimization. To verify reliability of the results, two combinations of the plane-wave cutoff energy and Monkhorst−Pack[36] grid spacing were applied (600 eV and 0.08 $Å^{-1}$ and 900 eV and 0.05 $Å^{-1}$; the latter setup providing better results on average is presented by default). With the optimized geometries, NMR shielding was calculated using the gauge-including projector-augmented wave approach (GIPAW)[37] and the "on the fly" pseudopotentials at 550 eV cutoff.[38]

dx.doi.org/10.1021/jp404382f | J. Phys. Chem. B 2013, 117, 7297−7307

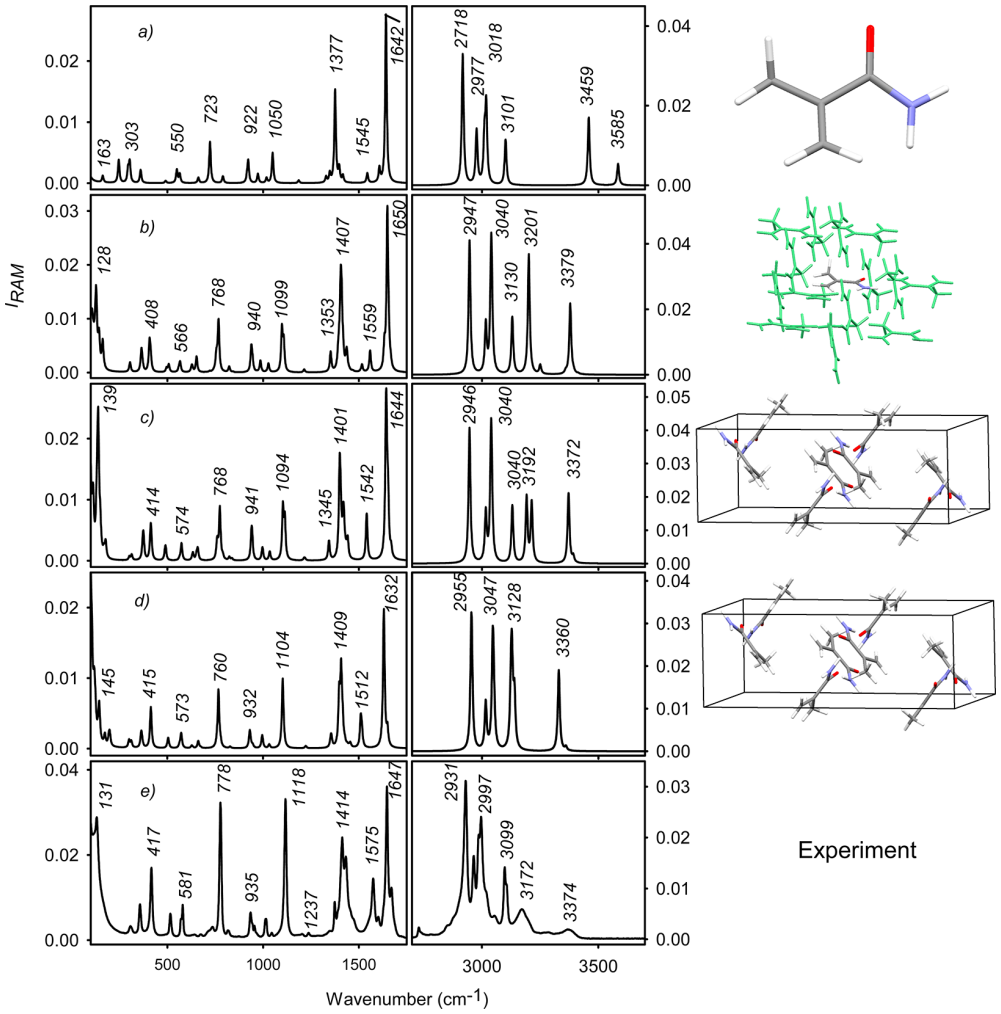


Figure 3. Raman spectra of methacrylamide, **A2**, computed for (a) single molecule and (b) a cluster. The (0,0,0) mode signals of the periodic elementary cell are in the (c) and (d) panels, while (e) is the experimental spectrum. Spectra a–c were obtained at the B3PW91/6-311++G**/6-31G/CPCM level and *x*-scaled by a factor of 0.96; (d) was obtained by the plane-wave simulation with the PBE functional.

As an alternative to the CASTEP "plane-wave" (PW) approach described above, we computed the Raman spectra with a cluster model and Gaussian atomic orbitals. First, the crystal cell obtained for each simulated system from the database was propagated to a "3 × 3 × 3" packing geometry, replicating elementary cell directions. Then clusters of a molecule chosen in the center and neighboring molecules closer than 4 Å were created with our own software. Finally, the geometries of the clusters were optimized in normal mode vibrational coordinates[39,40] with ω_{max}[41] of 300 cm^{-1}. Note that such constrained optimization leaves the crystal geometry virtually unchanged, whereas the vibrational coordinates relevant to Raman spectrum (frequencies typically greater than ω_{max}) can be relaxed.

The force field and polarizability derivatives of the clusters were then calculated by the Gaussian[42] program and transferred[43,44] back to the crystal cell. The BPW91[45] functional with the 6-31G and 6-31G** basis sets and the B3PW91[46,47] functional with the 6-311++G** and 6-31G basis sets for the central and neighboring molecules, respectively, were used. The reduction of the basis set was necessary to obtain results in a reasonable computational time. Supposedly, however, the effect of this approximation on the precision is limited as the diagonal force constants most important for the vibrational frequencies[48] were obtained from the central molecule with the larger basis set. The results were quite similar given the differences between the polymorphic forms; only the B3PW91/6-311++G**/6-31G calculations are shown, with the surrounding crystal environment simulated by placing the clusters in the CPCM model solvent[49] with relative electric permittivity ε_r = 78. We used the water permittivity as reported earlier[50] to approximately mimic the dielectric properties of polar organic crystals. While the CPCM correction provided slightly better results than vacuum computations, we verified that the simulated spectra were fairly independent of the variation of ε_r.

From the plane wave model and cluster force fields, zero (0, 0, 0) crystal phonon mode vibrational frequencies were calculated. Involvement of phonon modes with nonzero wave vectors had a negligible effect on the simulated intensities. Theoretical spectra were simulated with the Lorentzian bands, and the full width at half-maximum (Δ) was set 10 cm^{-1}, while applying the Boltzmann temperature correction. The spectrum at frequency ω was obtained from the computed backscattering Raman intensities (I_{180}) of each mode i with the harmonic

dx.doi.org/10.1021/jp404382f | J. Phys. Chem. B 2013, 117, 7297–7307

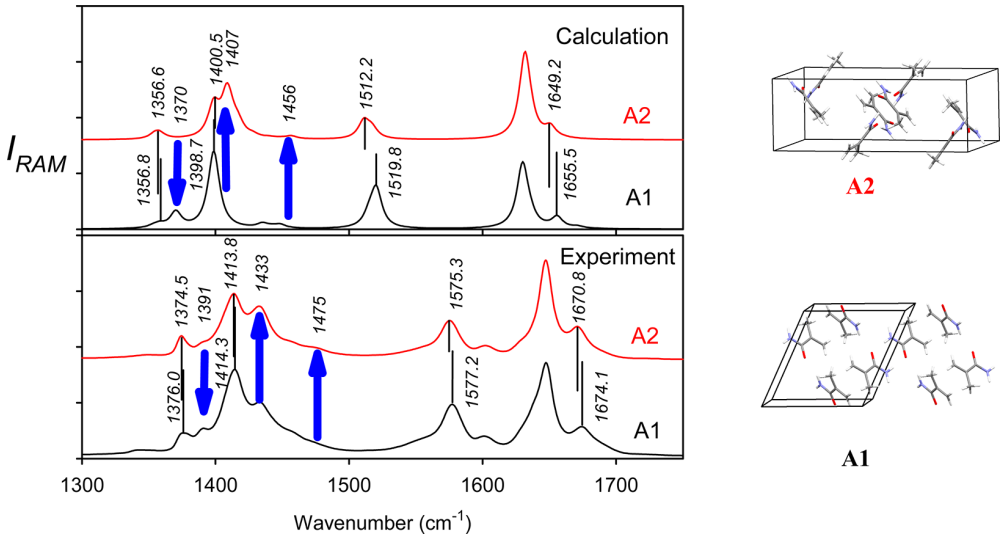


Figure 4. Calculated (top, plane wave) and experimental (bottom, NIR) Raman spectra of the two methacrylamide crystal forms, within 1300−1750 cm^{-1}. Main intensity changes are indicated by the blue arrows; selected peak maxima are indicated.

vibrational frequency ω_i as $S(\omega) = \omega_i^{-1}\{1 - \exp[-\omega_i/(kT)]\}^{-1}\{4[(\omega - \omega_i)/\Delta]^2 + 1\}^{-1}I_{180}$, where k is the Boltzmann constant and T is the temperature.

RESULTS

NMR Analysis. The NMR results were not the focus of the present study, as the detection of polymorphic forms by this technique has been previously described for a number of compounds.[5,26,51] In Figure 2, ^{13}C solid-state NMR spectra of polymorphic forms of methacrylamide (**A1**, **A2**), piracetam (**B1**, **B2**), and 2-thiobarbituric acid (**C1**, **C2**) are plotted. It can be clearly seen that the crystal packing induces specific shielding; i.e., the samples used in the Raman measurement correspond to the required forms.

Curiously, only nonpure preparations of methacrylamide batches, i.e., those also containing one or the other form, have been described so far.[3] As it is apparent from the NMR spectra (Figure 2), we did obtain pure forms of this compound. However, when sample **A2** (prepared from water solution) was not completely dried before the CP-MAS experiment, it slowly transformed to **A1**. The process could be monitored in a series of successive CP-MAS spectra, in which the **A2** signals were decreasing while the **A1** ones were increasing. The **A2** to **A1** transformation may have been facilitated by the combined effect of residual solvent and higher pressure in the MAS rotor spinning at 12 kHz because it was not observed when the sample was dry or wet and stored without spinning.

The calculations reproduce the experimental differences in chemical shifts of the polymorphic forms quite well. Except for one atom, all predicted signs agree with the experiment, with a small average error of 1.5 ppm (δ, see Table S1). These results are in accord with our earlier study, where the solid state NMR spectroscopy was found to be very suitable for identification of the polymorphic forms and the crystal packing.[51] The crystal packing is reflected in the chemical shift in two ways; it directly affects the electronic shielding by neighboring molecules and the geometry or conformational changes due to the packing itself.[52,53] Either of these contributions can be estimated by the comparison shifts calculated for an isolated molecule (while employing either X-ray data or optimized geometry; see second and third column in Table S1). However, as expected, the full periodic boundary computation (fourth column in Table S1) is needed to reach the best agreement with the experiment.

Theoretical Models Adopted for Raman Spectra. To understand the Raman spectral pattern, we adopted various models for its simulation. This is exemplified by the spectra of the **A2** form plotted in Figure 3. The single-molecule computation performed at the B3PW91/6-31++G**/CPCM level (panel a in Figure 3) is compared to the spectrum of the cluster (b). In addition, the elementary cell signal was obtained from (b) by the CCT transfer[44] (c) and as the plane-wave PBE result (d). The experiment is plotted in panel e. The B3PW91 frequencies were scaled by a factor of 0.96 for an easier comparison.

The effect of the packing interactions is clearly manifested in spectra shown in Figure 3. The isolated molecule (a) provides the basic spectral pattern; however, it deviates in some features from the experiment (e). For example, there is no signal (except for the weak band of methyl rotation at 163 cm^{-1}) at the lowest wavenumbers, unlike in the experiment with a strong peak at 131 cm^{-1}; the experimental band at 1575 cm^{-1} does not have a clear counterpart in the simulation, and the simulated bands at 3459 and 3585 cm^{-1} (in- and out-of-phase NH stretching, respectively) are too high.

The cluster model (b) provides a more realistic spectrum. In particular, the lowest wavenumber (∼100−600 cm^{-1}) signal rises and is better structured than on (a), which reflects the crystal packing forces. These modes are to a large extent composed of delocalized translations/rotations of various molecular parts. Also, the NH out-of-phase stretching frequency dramatically falls to 3379 cm^{-1}, and the in-phase mode (3201 cm^{-1} in (b)) is even below the CH_2 group stretching. This corresponds to the uneven length of the NH···O hydrogen bonds in the crystal, 2.029 and 2.093 Å,[3] contributing to the splitting of the NH stretching modes. It is also interesting to note that the PCM approximation of the bulk crystal adopted for the single-molecule spectrum a does not adequately represent the NH stretching frequency affected by the hydrogen bonding. Such inability of the polarizable continuum solvent models to fully describe the directional and

dx.doi.org/10.1021/jp404382f | J. Phys. Chem. B 2013, 117, 7297−7307

partially covalent hydrogen bonds has been well documented elsewhere.[54−60]

Finally, spectra c and d obtained from the periodic elementary cell and the (0, 0, 0) phonon mode are the most realistic ones. The cluster-based model (c) provides a more realistic hydrogen-stretching pattern, in particular the 3192 cm^{-1} signal observable at 3172 cm^{-1}, and intensities above 1200 and below 200 cm^{-1}. On the other hand, some features, mostly within 200−1200 cm^{-1}, are better reproduced by the plane-wave calculation (d).

A visualization of the normal mode displacement reveals an interesting shift of the amide I (C═O stretching) mode, in the single-molecule (a) hidden at the signal around 1642 cm^{-1}. This vibration heavily mixes with the in-phase NH bending and shifts downward in the crystal, to 1542 cm^{-1} for the cluster model c, experimentally at 1575 cm^{-1}. Note that the plane-wave approach can take into account the longer-scale electron conjugation and periodicity of the crystal structure, whereas the cluster model comprises interactions of the neighboring molecules only. On the other hand, the cluster approach enables one to use a more advanced functional (B3PW91, containing the Hartree−Fock exchange) and the more economical Gaussian basis set.

A1 and A2 Polymorphs. The differences between the individual polymorph spectra of methacrylamide (**A1** vs **A2**, calculation vs experiment) are documented in Figure 4 showing the 1300−1750 cm^{-1} interval where the most significant spectral changes occurred. As it can be seen, individual polymorphs of this molecule differ in minor shape changes of the Raman bands.

The relative intensity and frequency changes between individual peaks in **A1** and **A2** spectra can easily be inspected visually (Figure 4). Experimentally, for example, the intensity at 1374.5 cm^{-1} of **A2** partially moves to 1391 cm^{-1} for **A1**. Corresponding calculated peaks at 1356.6 and 1370 cm^{-1} reveal that this is caused by changes in mode coupling, rather than by shifts of the vibrational frequencies themselves. In particular, the CH_3 umbrella CH bending vibrations (experimentally ∼1376 cm^{-1}) couple differently with the other modes (NH_2, CH_2 bending, C−C stretching), providing thus the intensity boost at 1370 cm^{-1} in **A1**.

Similarly, another strong signal for **A2**, experimentally at 1433 cm^{-1}, calculated at 1407 cm^{-1}, originates in CH scissoring modes coupled with other CH bending and C−C stretching. A minor band of **A2** at 1475 cm^{-1} is caused primarily by CH_3 scissoring vibrations. The calculated band at 1512.2 cm^{-1} in **A2** shifts to 1519.8 cm^{-1} in **A1**, as in experiment, although the observed change is smaller and the absolute frequency higher, at ∼1576 cm^{-1}. Because this frequency is underestimated in the cluster model as well (e.g., at 1542 cm^{-1}, spectrum c, Figure 3), the lack of anharmonic corrections in the model is the most likely scenario. Presently, we are not aware of any other method providing reliable anharmonic corrections for systems of this size.

It is important to understand that the main differences in the polymorphic spectra have in this case their origin in the crystal packing. The cis and trans conformations of individual methacrylamide molecules provide different spectra (cf. Figure S1) but do not explain the crystal features (Figure 3) or the polymorphic differences (Figure 4). Interestingly, the **A1** and **A2** crystal Raman spectra (see Figure S2 for a comparison in the entire range of frequencies) resemble more each other than those of the isolated cis and trans conformers (Figure S1). In other words, crystal packing appears to smear the differences in the spectra of individual isomers.

Polymorphs B1 and B2. Piracetam behaves somewhat differently than methacrylamide. The **B1** and **B2** polymorphic forms are composed of very similar rotamers. The conformation in which the amide group is approximately perpendicular to the nearly planar five-membered ring is also conserved. The differences in the Raman spectra primarily arise from the different crystal packing, with the *P*-1 symmetry in **B1** and *P*21/*n* in **B2**. Even then, the packing is quite similar,[22] with a piracetam dimer being the basic unit held together by a dispersion attraction of the five-membered rings, and by hydrogen bonds between the NH_2 and exocyclic carbonyl groups. In **B1**, the dimer hydrogen bonds are slightly shorter (1.953 Å), and the ring is slightly more puckered than in **B2** (cf. the N−C_O−C−C angle of 7°, Figure 5).

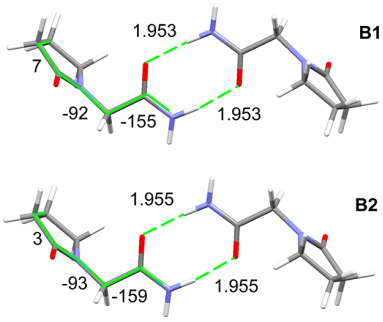


Figure 5. Piracetam dimer structure in two polymorphic crystal forms (distances in Å, torsional angles in deg).

The great degree of similarity between **B1** and **B2** causes only minor changes in the Raman spectra. The small effect of the packing on the spectra in an organic crystal is in agreement e.g. with a recent study[61] about vibronic effects and solid state circular dichroism. However, as it can be seen in Figure 6 where the experiment from the 1064 nm excitation is compared to the plane-wave calculation, the differences exhibited as fine frequency and intensity changes occur within the entire range of frequencies and are reasonably matched by the computation.

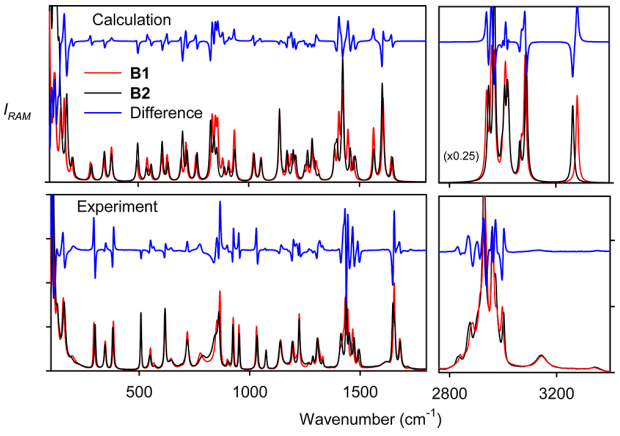


Figure 6. Calculated (top) and experimental (bottom) Raman spectra of the piracetam polymorphs **B1** and **B2**. The intensity scale is arbitrary, but consistent in calculation and experiment; the intensities in the upper right panel are scaled down by a factor of 0.25.

dx.doi.org/10.1021/jp404382f | *J. Phys. Chem. B* 2013, 117, 7297−7307

The experimental spectrum obtained with the 532 nm laser is quite similar, and its comparison with a computed spectrum is shown in Figure S3. The correspondence between the simulation and the experiment is less clear in the hydrogen stretching region (2800−3400 cm^{-1}, Figure 6), most likely due to anharmonic forces[62] that are difficult to include in the computation.[63] However, the theory provides correct signs of the intensity shifts of most bends including the most important intensity features of the difference spectrum plotted in Figure 6 in blue.

Alternatively, one can inspect the frequency change of individual vibrational bands for the two polymorphic forms. This analysis is more laborious than visual comparison of intensities as the corresponding peaks need to be identified. Nevertheless, as shown in Figure 7 where 10 vibrational bands

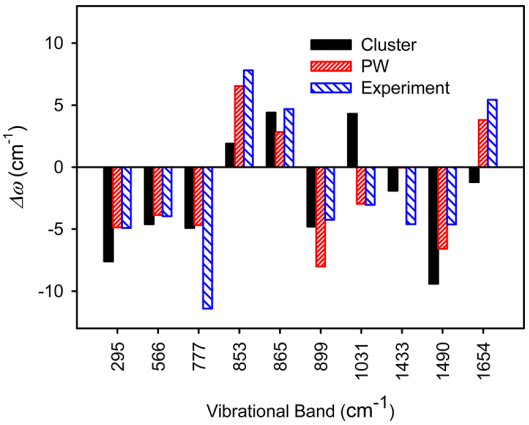


Figure 7. Ten largest frequency differences in the Raman bands between piracetam polymorphs, $\Delta\omega = \omega(\mathbf{B1}) - \omega(\mathbf{B2})$, calculated by the cluster and plane-wave method, and measured experimentally.

with the largest changes were selected, the frequency changes can also be quite reliably reproduced by the calculations. The plane-wave method appears to be more reliable than the cluster model; the latter one even yielded a frequency difference for the 1031 cm^{-1} Raman band with a wrong sign. Overall, however, the main trends are reproduced well. The large differences between the computation and experimental changes for 777 and 1433 cm^{-1} bands are at least partially explicable by an interference of close vibrations, indistinguishable at the currently available experimental resolution. On average, the changes computed by the PW method match the experiment with an error of only about 2 cm^{-1} and correlation coefficient of 0.739; the cluster results are noticeably worse (Figure S4).

The higher frequency (>2000 cm^{-1}) modes were not included in the statistics as they are difficult to assign and prone to anharmonic perturbations.[62,64] Quite often, modes involving large hydrogen atom movements provide the largest frequency differences. These include the backbone torsion and CH_2 group rotation (experimentally at 295 cm^{-1}), NH_2 bending (853 cm^{-1}), torsion (865 cm^{-1}), or CH_2 scissor modes (1433 and 1490 cm^{-1}). The modes involving the amide oxygen (out-of-plane deviation at 566 cm^{-1} and C═O stretching at 1606 cm^{-1}) also provide large frequency differences of about 5 cm^{-1} (Figure 7).

On piracetam, we can also demonstrate the effect of the energy cutoff and grid size on the computed Raman spectra (Figure S5). Both computational setups provided a reasonable spectral shape, and for some bands the cruder model (500 eV, 0.08 Å) gave a better agreement with the experiment than the finer one (900 eV, 0.05 Å). However, the finer setup clearly performed overall better in terms of frequencies and relative peak intensities. The differences between the two polymorphic forms were significantly overestimated by the cruder model. Higher computational precision is thus required to capture the tiny differences in Raman spectra, as caused by crystal packing.

2-Thiobarbituric Acid Polymorphs C1 and C2. This compound exhibits yet another behavior than the previous two systems, with the greatest differences among the Raman spectra of its polymorphs. The simulated (plane-wave) and NIR Raman experimental spectra are plotted in Figure 8. In the hydrogen stretching region (~2800−3300 cm^{-1}), for example, the **C1** form has only one distinct peak in the experimental spectrum (3104 cm^{-1}), whereas **C2** has three. This observation corresponds to the more extensive hydrogen bond network only compatible with the enol form **C1** (cf. Table 1).[26] Computation yields more peaks in this region, but we can suppose that the NH stretching signals (3110, 3126, and 3222 cm^{-1} in Figure 8) are manifested as a broad background only in the experiment, due to the strong anharmonic effects and polarity of the hydrogen bond, similarly as for the OH stretching.

The plane-wave computation provided an unreasonable value of the OH stretching frequency (~2000 cm^{-1}, not shown), most probably due to the inaccuracy of the PBE functional unable to describe the very short hydrogen bond and a rather long OH bond observed in the crystal. The =O···H and O−H distances were determined as 1.44 and 1.097 Å, respectively.[26] Indeed, the OH vibration can be better reproduced with the cluster model where the B3PW91 functional can be used and where the signal at 2000 cm^{-1} disappears. However, this does not improve the overall agreement with the experiment (cluster computation, Figure S6). As expected, the single-molecule spectra, also plotted in Figure S6, are not realistic either. Most probably, the weak experimental signals of the OH and NH stretching bands are caused by anharmonic interactions and consequent inhomogeneous band broadening; the experimental band at 3104 cm^{-1} can then be assigned to C−H stretching.

In the experimental Raman spectrum of **C2**, three strong bands appear in the high-frequency region (Figure 8). The peaks at 2874 and 2979 cm^{-1} are reproduced by the calculation (2886 and 3036 cm^{-1}) as C−H stretching bands. Interestingly, the rather unusual low frequency of the experimental 2874 cm^{-1} band seems to be caused by a weak hydrogen bond involving an aliphatic hydrogen and keto group oxygen. The third experimental band at the highest frequency (3097 cm^{-1}) is reproduced as a split 3108/3131 cm^{-1} signal, which unmasks yet another loophole in the theory regarding the polar hydrogen bonding.

The spectra of the **C1** and **C2** forms are even more complex in the lower-frequency region (<1800 cm^{-1}, Figure 8), and the theory can faithfully reproduce only some experimental features. It should be also noted that the experimental spectrum **C1** may contain incompletely subtracted contributions from **C2**. However, by a careful comparison, a reliable assignment of most peaks is possible. Only the **C2** form provides the C═O stretching ("amide I") peak at 1717 cm^{-1} in the experiment, reproduced at 1657 cm^{-1} by the model. In **C1**, this mode is mixed with C═C stretching, providing only a weak signal as reproduced by the calculation. The amide II mode (largely N−C stretching) shifts from the experimental 1553 cm^{-1} peak in

dx.doi.org/10.1021/jp404382f | J. Phys. Chem. B 2013, 117, 7297−7307

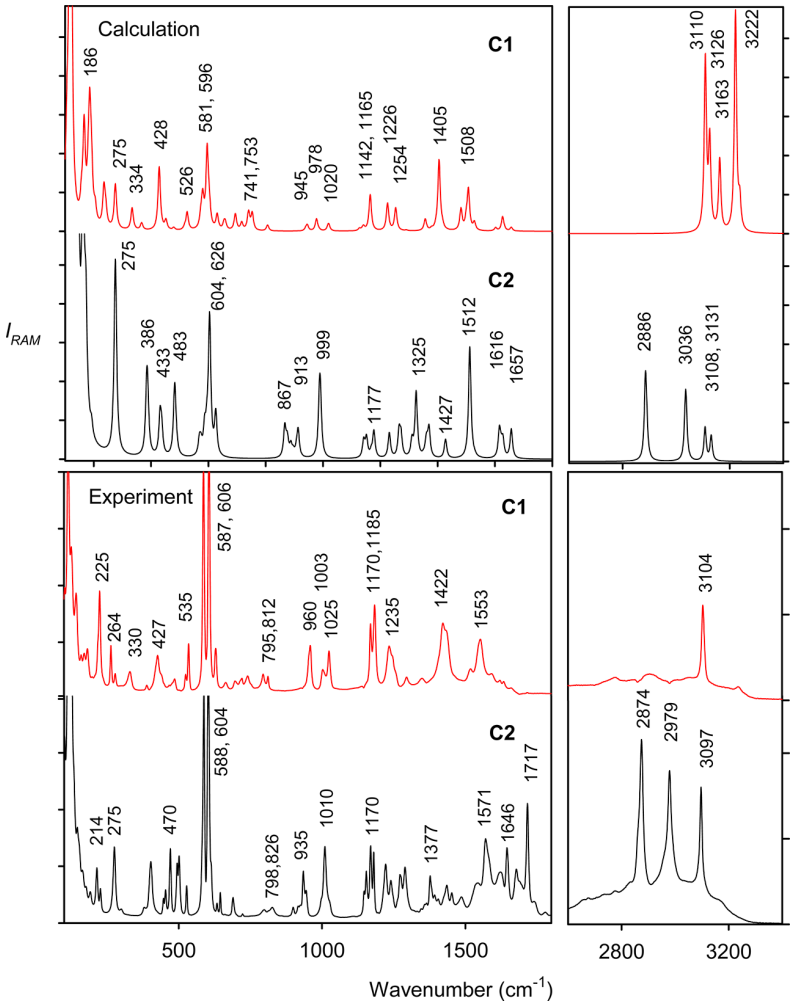


Figure 8. 2-Thiobarbituric acid: calculated (plane-wave) and NIR experimental Raman spectra of the two polymorphic forms **C1** and **C2**.

C1 to 1571 cm^{-1} in **C2**, similarly as in theory (1508 → 1512 cm^{-1}). A ring deformation in **C1** gives a strong signal both in the theoretical (1405 cm^{-1}) and observed (1422 cm^{-1}) spectrum, similarly as for the CH_2 scissoring mode in **C2** (exptl 1377 cm^{-1}, calcd 1325 cm^{-1}). The experimental C═S stretching band moves down from 1025 cm^{-1} (**C1**) to 1010 cm^{-1} (**C2**), as does the theoretical one (1020 → 999 cm^{-1}). A visual inspection of the normal modes reveals that C═S stretching is also mixed with C═N stretching, contributing to the signal around 1170 cm^{-1}.

A rather specific signal, clearly stronger in the **C2** form, is due to out-of-plane motions, such as NH bending; it is experimentally observed at 798−935 cm^{-1} and calculated at 867−913 cm^{-1}. The strongest experimental intensities, detected at 587−606 cm^{-1}, are underestimated by the computation. They originate in the out-of-plane C═O, C═S, and C−H group motions, which are notoriously difficult to describe within the harmonic approximation.[57,65] Finally, the lowest-wavenumber region is also rich in information, involving, for example, the 526/535 (calcd/exptl) cm^{-1} or the 428/427 cm^{-1} ring deformation band characteristic for **C1** and a very stable C═S bending frequency (experimentally at 264/275 cm^{-1} for **C1**/**C2**).

Although the polymorphic differences can be reliably reproduced by our models, an accuracy improvement is certainly still needed. It as very likely that the accuracy of the GGA force field and Raman intensities obtained within the PW computations represent the limiting factor; the cluster approach is more flexible in the basis set and functional choice but limited in the physical model. The absolute Raman intensities in particular are extremely sensitive to the basis set variations.[66] For the **C1**/**C2** polymorphic forms, a correction for anharmonic forces can potentially improve the hydrogen-stretching region (2800−3200 cm^{-1}). However, this is currently impossible for a system of this size. Additionally, larger deviations between the simulations and the experiment are also encountered in the lower-frequency region. Although this region is presumably well described at the harmonic level, frequency shifts of several cm^{-1} caused by anharmonic forces cannot be excluded. Future improvements of the method are thus dependent on the efficiency of implementing more precise electronic and vibrational quantum chemical methods.

Other structural information might also be hidden in the lattice modes, not analyzed in the present study. Our spectrometers do not enable measurement below ∼100 cm^{-1}. These vibrations are difficult to analyze because of the large background signal and a limited precision of calculated force

dx.doi.org/10.1021/jp404382f | J. Phys. Chem. B 2013, 117, 7297−7307

fields. Another important practical issue in the Raman spectra interpretation is the use of coordinates determined by X-ray diffraction. If one wants to replace X-ray spectroscopy, a pure a priory generation could be considered in the future for the initial coordinates as well. Currently, we propose the Raman technique as a convenient complementary way of examining the polymorphic forms in analytical practice.

CONCLUSIONS

We succeeded in preparing model polymorphic crystal forms of three model compounds including those of practical importance and characterized them by NMR and Raman spectroscopy. Pure polymorphs of methacrylamide were prepared, which has not been reported so far.

The NMR data confirmed sample purity and good performance of the technique in polymorphic discrimination. Raman spectroscopy proved to be an equally viable technique, as the spectra of different polymorphs exhibited distinct features that could be reliably reproduced by the density functional computations. Compared to green (532 nm) laser, excitation by a near-infrared one was somewhat more suitable as it suppressed sample fluorescence. The spectral shapes were similar in both experiments.

In spite of minor inconsistencies, the plane-wave approach, inherently comprising the crystal symmetry and periodicity, made it possible to reliably simulate spectral frequencies and intensities. Occasionally, the cluster-based computations were more advantageous as they could be combined with more theoretical approaches, in particular with different functionals. The studied examples represent a wide array of interpolymorphic differences manifested in the Raman spectra, from an almost pure crystal packing influence (piracetam) to profound changes in molecular structures (methacrylamide, 2-thiobarbituric acid). The computational models could successfully handle all cases, and we aim to further improve them so that even more complex systems can be reliably examined in the future. In any case, we demonstrated that Raman spectroscopy combined with computational modeling can significantly enhance the structural analysis of organic compounds' polymorphic forms.

ASSOCIATED CONTENT

Supporting Information

Calculated and experimental chemical shifts, additional computational and experimental details. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: bour@uochb.cas.cz (P.B.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was supported by the Academy of Sciences (M200551205), Czech Science Foundation (P208/11/0105 and 13-24880S), and Ministry of Education (LH11033). We thank Dr. Císařová for X-ray crystal characterization and Dr. Pelc for valuable comments.

REFERENCES

(1) Bernstein, E. R. Polymorphism - a Perspective. *Cryst. Growth Des.* **2011**, *11*, 632−650.
(2) Grant, D. J. W. Theory and Origin of Polymorphysm. In *Polymorphysm in Pharmaceutical Solids*; Brittain, H. G., Ed.; Marcel Dekker, Inc.: New York, 1999.
(3) Guo, C. Y.; Hickey, M. B.; Guggenheim, E. R.; Enkelmann, V.; Foxman, B. M. Conformational Polymorphism of Methacrylamide. *Chem. Commun.* **2005**, *17*, 2220−2222.
(4) Anderson, K. M.; Steed, J. W. Comment on "On the Presence of Multiple Molecules in the Crystal Asymmetric Unit (Z′ > 1)" by Gautam R. Desiraju. *CrystEngComm* **2007**, *9*, 328−330.
(5) Harris, R. K. NMR Studies of Organic Polymorphs & Solvates. *Analyst* **2006**, *131*, 351−373.
(6) Saurabh, G.; Kaushal, C. Pharmaceutical Solid Polymorphism in Abbreviated New Drug Application (ANDA) - a Regulatory Perspective. *J. Chem. Pharm. Res.* **2011**, *3*, 6−17.
(7) Datta, S.; Grant, D. J. W. Crystal Structures of Drugs: Advances in Determination, Prediction and Engineering. *Nat. Rev. Drug Discovery* **2004**, *3*, 42−57.
(8) Wu, V.; Rades, T.; Saville, D. J. Stability of Polymorphic Forms of Ranitidine Hydrochloride. *Pharmazie* **2000**, *55*, 508−512.
(9) Seddon, K. R. Perspective - Pseudopolymorph: A Polemic. *Cryst. Growth Des.* **2004**, *4*, 1087−1087.
(10) Anderson, K. M.; Probert, M. R.; Whiteley, C. N.; Rowland, A. M.; Goeta, A. E.; Steed, J. W. Designing Co-Crystals of Pharmaceutically Relevant Compounds That Crystallize with Z′ > 1. *Cryst. Growth Des.* **2009**, *9*, 1082−1087.
(11) Rummel, R. J. *Applied Factor Analysis*; Northwestern University Press: Evanston, 1970.
(12) Aliev, A. E.; Mann, S. E.; Rahman, A. S.; McMillan, P. F.; Cora, F.; Iuga, D.; Hughes, C. E.; Harris, K. D. M. High-Resolution Solid-State 2H NMR Spectroscopy of Polymorphs of Glycine. *J. Phys. Chem. A* **2011**, *115*, 12201−12211.
(13) Langan, P.; Mason, S. A.; Myles, D.; Schoenborn, B. P. Structural Characterization of Crystals of Alpha-Glycine during Anomalous Electrical Behaviour. *Acta Crystallogr., Sect. B* **2002**, *58*, 728−733.
(14) Pagliai, M.; Miranda, M. M.; Cardini, G.; Schettino, V. Raman and Infrared Spectra of Minerals from Ab Initio Molecular Dynamics Simulations: The Spodumene Crystal. *J. Mol. Struct.* **2011**, *993*, 151−154.
(15) Dračínský, M.; Benda, L.; Bouř, P. Comparison of Fused Silica, Crystal Quartz, and Water Raman Spectra. *Chem. Phys. Lett.* **2011**, *512*, 54−59.
(16) McGrane, S. D.; Barber, J.; Quenneville, J. Anharmonic Vibrational Properties of Explosives from Temperature-Dependent Raman. *J. Phys. Chem. A* **2005**, *109*, 9919−9927.
(17) Schweitzer-Stenner, R.; Sieler, G.; Mirkin, N. G.; Krimm, S. Intermolecular Coupling in Liquid and Crystalline States of Trans-N-Methylacetamide Investigated by Polarized Raman and FT-IR Spectroscopies. *J. Phys. Chem. A* **1998**, *102*, 118−127.
(18) Kim, I.; Krimm, S. Raman Longitudinal Acoustic Mode Studies of Poly(Ethylene Oxide) and Alpha, Omega-Methoxylated Poly(Ethylene Oxide) Fractions during Isothermal Crystallization from the Melt. *J. Polym. Sci., Part B: Polym. Phys.* **1997**, *35*, 1117−1126.
(19) Othman, S.; Richaud, P.; Vermeglio, A.; Desbois, A. Evidence for a Proximal Histidine Interaction in the Structure of Cytochromes C′ in Solution - a Resonance Raman Study. *Biochemistry* **1996**, *35*, 9224−9234.
(20) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd ed.; Academic: London, 2002.
(21) Fabbiani, F. P. A.; Allan, D. R.; David, W. I. F.; Davidson, A. J.; Lennie, A. R.; Parsons, S.; Pulham, C. R.; Warren, J. E. High-Pressure Studies of Pharmaceuticals: An Exploration of the Behavior of Piracetam. *Cryst. Growth Des.* **2007**, *7*, 1115−1124.
(22) Admiraal, G.; Eikelenboom, J. C.; Vos, A. Structures of the Triclinic and Monoclinic Modifications of (2-Oxo-1-Pyrrolidinyl)-Acetamide. *Acta Crystallogr., Sect. B* **1982**, *38*, 2600−2605.

(23) Bandoli, G.; Clemente, D. A.; Grassi, A.; Pappalardo, G. C. Molecular Determinants for Drug-Receptor Interactions. 1. Solid-State Structure and Conformation of the Novel Nootropic Agent 2-Pyrrolidone-N-Acetamide - X-Ray and Theoretical SCF-MO Studies. *Mol. Pharmacol.* **1981**, *20*, 558−564.
(24) Louër, D.; Louër, M.; Dzyabchenko, V. A.; Agafonov, V.; Céolin, R. Structure of a Metastable Phase of Piracetam from X-Ray-Powder Diffraction Using Atom-Atom Potential Method. *Acta Crystallogr., Sect. B* **1995**, *51*, 182−187.
(25) Bondock, S.; Tarhoni, A. E.; Fadda, A. A. *Phosphorus, Sulfur Silicon Relat. Elem.* **2007**, *182*, 1915−1936.
(26) Chierotti, M. R.; Ferrero, L.; Garino, N.; Gobetto, R.; Pellegrino, L.; Braga, D.; Grepioni, F.; Maini, L. The Richest Collection of Tautomeric Polymorphs: The Case of 2-Thiobarbituric Acid. *Chem.—Eur. J.* **2010**, *16*, 4347−4358.
(27) Calas, M. R.; Martinez, J. Determination de la Structure Cristalline de l'Acide Thiobarbiturique. *C. R. Acad. Sci., Ser. C* **1967**, *265*, 631−631.
(28) Altomare, A.; Cascarano, G.; Giacovazzo, C.; Guagliardi, A.; Burla, M. C.; Polidori, G.; Camalli, M. SIR92 - a Program for Automatic Solution of Crystal Structures by Direct Methods. *J. Appl. Crystallogr.* **1994**, *27*, 435.
(29) Sheldrick, G. M. A Short History of SHELX. *Acta Crystallogr., Sect. A* **2007**, *64*, 112−122.
(30) Hug, W.; Hangartner, G. A Novel High-Throughput Raman Spectrometer for Polarization Difference Measurements. *J. Raman Spectrosc.* **1999**, *30*, 841−852.
(31) Šebestík, J.; Bouř, P. Raman Optical Activity of Methyloxirane Gas and Liquid. *J. Phys. Chem. Lett.* **2011**, *2*, 498−502.
(32) Allen, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr., Sect. B* **2002**, *58*, 380−388.
(33) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. J.; Refson, K.; Payne, M. C. First Principles Methods Using CASTEP. *Z. Kristallogr.* **2005**, *220*, 567−570.
(34) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(35) Refson, K.; Tulip, P. R.; Clark, S. J. Variational Density-Functional Perturbation Theory for Dielectrics and Lattice Dynamics. *Phys. Rev. B* **2006**, *73*, 155114.
(36) Monkhorst, H. J.; Pack, J. D. Special Points for Brillouin-Zone Integrations. *Phys. Rev. B* **1976**, *13*, 5188−5192.
(37) Pickard, C. J.; Mauri, F. All-Electron Magnetic Response with Pseudopotentials: NMR Chemical Shifts. *Phys. Rev. B* **2001**, *6324*, 245101.
(38) Yates, J. R.; Pickard, C. J.; Mauri, F. Calculation of NMR Chemical Shifts for Extended Systems Using Ultrasoft Pseudopotentials. *Phys. Rev. B* **2007**, *76*, 024401.
(39) Bouř, P.; Keiderling, T. A. Partial Optimization of Molecular Geometry in Normal Coordinates and Use as a Tool for Simulation of Vibrational Spectra. *J. Chem. Phys.* **2002**, *117*, 4126−4132.
(40) Bouř, P. Convergence Properties of the Normal Mode Optimization and Its Combination with Molecular Geometry Constraints. *Collect. Czech. Chem. Commun.* **2005**, *70*, 1315−1340.
(41) Hudecová, J.; Hopmann, K. H.; Bouř, P. Correction of Vibrational Broadening in Molecular Dynamics Clusters with the Normal Mode Optimization Method. *J. Phys. Chem. B* **2012**, *116*, 336−342.
(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09, Revision B01*; Gaussian, Inc.: Wallingford, CT, 2009.
(43) Bouř, P.; Sopková, J.; Bednárová, L.; Maloň, P.; Keiderling, T. A. Transfer of Molecular Property Tensors in Cartesian Coordinates: A New Algorithm for Simulation of Vibrational Spectra. *J. Comput. Chem.* **1997**, *18*, 646−659.
(44) Yamamoto, S.; Li, X.; Ruud, K.; Bouř, P. Transferability of Various Molecular Property Tensors in Vibrational Spectroscopy. *J. Chem. Theory Comput.* **2012**, *8*, 977−985.
(45) Becke, A. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(46) Perdew, J. P.; Burke, K.; Wang, Y. Generalized Gradient Approximation for the Exchange-Correlation Hole of a Many-Electron System. *Phys. Rev. B* **1996**, *54*, 16533−16539.
(47) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(48) Bouř, P.; Andrushchenko, V.; Kabeláč, M.; Maharaj, V.; Wieser, H. Simulations of Structure and Vibrational Spectra of Deoxyoctanucleotides. *J. Phys. Chem. B* **2005**, *109*, 20579−20578.
(49) Klamt, A. COSMO and COSMO-RS. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; Vol. *1*, pp 604−615.
(50) Kapitán, J.; Baumruk, V.; Kopecký, V., Jr.; Pohl, R.; Bouř, P. Proline Zwitterion Dynamics in Solution, Glass and Crystalline State. *J. Am. Chem. Soc.* **2006**, *128*, 13451−13462.
(51) Dračínský, M.; Buděšínský, M.; Warzajtis, B.; Rychlewska, U. Solution and Solid-State Effects on NMR Chemical Shifts in Sesquiterpene Lactones: NMR, X-Ray, and Theoretical Methods. *J. Phys. Chem. A* **2012**, *116*, 680−688.
(52) Dračínský, M.; Bouř, P. Computational Analysis of Solvent Effects in NMR Spectroscopy. *J. Chem. Theory Comput.* **2010**, *6*, 288−299.
(53) Uldry, A. C.; Griffin, J. M.; Yates, J. R.; Perez-Torralba, M.; Maria, M. D. S.; Webber, A. L.; Beaumont, M. L. L.; Samoson, A.; Claramunt, R. M.; Pickard, C. J.; Brown, S. P. Quantifying Weak Hydrogen Bonding in Uracil and 4-Cyano-4'-Ethynylbiphenyl: A Combined Computational and Experimental Investigation of NMR Chemical Shifts in the Solid State. *J. Am. Chem. Soc.* **2008**, *130*, 945−954.
(54) Ganim, Z.; Chung, H. S.; Smith, A. W.; Deflores, L. P.; Jones, K. C.; Tokmakoff, A.; Amide, I. Two-Dimensional Infrared Spectroscopy of Proteins. *Acc. Chem. Res.* **2008**, *41*, 432−441.
(55) Lee, K. K.; Hahn, S.; Oh, K. I.; Choi, J. S.; Joo, C.; Lee, H.; Han, H.; Cho, M. Structure of N-Acetylproline Amide in Liquid Water: Experimentally Measured and Numerically Simulated Infrared and Vibrational Circular Dichroism Spectra. *J. Phys. Chem. B* **2006**, *110*, 18834−18843.
(56) Choi, J. H.; Kim, J. S.; Cho, M.; Amide, I. Vibrational Circular Dichroism of Polypeptides: Generalized Fragmentation Approximation Method. *J. Chem. Phys.* **2005**, *122*, 174903.
(57) Andrushchenko, V.; Matějka, P.; Anderson, D. T.; Kaminský, J.; Horníček, J.; Paulson, L. O.; Bouř, P. Solvent Dependence of the N-Methylacetamide Structure and Force Field. *J. Phys. Chem. A* **2009**, *113*, 9727−9736.
(58) Bouř, P.; Michalík, D.; Kapitán, J. Empirical Solvent Correction for Multiple Amide Group Vibrational Modes. *J. Chem. Phys.* **2005**, *122*, 144501.
(59) Bouř, P. On the Influence of the Water Electrostatic Field on the Amide Group Vibrational Frequencies. *J. Chem. Phys.* **2004**, *121*, 7545−7548.
(60) Bouř, P.; Keiderling, T. A. Empirical Modeling of the Peptide Amide I Band IR Intensity in Water Solution. *J. Chem. Phys.* **2003**, *119*, 11253−11262.
(61) Pescitelli, G.; Padula, D.; Santoro, F. Intermolecular Exciton Coupling and Vibronic Effects in Solid-State Circular Dichroism: A Case Study. *Phys. Chem. Chem. Phys.* **2013**, *15*, 795−802.
(62) Daněček, P.; Bouř, P. Comparison of the Numerical Stability of Methods for Anharmonic Calculations of Vibrational Molecular Energies. *J. Comput. Chem.* **2007**, *28*, 1617−1624.
(63) Dračínský, M.; Bouř, P. Vibrational Averaging of the Chemical Shift in Crystalline α-Glycine. *J. Comput. Chem.* **2012**, *33*, 1080−1089.
(64) Temelso, B.; Shields, G. C. The Role of Anharmonicity in Hydrogen-Bonded Systems: The Case of Water Clusters. *J. Chem. Theory Comput.* **2011**, *7*, 2804−2817.
(65) Kapitán, J.; Hecht, L.; Bouř, P. Raman Spectral Evidence of Methyl Rotation in Liquid Toluene. *Phys. Chem. Chem. Phys.* **2008**, *10*, 1003−1008.

(66) Zuber, G.; Hug, W. Rarefied Basis Sets for the Calculation of Optical Tensors. 1. The Importance of Gradients on Hydrogen Atoms for the Raman Scattering Tensor. *J. Phys. Chem. A* **2004**, *108*, 2108−2118.

9.2 Appendix B – *Chirality* 2016, 28, 420.

Circular Dichroism is Sensitive to Monovalent Cation Binding in Monensin Complexes

AHMED NEDZHIB,[1] JIŘÍ KESSLER,[2] PETR BOUŘ,[2]* BÉLA GYURCSIK,[3] AND IVAYLA PANTCHEVA[1]*

[1]*Department of Analytical Chemistry, Faculty of Chemistry and Pharmacy, Sofia University "St. Kl. Ohridski,", Sofia, Bulgaria*
[2]*Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Prague, Czech Republic*
[3]*Department of Inorganic and Analytical Chemistry, University of Szeged, Szeged, Hungary*

ABSTRACT Monensin is a natural antibiotic that exhibits high affinity to certain metal ions. In order to explore its potential in coordination chemistry, circular dichroism (CD) spectra of monensic acid A (MonH) and its derivatives containing monovalent cations (Li^+, Na^+, K^+, Rb^+, Ag^+, and Et_4N^+) in methanolic solutions were measured and compared to computational models. Whereas the conventional CD spectroscopy allowed recording of the transitions down to 192 nm, synchrotron radiation circular dichroism (SRCD) revealed other bands in the 178–192 nm wavelength range. CD signs and intensities significantly varied in the studied compounds, in spite of their similar crystal structure. Computational modeling based on the Density Functional Theory (DFT) and continuum solvent model suggests that the solid state monensin structure is largely conserved in the solutions as well. Time-dependent Density Functional Theory (TDDFT) simulations did not allow band-to-band comparison with experimental spectra due to their limited precision, but indicated that the spectral changes were caused by a combination of minor conformational changes upon the monovalent cation binding and a direct involvement of the metal electrons in monensin electronic transitions. Both the experiment and simulations thus show that the CD spectra of monensin complexes are very sensitive to the captured ions and can be used for their discrimination. *Chirality 28:420–428, 2016.*

KEY WORDS: monovalent polyether ionophore; metal complexes; synchrotron radiation circular dichroism; time-dependent density functional theory

Monensin is a natural antibiotic produced by *Streptomyces cinnamonensis*.[1–3] It is widely applied in stock farming and veterinary medicine due to its pronounced coccidiostatic and antibacterial properties.[4–14] The main form of the ionophore is monensin A (monensic acid, MonH), accompanied by two minor factors, monensin B and monensin C, also produced by the *Streptomyces* bacteria. From a chemical point of view, monensin A is a polyether derivative of a monocarboxylic acid (Scheme 1). Its monohydrated form ($MonH \times H_2O$) exists in a pseudo-cyclic conformation secured by head-to-tail H-bonding between the carboxylic moiety and the alcoholic hydroxyl group of the last six-membered ring (O11), with a supplementary binding of a water molecule.[15,16] Oxygen atoms pointing inside the cavity ensure its hydrophilic character, while the alkyl-rich polyether backbone provides antibiotic lipophilicity and corresponding cell membrane activity.

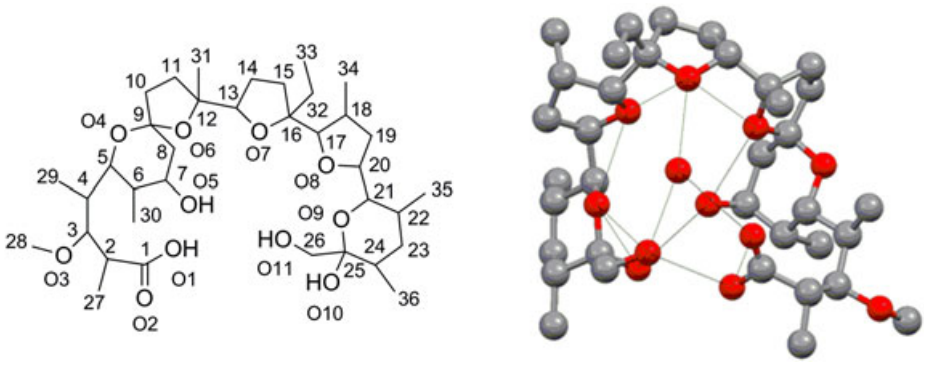
Another interesting property of monensin is its ability to form complexes with certain monovalent metal cations. The antibiotic acts as a monoanion through deprotonated carboxylic function, assuring an overall neutral charge of the complex. These complexes can also easily penetrate bacteria's cell membranes via the so-called electrogenic and nonelectrogenic mechanisms.[17–24] Inside the cell dissociation processes occur, leading to disturbance of pH and metal ion equilibria. Subsequent changes activate a variety of further events, ultimately leading to cell death. Better understanding of the metal ion complexation of monensin in solution will contribute to elucidation of the details of the above processes.

The affinity of monensin to bind monovalent metal ions decreases in the order of $Ag^+ > Na^+ > K^+ > Rb^+ > Li^+ \sim Cs^+$.[25] Molecular geometries of complexes with lithium, sodium, potassium, rubidium, and silver cations were determined by X-ray diffraction on single crystals.[26–35] The crystal forms are very similar, with the monovalent metal ion trapped in the central cavity and coordinated with at least six oxygen (-O-, -OH) donor atoms. Carboxylate oxygens do not participate in the binding, but two hydrogen bonds between the carboxylate group and alcoholic OH groups at the opposite end of the molecule stabilize the pseudo-cyclic structure, which is slightly differently than in free ligand.

MonH and its monovalent metal complexes exhibit a very low solubility in water, but are readily soluble in methanol, ethanol, acetone, or chloroform. These solutions were studied in potentiometric,[36–42] NMR,[43–47] polarographic,[48] and mass-spectrometric[49–54] experiments. Natural polyether ionophores such as lasalocid and salinomycin were studied by circular dichroism (CD) in the near UV range,[55–59] but because of the lack of suitable chromophores this technique was not employed so far for monensin.

Let us recall that the use of CD spectroscopy, at least as a complementary method, appears convenient for many reasons. It is in general more sensitive with respect to the sample amount and structural changes than infrared absorption, the experiment is simpler than X-ray diffraction, NMR, or mass-spectroscopy, it is applicable to solutions, unlike

*Correspondence to: P. Bouř, Institute of Organic Chemistry and Biochemistry, Academy of Sciences, 2, Flemingovo nám., 16610 Prague, Czech Republic. E-mail: petr.bour@uochb.cas.cz; I. Pantcheva, Department of Analytical Chemistry, Faculty of Chemistry and Pharmacy, Sofia University "St. Kl. Ohridski", 1, J. Bourchier blvd., 1164 Sofia, Bulgaria. E-mail: ipancheva@chem.uni-sofia.bg
Received for publication 2 November 2015; Accepted 25 February 2016
DOI: 10.1002/chir.22597
Published online 7 April 2016 in Wiley Online Library (wileyonlinelibrary.com).



Scheme 1. Numbering scheme of monensin A (left) and MonH × H_2O crystal structure [16] (right). Protons are omitted for clarity and H-bond distances shorter than 3.00 Å are indicated.

X-ray diffraction, etc. On the other hand, CD often provides rather limited resolution and a number of spectral features; these disadvantages can be at least partially sorted out by the possibility to interpret the spectra on the basis of parameter-free quantum-chemical computations.

In the present article the potency of conventional CD and synchrotron radiation circular dichroism (SRCD) spectroscopy is explored to evaluate the complexation ability of monensin A with respect to monovalent cations in solution. The experimental data are discussed on the basis of computational modeling. Density Functional Theory (DFT) and a dielectric solvent model were used to estimate solution geometries, and time-dependent DFT (TDDFT) was used to simulate the absorption and CD spectra.

MATERIALS AND METHODS

Materials

Sodium monensinate (MonNa) was kindly provided by Biovet (Peshtera, Bulgaria). Metal(I) salts, Et_4NOH, and methanol of analytical grade were supplied by Merck/Fluka (Darmstadt, Germany).

Monensic acid (MonH × H_2O), tetraethylammonium monensinate ($MonNEt_4$), and monovalent metal complexes MonM (M = K, Rb, Li, Ag) were prepared as described previously.[26,40] The complex formation was confirmed by IR spectroscopy (Supplementary Fig. S1, FT-IR Nikolet 6700 spectrophotometer, Thermo Scientific, Waltham, MA, KBr pellet). Isolated solid-state monensin complexes were dissolved in methanol for subsequent measurements. The data from titrations of MonH with monovalent metal ions fit well these (Fig. S2). $MonNEt_4$ was obtained *in situ*.

CD Spectroscopy

CD spectroscopic measurements were performed on a Jasco (Tokyo, Japan) J-815 spectrometer with solution samples (concentration of 5–20 mmol dm^{-3}, temperature of 25 °C) kept in a fused silica cuvette of 0.2 mm optical pathlength. The spectra were recorded in the 180–300 nm range, using 0.5 nm resolution, 2 sec response time, and a scanning speed of 20 nm/min.

Synchrotron radiation CD (SRCD) spectra were recorded at the AU-CD beam line SRCD facility, part of the ASTRID2 storage ring at the Institute for Storage Ring Facilities (ISA), University of Aarhus, Aarhus, Denmark.[60,61] The compounds were dissolved in methanol to concentrations of 40–100 mmol dm^{-3}. All spectra were recorded at 24.4 °C in 1-nm steps with a dwell time of 2 sec per step, in the wavelength range of 170–300 nm and with resolution of 0.5 nm. Spectra of sodium monensinate (MonNa) were recorded using both 0.2 mm and 0.014 mm cuvettes, whereas 0.014 mm only was used for the rest.

Two accumulations were averaged both for the CD and SRCD measurements. The molar absorbance and molar ellipticity of compounds were calculated after subtraction of the solvent (methanol) spectra acquired at identical conditions.

Calculations

X-ray structures of monensic acid (MonH × H_2O - MONSND)[15] and its monovalent metal complexes MonM (M = Li^+ - MIPSIO,[33] Na^+ - DEYGAQ,[32] K^+ - FECROU10,[30] Rb^+ - RITLIQ,[34] Ag^+ - MONSIN10[26]) were used as starting geometries. The structures were fully optimized with the Gausian09.Rev.D01 program[62] using the B3LYP functional[63] and the conductor-like polarizable continuum solvent model (CPCM)[64] to account for the methanol environment. CAM-B3LYP, invented to improve B3LYP, B3PW91, LC-WPBE, and WB97XD functionals were also applied, but did not give better results than the standard B3LYP (Fig. S3).

Alternatively to the full optimization, X-ray geometries, partially optimized in the normal mode coordinates were used as well; normal modes with frequencies $|\omega_i| < 300\ cm^{-1}$ were fixed.[65] The partial optimization corrected in particular bond lengths and angles of the hydrogen atoms, determined with a big error or completely missing in the crystal structures.

The 6-311++G** basis set was used for the carboxyl group atoms, the MWB28 pseudopotential[66] and basis set were used for silver and rubidium atoms, and the 6-31G** basis set was used for the rest. For the optimized structures, UV and CD spectra were calculated at the TDDFT[67]/CPCM level. For each system 100 electronic excited states were obtained to cover the experimentally observable spectral range.

RESULTS AND DISCUSSION

To the best of our knowledge, we report for the first time the spectral changes occurring upon complexation of monensin A, evaluated by means of CD spectroscopy. As some readers might not be familiar with the SRCD technique, we compare classical CD and SRCD measurements done using the same 0.2 mm optical pathlength cell. The CD and SRCD spectra of sodium monensinate (MonNa), as well as the total absorbances derived from the same measurements, are plotted in Figure S4. The careful comparison of the two techniques in this setup showed only minor differences. Keeping the total absorbance value below 2 is a prerequisite to obtain reliable CD spectra. By means of conventional CD and SRCD under the same experimental conditions, such data were obtained in the UV range down to 192 and 187 nm, respectively. In both CD and SRCD the absorbance of the solvent used (MeOH) with and without MonNa is almost the same, demonstrating the negligible contribution of the dissolved substance to the total absorbance. The CD and SRCD spectra of MonNa correspond reasonably well to each other. Below 192 nm, the conventional CD quickly deteriorates due to the high absorption and low sensitivity.

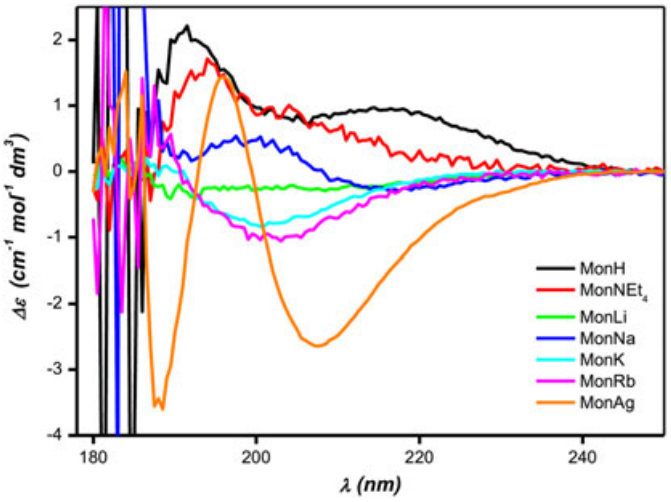


Fig. 1. Experimental CD spectra of MonH and its complexes with six monovalent ions (0.2 mm optical pathlength). The intensities within 180–192 nm are not reliable; they are shown for comparison with the SRCD spectra in Figure 2.

The CD spectra of MonH and all the monovalent complexes (recorded at 0.2 mm optical pathlength) are presented in Figure 1. The results show that the position and sign of CD signals significantly depend on the protonation state of the ligand and coordinated cation. MonH, for example, exhibits a positive sign within 190–245 nm with two maxima at 218 and 192 nm, while the replacement of H^+ with the Et_4N^+ cation diminishes the 218 nm band (although it remains still positive) and gives rise to a positive maximum at 195 nm. The observed spectral variance can be probably attributed to formation of an alternative hydrogen-bonding network. Most probably, MonH stays in the "closed" conformation, as it was found most favorable for the anionic form.[15,68]

The trapping of monovalent metal ions into the ligand cavity causes significant changes in the CD spectra. The complexation of sodium ions is accompanied by the appearance of two bands, at 200 nm (positive) and 216 nm (negative). The coordination of lithium ions provides a negative signal within 190–230 nm, while the CD spectrum of the silver complex contains positive (196 nm) and negative (208 nm)

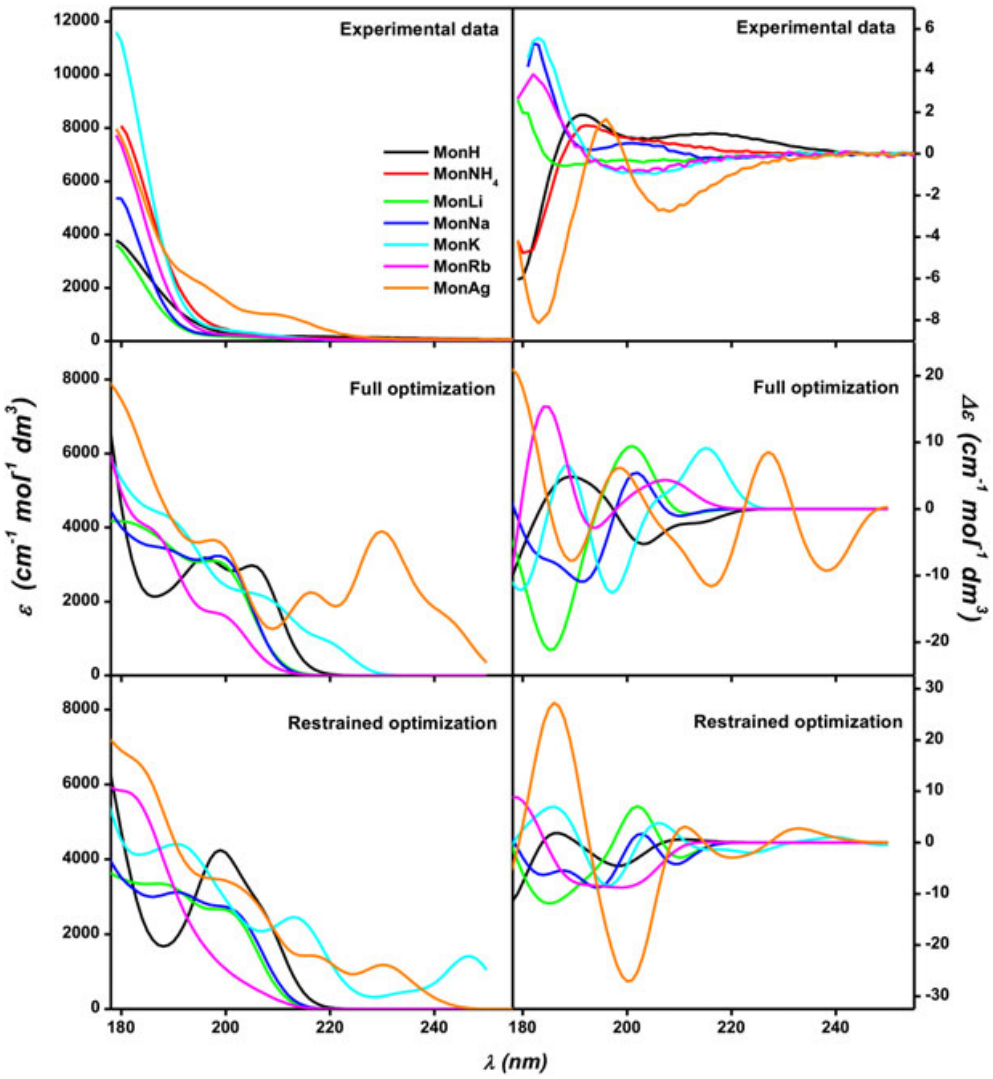


Fig. 2. Experimental and calculated absorption (left) and SRCD (right) spectra of MonH and MonM (M = Li^+, Na^+, K^+, Rb^+, Ag^+). The measurement was performed with 0.014 mm optical pathlength. The B3LYP/PCM/6-311++G**/6-31G** level was used in the computations; results for fully and partially ("X-ray") optimized geometries are shown.

bands. The potassium and rubidium complexes of monensin are mostly characterized by negative bands at 201–203 nm.

We were able to obtain additional data in the far UV region by SRCD spectroscopy using a shorter optical pathlength of 0.014 mm (Fig. 2). The spectra were cut at 178 nm by the limitation due to the high absorbance values (hight tension voltage was greater than 5 kV below this wavelength). Absorption spectra of all compounds are quite similar, except for the MonAg, with a characteristic shoulders at longer wavelengths (200 and 218 nm). The absorption intensity generally increased upon exchange of H^+ with monovalent cations. Low-wavelength CD bands unseen by conventional CD are of high intensity and provide an additional possibility to distinguish different monensin complexes. The experimental spectra of MonH and $MonNEt_4$ possess negative signal below 190 nm, and silver complex has a strong negative band at 185 nm. The coordination of lithium ions leads to the appearance of a negative band at 187 nm accompanied by a negative shoulder at ~200–220 nm. Spectral shapes of the MonK and MonRb complexes in the range of 180–185 nm differ, in spite of the similarities observed within the 190–220 nm interval. The sodium complex of monensin provides a unique spectrum too.

The comparison of CD and SRCD spectra reveals a very good agreement in the range of 192–300 nm. The SRCD technique leads to a higher signal-to-noise ratio, and the SRCD setup with the thin 0.014 mm cell provides characteristic, high intensity signal even within the 178–192 nm region of high absorption.[60,61]

Structural changes upon metal binding may account for the variations of CD spectra. However, these are quite small. The structures of monensin A and its metal complexes were compared; the root-mean-square deviations (RMSD)[69] obtained using the PyMol alignment procedure are collected in Table 1. We see that the overall conformation of the ligand is very similar in each complex. Despite this similarity, observed also in the IR and NMR spectra of monovalent monensinates,[32,33,35,47] the CD spectra differ significantly in band positions, signs, and intensities. Nevertheless, some trends can be observed. The fine structure of MonH differs from those of the metal complexes (RMSD ~0.7; Fig. S5), which may be reflected in the CD intensities (Fig. 3). Both the MonLi and MonNa, and MonK and MonRb crystal pairs look rather similar (RMSD ~0.2; Fig. S6), in accordance with the good agreement in the wavelengths of CD minima and maxima.

Considering that there is no significant change in the overall structures of the species when dissolved in methanol, which can also be supposed from NMR data published earlier,[45–47] at least some differences in the CD spectra might be explained by the small conformational changes of the ligand molecule upon complexation. On the other hand, the crystal structure of MonAg, which is also different from MonH and is rather similar to MonM structures (e.g., RMSD ~0.1 for the comparison with MonNa, see Fig S7), has a rather unique CD spectrum.

Calculated absorption and CD spectra (Fig. 2) may provide a better understanding of the problem, although they do not reproduce the experiment quantitatively. The absorption spectra exhibit a limited number of features to be compared, except for the outstanding high-wavelength absorption of the Ag-derivative. The fully optimized geometries (middle panel) provide better absorption profile of the K-derivative, otherwise they do not exhibit a clear advantage against the X-ray model (bottom).

Occasional agreement of calculated and experimental CD spectra can be observed. For example, both spectra of MonH are negative around 180 nm. The measured curve is positive above ~185 nm, but the calculation predicts a negative signal at higher wavelengths that is not observed experimentally. Part of the disagreement can be given by the position of hydrogen atoms, not clearly given by the X-ray data. The usual error of the TDDFT method[70] and complicated conformational and hydration equilibria that could not be included in the calculations due to excessive computational demands also hinder a more detailed comparison of calculated and experimental CD intensities.

On the other hand, the calculations well document the sensitivity of monensin to the metal binding. Also, based on the orbital analysis, we could assign the most prominent spectral features. Thus, most transitions around 180 nm in monovalent monensinates (except the Ag^+-derivative) are $\sigma\rightarrow\sigma^*$; transitions within 190–200 nm can be approximately thought of as $\sigma\rightarrow\sigma^*$, and within 200 to 206 nm $n\rightarrow\sigma^*$ transitions dominate ("n" means a nonbonding [lone pair] orbital on oxygen in hydroxyl or carboxyl). Most but not all transitions above 180 nm are located around the carboxyl residue. The highest wavelength bands around 210–215 nm are attributed to $\sigma\rightarrow\pi^*$ and $n\rightarrow\pi^*$ transitions, where the π-orbitals mostly belong to the carboxyl, although a small participation of the lone pairs on hydroxyl oxygen can be also counted as π. With silver cation, the situation is different, as the silver transitions are stronger and dominate, all the 180, 215, 230, and 242 nm intense bands are assigned to $4d\rightarrow5s$ transition; the n, π, and σ orbitals of monensin also contribute, as above, but their intensities are weaker.

The spectra generated with the partially optimized structures (lower part of Fig. 2) provide very similar results in

TABLE 1. Root-mean-square deviations (RMSD in Å units) as a measure of similarity between 47 nonhydrogen backbone atoms of the superimposed monensin A complexes

Comparison by LigAlign[69]	Optimized structures	MonH	MonLi	MonNa	MonK	MonRb
MonH	0.188					
MonLi	0.116	0.715 (0.650)				
MonNa	0.152	0.685 (0.555)	0.187 (0.198)			
MonK	0.438	0.648 (0.472)	0.558 (0.611)	0.404 (0.521)		
MonRb	0.477	0.674 (0.551)	0.658 (0.798)	0.503 (0.646)	0.131 (0.660)	
MonAg	0.460	0.649 (0.283)	0.313 (0.597)	0.134 (0.490)	0.231 (0.567)	0.348 (0.597)

Crystal structures obtained with various monovalent ions were compared among each other, as well as, to their optimized structures (first column). The numbers in parentheses relate to the optimized structures.

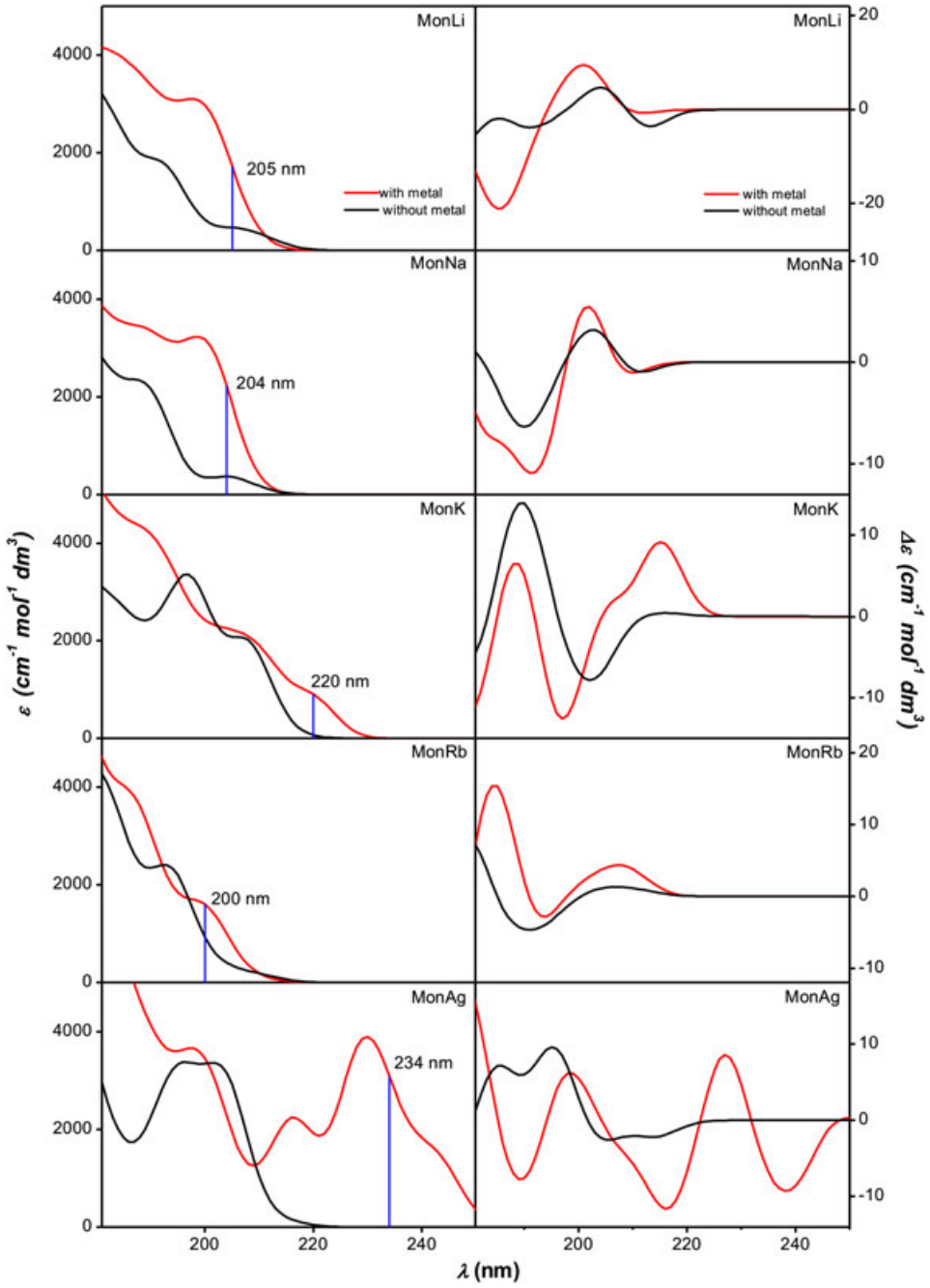


Fig. 3. Calculated absorption (left) and CD (right) spectra of monensin complexes for optimized structures including the metal ion (red) and for the same structures with metal ion removed (black). Transitions with a significant participation of the HOMO–LUMO excitation are indicated for absorption by the vertical lines

terms of agreement/disagreement with the experiment, which also suggest that the overall ligand geometries are rather similar. However, it should be mentioned that, e.g., for MonK and MonRb, the spectra calculated for the partially optimized ("crystal") geometry compare better to experimental CD than those obtained for the fully optimized structure.

Thus, although the calculations do not reproduce well detailed experimental CD patterns, they confirm that the metal ions can induce specific CD shapes under a minimal change of conformation. The limited accuracy can be explained by the complexity of the system and accumulation of computational error stemming from the DFT and TDDFT approximations, approximate solvent model, and lack of dynamics in the modeling. Yet several trends could be observed, such as the profound difference in the behavior of the Ag^+-ion if compared to the others.

To understand better the link between the spectrum and the structure, we performed various computational experiments. In Figure 3, simulated absorption and CD spectra for the Li^+, Na^+, K^+, Rb^+, and Ag^+ complexes, and their counterparts with the same ligand structure but metal ions removed, are plotted. The spectral shapes and the pairwise comparison demonstrate that the metal ion can significantly affect the spectrum participating in electronic transitions and induce changes in the ligand fine structure. We marked the position of the highest-wavelength (lowest-energy) electronic transition for different metal complexes.

As expected, this "threshold" transition largely involves the HOMO and LUMO orbitals. Both the position and relative intensity of these bands vary for different metal ions, with

Fig. 4. HOMO (left) and LUMO (right) molecular orbitals of MonH and its monovalent metal complexes calculated for the optimized structures.

TABLE 2. Distances (in Å) of monensin A oxygen atoms from the metal ion, and the distances between the oxygens at the opposite ends of the molecule obtained from Pymol[a]

Donor atoms	MonH	MonLi	MonNa	MonK	MonRb	MonAg
O1	—	*4.11 (4.22)*	*4.11 (3.99)*	*3.46* **(2.86)**	*3.29 (3.18)*	*3.86 (3.91)*
O2	—	*4.30 (4.47)*	*4.08 (4.18)*	*3.46* **(2.89)**	*3.34 (4.04)*	*3.84 (4.35)*
O3	—	*7.25 (7.36)*	*7.22 (7.22)*	*6.77 (6.47)*	*6.63 (6.82)*	*7.05 (7.30)*
O4	—	*4.03 (4.09)*	*4.16 (4.17)*	*4.22 (4.31)*	*4.22 (4.26)*	*4.15 (4.33)*
O5	—	2.02 (2.06)	2.34 (2.34)	2.65 (2.69)	2.79 (2.98)	2.43 (2.46)
O6	—	2.17 (2.15)	2.33 (2.32)	2.66 (2.81)	2.80 (2.97)	2.41 (2.52)
O7	—	2.63 (2.32)	2.60 (2.49)	2.78 (2.83)	*2.87 (3.05)*	2.69 (2.60)
O8	—	2.32 (2.22)	2.45 (2.44)	2.72 (2.89)	2.82 (2.93)	2.58 (2.62)
O9	—	2.20 (2.24)	2.41 (2.41)	2.80 (2.92)	2.91 (*3.18*)	2.56 (2.75)
O10	—	*3.23 (3.25)*	*3.10 (3.03)*	*3.01 (3.27)*	*3.14 (3.46)*	*3.04 (3.46)*
O11	—	2.11 (2.16)	2.38 (2.40)	2.78 (2.87)	2.92 (2.96)	2.46 (2.37)
O10-O1,2	*3.77, 5.32 (3.32, 5.17)*	2.62, *3.93* (2.54, *3.85*)	2.68, *3.94* (2.55, *3.83*)	2.62, *3.84* (2.60, *4.31*)	2.65, *3.89* (2.62, *3.73*)	2.65, *3.87* (***3.73, 5.56***)
O11-O1,2	2.65, *3.34* (2.67, *3.29*)	2.58, *3.34* (2.61, *3.38*)	2.55, *3.33* (2.60, *3.40*)	2.51, *3.26* (***3.65***, *3.76*)	2.54, *3.41* (2.66, *3.51*)	2.51, *3.30* (2.57, *3.34*)

The numbers in parentheses relate to the optimized structures.
[a]The nonbonding distances are in italic, arbitrarily chosen at 3.00 Å. Bold numbers show significant differences between the crystal structures and the optimized structures, mostly yielding new bonding or nonbonding distances. The underlined numbers show the newly established bonding in the optimized structure.

the Ag^+-ion causing the largest shift of the absorption band to longer wavelength, in agreement with the experiment.

On a qualitative level, one can see the influence of the metal ion binding on electronic structure in Figure 4, where the HOMO and LUMO orbitals are exemplified for MonH and its monovalent derivatives. Apparently, in most metal complexes, the orbitals, in particular HOMO, shift closer to the molecular site where the metals are bound. The Ag^+ complex is an exception, with HOMO being relatively far from the metal, but the LUMO shape is unique due to the participation of the Ag 5s orbital. The delocalization of the orbitals lends CD spectra additional sensitivity to fine conformational changes.

The comparison of the optimized structures gives a similar picture as for the crystal structures (Table 1). The metal complexes do not deviate much from MonH and the pairwise similarity of MonLi and MonNa can still be recognized (RMSD ~0.2). The increase of the size of the metal ion, however, seems to increase discrepancies between the X-ray and computed geometries. This can be demonstrated by the RMSD values derived from the comparison of each crystal structure and its optimized counterpart (first column in Table 1). While these numbers for MonH, MonLi, and MonNa are below 0.2, they increase for MonK, MonRb, and also for MonAg. The strict similarity between MonK and MonRb has been lost (RMSD ~0.7), as well.

Table 2 and Figure S8 provide more details of the fine structural changes during the optimization process. By comparison of the probable hydrogen bonding scheme between the two termini of the monensinate ion in different complexes, we can state that the most significant fine changes occurred in the MonK and MonAg structures. In addition, the M-O atomic distances also vary in the MonK complex as the carboxylate oxygens approached the metal ion during the optimization. The original distance of K^+ from both carboxylate oxygens is ~3.5 Å and from the coordinated oxygens is around 2.7 Å, while in the optimized structure the carboxylate oxygen-metal ion distances are ~2.9 Å and the other oxygens are at 2.8–2.9 Å, a little bit further from the metal ion than in the initial structure. In the metal ion bonding network of the MonRb complex there are gradual changes of the distances, but the overall structure and the coordination mode seems to be the same as in the crystal.

The differences might reflect approximations (solvent model, limited basis set, functional) used in the calculations, but such small deviations between the solution and crystal structures may also be realistic. As was pointed out in the reviewing process, the experiments in methanol may also be biased due to residual traces of water (potentially bound to monensin) that are difficult to control. The overall similarity of the crystal and fully optimized DFT structures of the metal complexes of monensin thus contrasts with their different CD spectral patterns. In general, we can thus conclude that while it is difficult to estimate the extent of the geometry change when the crystals are dissolved in methanol, the comparison of calculated and experimental spectral patterns can provide useful indications.

CONCLUSION

In order to explore the unique metal-binding properties of the monensin A antibiotic, we recorded and analyzed CD and SRCD spectra of its complexes with ammonium, light, and heavy monovalent metal ions. The SRCD technique provided a higher signal-to-noise ratio and enabled measurement in a wider wavelength range than CD. Except for the Ag^+-ion, the metal ion binding did not significantly influence the absorption spectrum, whereas significant changes occurred in CD. This behavior was on a qualitative level explained by time-dependent density functional computations of solution geometries and excitation spectra. These confirmed that incorporation of monovalent cations into the antibiotic structure does not significantly change the solid-state conformation, but that the cation directly participates in the electronic transitions, which may be largely responsible for the CD pattern changes. Because of the large and metal-specific spectral variations under the binding, we can thus conclude that the CD spectroscopy can be used as a sensitive indicator of monensin A monovalent cation binding.

ACKNOWLEDGMENTS

Financial support of TÁMOP-4.2.4.A/2, John von Neumann International Scholarship for senior foreign teachers-researchers & CALIPSO Programme (FP7/2007-2013, grant no. 312284) is acknowledged. In the Czech Republic, the work was also supported by the Grant Agency (16-05935S, and 15-09072S), and MetaCentrum computational grants LM2010005 and CZ.1.05/3.2.00/08.0144. The authors thank Eszter Nemeth for valuable discussion on the PyMol alignment procedure.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

LITERATURE CITED

1. Haney ME Jr, Hoeln MM. Monensin, a new biologically active compound. I. Discovery and isolation. Antimicrob Agents Chemother (Bethesda) 1967;7:349–352.
2. Agtarap A, Chamberlin JW. Monensin, a new biologically active compound. IV. Chemistry. Antimicrob Agents Chemother (Bethesda) 1967;7:359–362.
3. Agtarap A, Chamberlin JW, Pinkerton M, Steinrauf LK. The structure of monensin acid, a new biologically active compound. J Am Chem Soc 1967;89:5737–5739.
4. Pressman BC. Biological application of ionophores. Annu Rev Biochem 1976;45:501–529.
5. Stern PH. Ionophores: chemistry, physiology and potential applications to bone biology. Clin Orthop Rel Res 1977;122:273–298.
6. Long PL, Jeffers TK. Studies on the stage of action of ionophorous antibiotics against Eimeria. J Parasitol 1982;68:363–371.
7. Augustine PC, Smith CK, Danforth HD, Ruff MD. Effect of ionophorous anticoccidials on invasion and development of Eimeria: comparison of sensitive and resistant isolates and correlation with drug uptake. Poultry Sci 1987;66:960–965.
8. Koinarski V, Sherkov SN. Effect of anticoccidial preparations in the prevention of coccidiosis in turkeys caused by Eimeria adenoides. Vet Med Nauki 1987;24:81–85.
9. Folz SD, Lee BL, Nowakowski LH, Conder GA. Anticoccidial evaluation of halofuginone, lasalocid, maduramicin, monensin and salinomycin. Vet Parasitol 1988;28:1–9.
10. Augustine PC, Watkins KL, Danforth HD. Effect of monensin on ultrastructure and cellular invasion by the turkey coccidia Eimeria adenoeides and Eimeria meleagrimitis. Poultry Sci 1992;71:970–978.
11. Dutton CJ, Banks BJ, Cooper CB. Polyether ionophores. Nat Prod Rep 1995;12:165–181.
12. Page SW. The role of enteric antibiotics in livestock production. Avcare: Canberra, Australia;2003.

13. Wang Z, Suo X, Xia X, Shen J. Influence of monensin on cation influx and Na^+-K^+- ATPase activity of Eimeria tenella sporozoites in vitro. J Parasitol 2006;92:1092–1096.
14. Kevin DA II, Meujo DAF, Hamann MT. Polyether ionophores: broad-spectrum and promising biologically active molecules for the control of drug resistant bacteria and parasites. Expert Opin Drug Discov 2009;4:109–146.
15. Lutz WK, Winkler FK, Dunitz JD. Crystal structure of the antibiotic monensin: similarities and differences between free acid and metal complex. Helv Chim Acta 1971;54:1103–1108.
16. Anteunis MJO, Rodios NA. Solution conformation of monensin free acid, a typical representative of the polyetherin antibiotics. Bioorg Chem 1978;7:47–55.
17. Pressman BC. Ionophorous antibiotics as models for biological transport. Fed Proc 1968;27:1283–1288.
18. Lutz WK, Wipf HK, Simon W. Alkali cation specificity and carrier properties of antibiotics nigericin and monensin. Helv Chim Acta 1970;53:1741–1746.
19. Pressman BC, Deguzman NT. Biological applications of ionophores — theory and practice. Ann N Y Acad Sci 1975;264:373–386.
20. Westley JW. Polyether antibiotics, naturally occurring acid ionophores I and II. New York: Marcel Dekker; 1982 Vols.
21. Antonenko YN, Yaguzhinsky LS. The ion selectivity of nonelectrogenic ionophores measured on a bilayer lipid membrane: nigericin, monensin, A23187 and lasalocid A. Biochim Biophys Acta 1988;939:125–130.
22. Prabhananda BS, Kombrabail MH. Monensin-mediated transports of H+, Na+, K and Li + ions across vesicular membranes — T-jump studies. Biochim Biophys Acta 1992;1106:171–177.
23. Huczynski A, Janczak J, Lowicki D, Brzezinski B. Monensin A acid complexes as a model of electrogenic transport of sodium cation. Biochim Biophys Acta Biomembr 2012;1818:2108–2119.
24. Antonenko YN, Rokitskaya TI, Huczy ski A. Electrogenic and nonelectrogenic ion fluxes across lipid and mitochondrial membranes mediated by monensin and monensin ethyl ester. Biochim Biophys Acta 2015;1848:995–1004.
25. Cox BG, Van Truon N, Rzeszotarska J, Schneider H. Rates and equilibria of alkali metal and silver ion complex formation with monensin in ethanol. J Am Chem Soc 1984;106:5965–5969.
26. Pinkerton M, Steinrauf LK. Molecular structure of monovalent metal cation complexes of monensin. J Mol Biol 1970;49:533–546.
27. Duax WL, Smith GD, Strong PD. Complexation of metal ions by monensin — crystal and molecular structure of hydrated and anhydrous crystal forms of sodium monensin. J Am Chem Soc 1980;102:6725–6729.
28. Barrans PY, Alleaume M, Jeminet G. Complexe de sodium de l'ionophore monensine B monohydrate. Acta Cryst 1982;B38:1144–1149.
29. Walba DM, Hermsmeier M, Haltiwanger RC, Noordik JH. Crystal structures of monensin B lithium and silver salts. J Org Chem 1986;51:245–247.
30. Pangborn W, Duax WL, Langs D. The hydrated potassium complex of the ionophore monensin A. J Am Chem Soc 1987;109:2163–2165.
31. Paz FAA, Gates PJ, Fowler S, Gallimore A, Harvey B, Lopes NP, Stark CBW, Staunton J, Klinowski J, Spencer JB. Sodium monensin dihydrate. Acta Cryst 2003;E59:m1050–m1052.
32. Huczynski A, Ratajczak-Sitarz M, Katrusiak A, Brzezinski B. Molecular structure of the 1:1 inclusion complex of monensin A sodium salt with acetonitrile. J Mol Struct 2007;832:84–89.
33. Huczynski A, Ratajczak-Sitarz M, Katrusiak A, Brzezinski B. Molecular structure of the 1:1 inclusion complex of monensin A lithium salt with acetonitrile. J Mol Struct 2007;871:92–97.
34. Yildirim SO, McKee V, Khardli FZ, Mimouni M, Hadda TB. Rubidium (I) monensinate dihydrate. Acta Cryst 2008;E64:m154–m155.
35. Huczynski A, Ratajczak-Sitarz M, Katrusiak A, Brzezinski B. Molecular structure of rubidium six-coordinated dihydrate complex with monensin A. J Mol Struct 2008;888:224–229.
36. Hoogerheide JG, Popov AI. Study of monensin complexes with monovalent metal ions in anhydrous methanol solutions. J Sol Chem 1978;7:357–372.
37. Garcia-Rosas J, Schneider H, Cox BG. Silver complexation by the ionophorous antibiotic monensin in nonaqueous solvents. J Phys Chem 1983;87:5467–5472.
38. Cox BG, Van Truong N, Rzeszotarska J, Schneider H. Stability constants of complexes of monensin and lasalocid with alkali-metal and alkaline-earth-metal ions in protic and polar aprotic solvents. J Chem Soc Faraday Trans I 1984;80:3275–3284.
39. Mimouni M, Perrier S, Pointud I, Juillard J. Selectivity of bacterial ionophore monensin for monovalent metal cations. Solvent effects in methanol and biphasic water-organic systems. J Sol Chem 1993;22:769–785.
40. Gertenbach PG, Popov AI. Solution chemistry of monensin and its alkali metal ion complexes. Potentiometric and spectroscopic studies. J Am Chem Soc 1975;97:4738–4744.
41. Hoogerheide JG, Popov AI. A study of metal complexes of a naturally occurring macrocyclic ionophore Monensin. J Sol Chem 1979;8:83–95.
42. Pointud Y, Bernard C, Touzain S, Astier L, Sabatier B, Juillard J. Thermodynamics of complexation of monovalent metal cations by the ionophore monensin free acid in acetonitrile. J Sol Chem 1997;26:479–495.
43. Degani H. Kinetics of monensin complexation with sodium ions by ^{23}Na NMR spectroscopy. Biophys Chem 1977;6:345–349.
44. Briggs RW, Hinton JF. Thallium-205 and proton nuclear magnetic resonance investigation of the complexation of thallium by the ionophores monensin and nigericine. Biochemistry 1978;17:5576–5582.
45. Turner DL. The conformation of the monensin-A-sodium Complex in solution determined from self-consistent NOE distance constraints. J Magn Reson B 1995;108:137–142.
46. Mimouni M, Hebrant M, Dauphin G, Juillard J. Monovalent cation salts of the bacterial ionophore monensin in methanol. Structural aspects from NMR experiments. J Chem Res 1996;S6:278–279.
47. Martinek T, Riddell FG, Wilson C, Weller CT. The conformations of monensin-A metal complexes in solution determined by NMR spectroscopy. J Chem Soc Perkin Trans 2000;2:35–41.
48. Filipek S, Rzeszotarska J, Kalinowski MK. Polarographic study of Tl^+, Li^+, Na^+, K^+, and Cs^+ complexes with monensin anion in dipolar aprotic solvents. Monat Chem 1994;125:801–809.
49. Chamberlin JW, Agtarap A. Observations on the mass spectrometry of monensin and related compounds. Org Mass Spectrom 1970;3:271–285.
50. Havlicek V, Ryska M, Pospisil S. Negative-ion fast atom bombardment tandem mass spectrometry of sodium salts of monensins and related compounds. J Mass Spectrom 1995;30:1089–1094.
51. Lopes NP, Stark CBW, Hong H, Gates PJ, Staunton J. A study of the effect of pH, solvent system, cone potential and the addition of crown ethers on the formation of the monensin protonated parent ion in electrospray mass spectrometry. Analyst 2001;126:1630–1632.
52. Lopes NP, Stark CBW, Hong H, Gates PJ, Staunton J. Fragmentation studies on monensin A and B by accurate-mass electrospray tandem mass spectrometry. Rapid Commun Mass Spectrom 2002;16:414–420.
53. Miao XS, March RE, Metcalfe CD. Fragmentation study of salinomycin and monensin A antibiotics using electrospray quadrupole time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 2004;17:149–154.
54. Lopes NP, Almeida-Paz FA, Gates PJ. Influence of the alkali metal cation on the fragmentation of monensin in ESI-MS/MS. Braz J Pharm Sci 2006;42:363–367.
55. Degani H, Friedman HL. Ion binding by X-537A. Formulas, formation constants, and spectra of complexes. Biochemistry 1974;13:5022–5032.
56. Painter G, Pressman BC. Induced fit as a determinative of ionophore selectivity. Biochem Biophys Res Comm 1979;91:1117–1122.
57. Painter G, Pressman BC. Circular dichroic discrimination of inclusion and peripheral complexes of the ionophore lasalocid A. Biochem Biophys Res Comm 1980;97:1268–1275.
58. Painter GR, Pollack R, Pressman BC. Conformational dynamics of the carboxylic ionophore lasalocid A underlying cation complexation-decomplexation and membrane transport. Biochemistry 1982;21:5613–5620.
59. Shastri BP, Easwaran KRK. Conformations of lasalocid A-lithium complexes in acetonitrile. Int J Biol Macromol 1984;6:219–223.
60. Miles AJ, Hoffmann SV, Tao Y, Janes RW, Wallace BA. Synchrotron radiation circular dichroism (SRCD) spectroscopy: new beamlines and new applications in biology. Spectroscopy 2007;21:245–255.
61. Miles AJ, Janes RW, Brown A, Clarke DT, Sutherland JC, Tao Y, Wallace BA, Hoffmann SV. Light flux density threshold at which protein denaturation is induced by synchrotron radiation circular dichroism beamlines. J Synchrotron Radiat 2008;15:420–422.

62. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ. Gaussian 09, Revision D01. Wallingford CT: Gaussian, Inc.; 2009.

63. Becke AD. Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 1993;98:5648–5652.

64. Klamt A. COSMO and COSMO-RS. In: Schleyer PR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF III, Schreiner PR editors, Encyclopedia of computational chemistry. Chichester, UK: John Wiley & Sons; 1998. p 604–615.

65. Bou P, Keiderling TAJ. Partial optimization of molecular geometry in normal coordinates and use as a tool for simulation of vibrational spectra. Chem Phys 2002;117:4126–4132.

66. Dolg M. Relativistic effective core potentials. In: Schwerdtfeger P, editor. Relativistic electronic structure theory, Part I: Fundamentals. Amsterdam: Elsevier; 2002. p 793–862.

67. Furche F, Ahlrichs R, Wachsmann C, Weber E, Sobanski A, Vögtle F, Grimme S. Circular dichroism of helicenes investigated by time-dependent density functional theory. J Am Chem Soc 2000;122:1717–1724.

68. Malfreyt P, Juillard J, Pascal Y. Structure of the ionophore monensin and of its tetramethylammonium salt — molecular modeling using various computational methods. New J Chem 1997;21:317–328.

69. Heifets A, Lilien R. LigAlign: flexible ligand-based active site alignment and analysis. J Mol Graph Modell 2010;29:93–101.

70. Bruhn T, Schaumlöffel A, Hemberger Y, Bringmann G. SpecDis: Quantifying the comparison of calculated and experimental electronic circular dichroism spectra. Chirality 2013;25:243–249.

9.3 Appendix C – *PCCP* 2016, 18, 23803.

PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2016, **18**, 23803

Chiral sensing of amino acids and proteins chelating with Eu^{III} complexes by Raman optical activity spectroscopy†

Tao Wu,* Jiří Kessler and Petr Bouř

Received 7th June 2016,
Accepted 1st August 2016

DOI: 10.1039/c6cp03968e

www.rsc.org/pccp

Chiroptical spectroscopy of lanthanides sensitively reflects their environment and finds various applications including probing protein structures. However, the measurement is often hampered by instrumental detection limits. In the present study circularly polarized luminescence (CPL) of a europium complex induced by amino acids is monitored by Raman optical activity (ROA) spectroscopy, which enables us to detect weak CPL bands invisible to conventional CPL spectrometers. In detail, the spectroscopic response to the protonation state could be studied, *e.g.* histidine at pH = 2 showed an opposite sign of the strongest CPL band in contrast to that at pH = 7. The spectra were interpreted qualitatively on the basis of the ligand-field theory and related to CPL induced by an external magnetic field. Free energy profiles obtained by molecular dynamic simulations for differently charged alanine and histidine forms are in qualitative agreement with the spectroscopic data. The sensitivity and specificity of the detection promise future applications in probing peptide and protein side chains, chemical imaging and medical diagnosis. This potential is observed for human milk and hen egg-white lysozymes; these proteins have a similar structure, but very different induced CPL spectra.

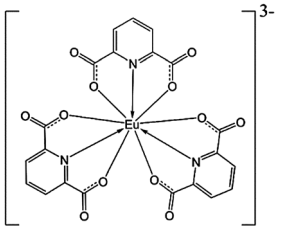
Introduction

Lanthanide complexes are popular probes used in analytical and medicinal chemistry.[1,2] They exhibit unique spectroscopic properties due to their electronic structure, such as the relatively isolated 4f-shell electrons and a wealth of spectroscopically active electronic transitions.[3] Their luminescence is particularly sensitive to the environment of the lanthanide metal; circularly polarized luminescence (CPL), difference in the emission of left- and right-circularly polarized light, provides additional information about molecular chirality, and is thus commonly employed in biologically oriented applications.[4–7]

However, the development of molecular biologically relevant probes based on lanthanides is not straightforward. For example, many europium(III) complexes decompose in the aqueous environment[8] or the luminescence in water solutions is strongly affected by interaction with the O–H groups. The $Na_3[Eu(DPA)_3]$ complex (Scheme 1) used here and similar tris-dipicolinate (DPA) complexes of lanthanides appear to be more universal because they are stable also in water. Such complexes can also be prepared relatively easily and their chiral spectroscopic response sensitively reflects their chemical neighborhood.

The sensitivity comes from the unperturbed geometry of the complex that possesses the D_3 symmetry and forms enantiomers of opposite helicity. They can be imagined as left- and right-propellers and in a standard notation they are referred to as "Δ" and "Λ". In achiral samples they are present in equal amounts. However, the $\Delta \leftrightarrow \Lambda$ equilibrium can be perturbed by interactions with chiral ligands, as revealed by previous CPL experiments.[9–14] In general, such chirality transfer is referred to as the Pfeiffer effect.[15]

So far, CPL experiments on induced chirality in metal complexes used weak excitation sources and were mostly oriented to detect strong CPL bands in crystals exhibiting high CPL to total luminescence ratios.[16] Only recently, solution complexes providing a strong signal were targeted by this



Scheme 1 The $[Eu(DPA)_3]^{3-}$ ion.

Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Flemingovo náměstí 2, 16610 Prague 6, Czech Republic. E-mail: njutwu@gmail.com

† Electronic supplementary information (ESI) available: Raman and ROA spectra of alanine, histidine and arginine, Raman and ROA spectra of the L-His/$[Eu(DPA)_3]^{3-}$ complex at different temperatures, CID dependence on molar ratios of L-Arg/$[Eu(DPA)_3]^{3-}$ at pH 7 and L-His/$[Eu(DPA)_3]^{3-}$ at pH 2, calculated $[Eu(DPA)_3]^{3-}$ electrostatic potential, Raman and degree of circularity spectra of $[Eu(DPA)_3]^{3-}$ chelated with human milk lysozyme, and CID ratios for $^5D_0 \rightarrow {}^7F_1$ transition of the $[Eu(DPA)_3]^{3-}$ complex induced by amino acids. See DOI: 10.1039/c6cp03968e

technique as well.[14,17] In the Raman optical activity (ROA) instrument the luminescence is excited by more intense laser radiation; in addition the circularly polarized components are detected using a sensitive artifact-resistant detection scheme.[18] Thus very weak fluorescence bands can be detected as well. In a recent study, we discussed how the fluorescence can be distinguished from the ''true'' vibrational Raman bands.[19] A disadvantage of the ROA–CPL approach lies in a limited spectral range of the detected fluorescence (approximately 535–610 nm), *i.e.* the transitions need to be close to the laser 532 nm excitation radiation. For europium(III) this is a minor drawback only because the metal has a large number of luminescence bands, including about sixty $^5D \rightarrow {}^7F$ transitions in the measurable region. For the systems investigated in the present study, only the strongest bands would be detectable on traditional CPL instruments.

The ROA technique has been developed to measure tiny differences in scattering of the right- and left-circularly polarized light caused by vibrational transitions in chiral molecules.[20,21] It has been applied to a wide range of systems including small organic compounds, complex proteins, saccharides and nucleic acids.[22–26] The difference is difficult to measure; for a typical molecule the ratio of the ROA and Raman signal (''CID'', circular intensity difference) is about 10^{-4}. Therefore, a strong and easily detectable signal in the presence of lanthanide metal attracted attention for both the plain Raman[27,28] and the ROA[29–31] spectra. As pointed out above, the strongest ''ROA'' bands correspond neither to Raman scattering nor to vibrational transitions, but they originate from circularly polarized luminescence initiated by the laser radiation.[19]

Before, we could detect a record-high CPL activity of a Cs–Eu^{III} complex by ROA.[19] The dissymmetry factor (''*g*'', CPL analogue of the CID) of this complex was 1.4, corresponding to a CID of 1.4/2 = 0.7.[32,33] The aqueous complexes studied here provide much smaller *g*; however, they exhibit remarkable sensitivity to the type and the protonation state of amino acid chain.

At present, *a priori* theoretical tools providing the link between the spectra and the structure are not available. The principle problems lie in the complicated lanthanide electronic structure, so far fairly inaccessible by contemporary quantum chemistry. Nevertheless, we use the empirical crystal field theory[34–36] enabling us to assign the observed transitions, and on a semi-quantitative basis model the induced chirality.

To document that the structural sensitivity remains even when the europium compound reacts with proteins and thus indicate a more wide range of potential applications of this technique, we also report induced CPL spectra of two structurally very close proteins, human milk and hen egg-white lysozymes. They provide opposite spectral patterns; in fact, we are not aware of any other spectroscopy that would provide signals so different as CPL.

Results and discussion

Eu^{III} CPL induced by alanine

CPL spectra of complex $Na_3[Eu(DPA)_3]$ induced by alanine provide rather weak signals. The Raman (luminescence) and ROA (CPL) spectra are plotted in Fig. 1 (top). At pH 7 and 10, the (total) luminescence is dominated by two strong 1864 and 1976 cm^{-1} bands, both corresponding to $^5D_0 \rightarrow {}^7F_1$ transitions of Eu^{3+}.[3] These are accompanied by corresponding circularly polarized luminescence bands. At pH 2, the fluorescence splits into more bands, and a band at 1546 cm^{-1} significantly gains in intensity. The highest CID ratio of the polarized and total luminescence components is about 1×10^{-4} and corresponds to the weak interaction between alanine and the complex (selected CID ratios are listed in Table S1, ESI†). Still, the spectral intensities are much higher than Raman and vibrational ROA of alanine. In the zoomed insets, a few Raman bands may be recognized while ROA is invisible at this scale. True vibrational ROA of alanine is also impossible to measure at such low concentrations. At high concentrations of alanine, observed ROA CID ratios of $\sim 3 \times 10^{-4}$ (Fig. S1, ESI†) are larger than those corresponding to Eu(III) CPL.

Eu^{III} CPL induced by histidine and arginine

As discussed before amino acids with the polar side chains interact much more strongly with $Na_3[Eu(DPA)_3]$ and similar complexes than alanine.[16,37] The luminescence, and in particular CPL, is also very sensitive to pH and the protonation state. One can see (Fig. 1), for example, that the low pH of 2 supports a high luminescence intensity of the 1546 cm^{-1} ($^5D_0 \rightarrow {}^7F_0$) band. Neutral (7) and basic (10) pH provide very similar Raman spectra. In the magnified ''Raman'' spectra the true vibrational Raman scattering signal specific for each amino acid and the complex is apparent as well: dipicolinate centers around 1420 and 1023 cm^{-1}, and Raman bands of histidine and arginine can be recognized, for example, as CH_2 wagging around 1439 cm^{-1}, C–N valence, ring vibration at 999 cm^{-1}, *etc.*[38,39] Vibrational Raman and ROA spectra of pure histidine and arginine can be found in Fig. S2 and S3 (ESI†).

The ROA intensities in Fig. 1 comprise CPL only; vibrational ROA at these low concentrations is invisible as for alanine. But the CID ratios are much higher for the charged amino acids, with a maximum of ~ 0.007 for the protonated (pH = 2) form of histidine (which can be gradually increased to ~ 0.014 with higher concentration of histidine, Fig. S5, ESI†), indicating a much stronger interaction with the europium complex than for alanine. The weakest interaction, at least as judged from the low CID of 4×10^{-5}, is indicated between $[Eu(DPA)_3]^{3-}$ and histidine at pH = 10 (Table S1, ESI†). For the protonated form of histidine and the neutral form of arginine $^5D_1 \rightarrow {}^7F_2$ transition gives rise to a weak but measurable CPL. This signal appears in a form of couplets (close positive and negative band), different from the one-sign pattern of the strongest bands, which suggests a different mechanism of chirality induction, most probably associated with a perturbation of the D_3 symmetry of the Λ and Δ forms of the complex. Above 2400 cm^{-1}, the signal of a $^5D_0 \rightarrow {}^7F_2$ band is partially observable as a shoulder, which is at the limit of the operational instrumental range and the intensity might not be reliable.

The complex chirality control achieved by the environmental pH value appears to be very interesting. Let us arbitrarily

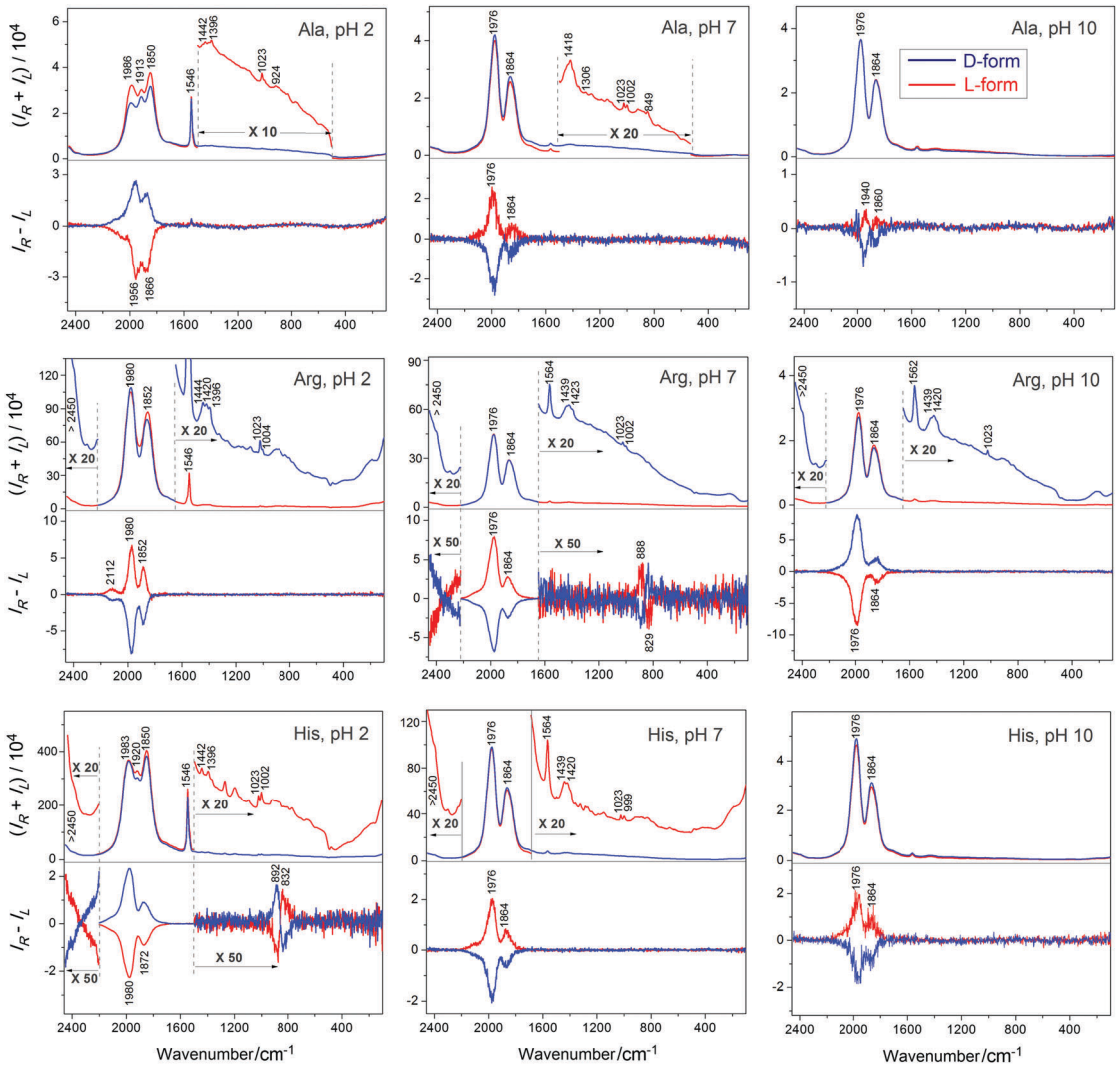


Fig. 1 $[Eu(DPA)_3]^{3-}$ (2 mM) Raman and ROA spectra in the presence of L- and D-alanine (20 mM), L- and D-arginine (12 mM) and L- and D-histidine (24 mM) at pH 2, 7 and 10 (experimental details are listed in the Experimental section).

associate the *Λ*-configuration of the complex with the positive signal at 1970 cm^{-1}. Then for arginine we observe the following possibility of chirality changes by pH:

$$\Lambda\ (\mathrm{pH} = 2) \leftrightarrow \Lambda\ (\mathrm{pH} = 7) \leftrightarrow \Delta\ (\mathrm{pH} = 10)$$

For alanine and histidine, we get:

$$\Delta\ (\mathrm{pH} = 2) \leftrightarrow \Lambda\ (\mathrm{pH} = 7) \leftrightarrow \Lambda\ (\mathrm{pH} = 10)$$

Spectral changes beyond this scheme, such as slight intensity redistribution, band shifts and the occasional appearance of the weaker bands, can be considered as second-order effects.

Eu^{III} CPL induced by tartaric acid and an external magnet

The spectrum induced by tartaric acid (Fig. 2) is similar to that of alanine, but the two strongest luminescence bands are broader and shifted (to 1978 and 1858 cm^{-1}) and a narrow band appears at 1546 cm^{-1} ($^5D_0 \rightarrow {}^7F_0$ transition). CPL bands are broader as well, and the CID of the strongest bands increases to 1.7×10^{-4}.

Raman and magnetic ROA spectra[40] of aqueous solution of plain $Na_3[Eu(DPA)_3]$ (Fig. 3) confirm that the bands observed in the complexes with amino acids originate in europium electronic transitions and facilitate their assignment. Although pure electronic magnetic Raman optical activity has been reported for heavy metal complexes as well,[21,41] control measurements with a 404 nm excitation laser and the degree of circularity confirm[19] that at least the strongest bands are almost exclusively formed by the luminescence and correspond to those observed in the chiral complexes. Only few bands corresponding to the "true" vibrational Raman scattering of the tris(pyridine-2,6-dicarboxylate) ligand can be recognized, such as the CC ring stretch at ~1423 cm^{-1} and symmetric breathing mode of the pyridine ring at ~1023 cm^{-1}.[42,43]

The magnetic ROA spectra (Fig. 3, lower) are again mostly formed by circularly polarized luminescence. Band positions correspond to those in the complexes with amino acids, but the

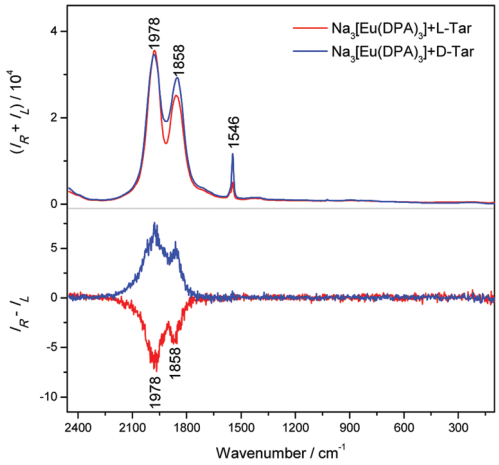


Fig. 2 Raman (upper) and ROA (lower) spectra of complexes with $[Eu(DPA)_3]^{3-}$ (2 mM) with L- and D-forms of tartaric acid (20 mM); the accumulation time was 2 hours.

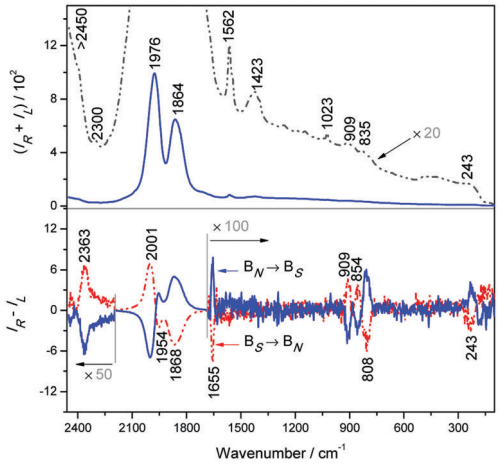


Fig. 3 Raman (upper) and magnetic ROA (lower) of $Na_3[Eu(DPA)_3]$ (2 mM) water solution; the accumulation time was 20 minutes.

sign pattern and relative intensities are different, which reflects the different mechanisms of optical activity induced by the magnetic field and by the complexations with chiral ligands. This is best documented on the strongest signal within 1800–2100 cm^{-1} providing a multiple sign pattern in magnetic ROA/CPL, where the amino acids induced one sign pattern only. In the magnetic field, additional weaker magnetic ROA/CPL bands below 250 cm^{-1} can be observed, assignable to $^5D_1 \rightarrow {}^7F_1$ transitions within the Eu^{3+} f-shell electronic system.[3,13,19,44] The strongest observed transitions of $Na_3[Eu(DPA)_3]$ and its complexes with histidine and arginine are summarized in Table 1. The transition assignment is based on the crystal field theory[3,45] and the electronic structure of free Eu^{3+} ions, which is not much perturbed in the complexes.

Theoretical analysis

The crystal field theory also provides a handy tool to qualitatively model the Eu^{3+} spectral intensities and assign observed transition.

Table 1 Assignment of the observed $Na_3[Eu(DPA)_3]$ transition in the magnetic and induced CPL experiments, the Raman shift (δ) from the 532 nm laser excitation and the corresponding wavelength (λ) of the emitted light

	Magnetic CPL		Induced CPL-His		Induced CPL-Arg	
Transition	δ/cm^{-1}	λ/nm	δ/cm^{-1}	λ/nm	δ/cm^{-1}	λ/nm
$^5D_0 \rightarrow {}^7F_2$	>2450	615	>2450	615	>2450	615
	2362	608				
	2300	606				
$^5D_0 \rightarrow {}^7F_1$	2001	595	1983	595	1976	595
	1976	594	1980	595		
	1954	594	1920	592		
$^5D_0 \rightarrow {}^7F_1$	1868	591	1872	591	1864	591
	1864	591	1850	590		
$^5D_0 \rightarrow {}^7F_0$	1655	583	1546	580	1546	580
$^5D_1 \rightarrow {}^7F_2$	909	550	892	558	888	558
	854	557	832	557	829	557
	835	557				
	808	556				
$^5D_1 \rightarrow {}^7F_1$	243	539				

The calculated luminescence, magnetic and ordinary CPL of the $[Eu(DPA)_3]^{3-}$ ion are plotted in Fig. 4. As expected for an empirical model, the calculated transition energies differ from the observed ones by 200–300 cm^{-1} and relative spectral intensities are not always matching the experiment, otherwise the simulated spectra well correspond to the observation. For example, the highest-wavenumber transition observed at ~2000 cm^{-1} is reproduced at 1740 cm^{-1}, although its relative intensity is significantly underestimated, the lowest one observed at ~240 cm^{-1} is reproduced at ~0 cm^{-1} (= 532 nm), *etc.* As in experiment, CPL induced by the magnetic field is different from that exhibited by an enantiomeric (Δ) form, magnetic CPL above 2300 cm^{-1} is relatively strong compared to (ordinary) CPL, and

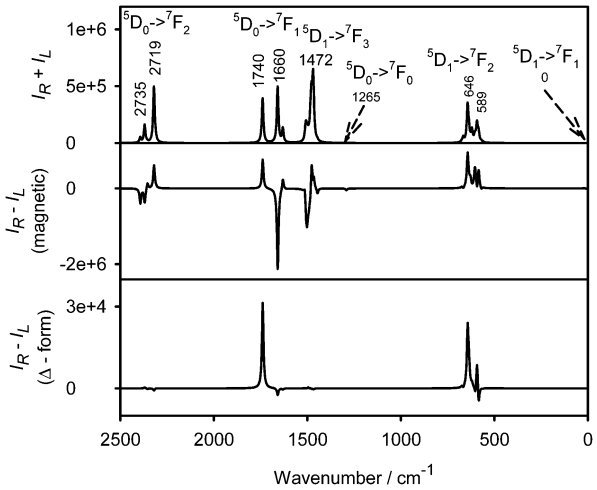


Fig. 4 Raman, magnetic and ROA/CPL Δ-$Na_3[Eu(DPA)_3]$ spectra and state assignment simulated using the crystal field theory. The *y*-scale is arbitrary, but the $(I_R - I_L(\Delta))/(I_R + I_L)$ ratio is meaningful.

the strongest $^5D_0 \rightarrow {}^7F_1$ region exhibits both positive and negative magnetic CPL, while for CPL rather one-sign pattern is simulated. Note that the *y*-scale has no meaning for magnetic CPL because absolute magnitudes of transition matrix elements are not known.

Chiral recognition mechanism

In spite of the large chiroptical effect, the complexes of $[Eu(DPA)_3]^{3-}$ with the amino acids do not appear particularly stable. For example, Raman and ROA spectra of the L-His/$Na_3[Eu(DPA)_3]$ complex measured at different temperatures (Fig. S4, ESI†) reveal that the induced CPL almost vanishes at 90 °C. Also, the dependence of CID on the amino acid/$Na_3[Eu(DPA)_3]$ molar ratio (Fig. S5, ESI†) exhibits a slow convergence and indicates a small stability constant of the resultant "super-complex".

At present, we do not know the accurate structure of the amino acid associated with the $[Eu(DPA)_3]^{3-}$ ion. Clearly, the positively charged histidine and arginine forms are most attracted to the negatively charged complex, *cf.* Fig. S6 (ESI†) with its electrostatic potential. The ion pairing, however, can produce high CID and strong CPL only for an interaction additionally sensitive to $[Eu(DPA)_3]^{3-}$ chirality.[46] This can be provided by hydrogen bonding and electrostatic attractions between the COO^- groups of $[Eu(DPA)_3]^{3-}$ and polar amino acid parts, as well as by π–π interactions and van der Waals attraction.

For selected systems we ran molecular dynamics simulations and obtained free energy profiles using the weighted histogram analysis method (WHAM).[35,36] For alanine (Fig. 5) we can see that the protonated (Ala^+) form is predicted to interact most strongly with the $[Eu(DPA)_3]^{3-}$ ion, with two minima of the energy at approximately 5 and 7 Å. For the Δ-form, the minima have about the same value (−0.8 kcal mol^{-1}), whereas for the Λ-form the distance of 7 Å is more favoured. Note also the energy maximum at ∼9.5 Å, which corresponds to disruption of the hydration spheres of $[Eu(DPA)_3]^{3-}$ and Ala^+ when they are coming closer.

For alanine in the zwitterionic form (Ala_{zw}, middle in Fig. 5) only a very shallow (−0.2 kcal mol^{-1}) minimum appears at ∼7.5 Å for the Λ-enantiomer of $[Eu(DPA)_3]^{3-}$. For Ala_{zw} and the Δ-enantiomer, as well as for the deprotonated amino acid (Ala^-, bottom of the figure), no significant minima of the energy are apparent. This behavior qualitatively well corresponds to the spectroscopic results (Fig. 1), where Ala^+ provided the largest chiral response while the other forms gave much weaker CPL intensities.

Similar behaviour of the free energy was obtained also for histidine (Fig. 6). However, its energy profiles for the Δ and Λ enantiomers are more different than those for alanine, and the energy scale suggests that the interaction with $[Eu(DPA)_3]^{3-}$ is in general stronger. For example, the double-protonated form (His^{2+}) provides only a weak (∼−0.7 kcal mol^{-1}) energy minimum with the Λ-enantiomer, but its associate with the Δ-enantiomer of $[Eu(DPA)_3]^{3-}$ is stabilized by 3.1 kcal mol^{-1} around 8 Å. This is in agreement with both the spectroscopic results described above (Fig. 1) and the different histidine structure providing more binding sites than alanine.

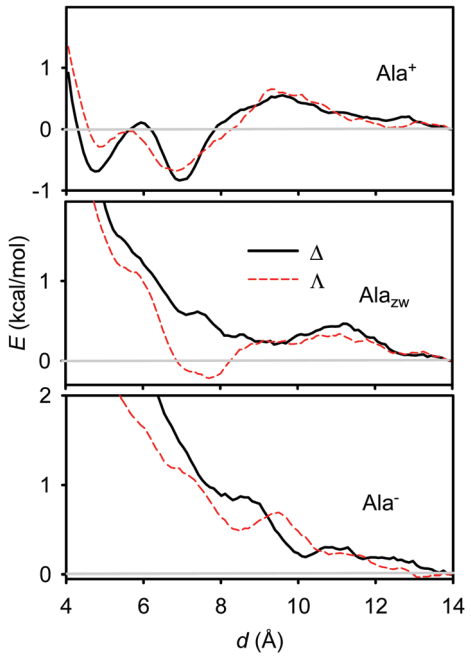


Fig. 5 Calculated free energy profiles for different L-alanine forms interacting with Δ and Λ enantiomers of $[Eu(DPA)_3]^{3-}$.

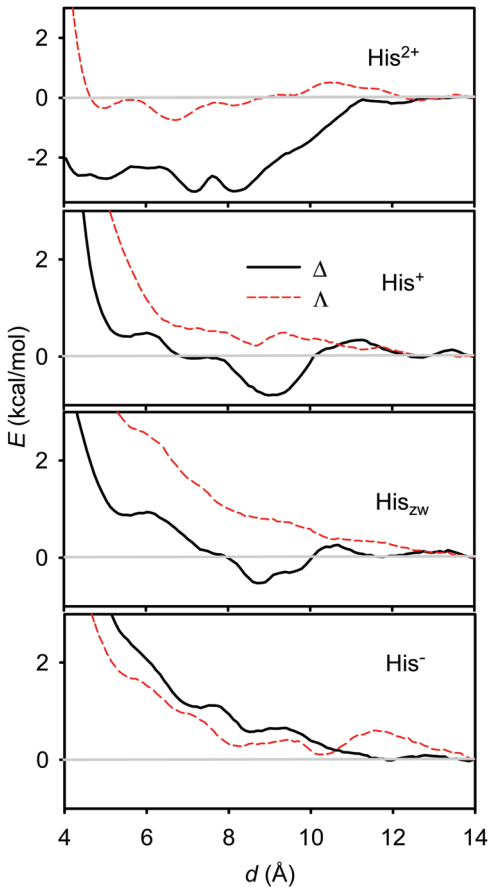


Fig. 6 Calculated free energy profiles for different L-histidine forms interacting with Δ and Λ enantiomers of $[Eu(DPA)_3]^{3-}$.

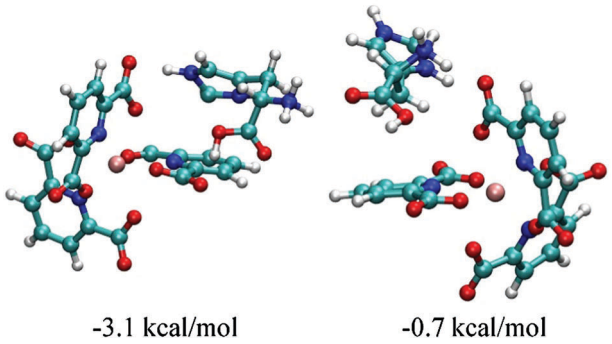


Fig. 7 Geometries and formation free energies of L-histidine complexes with the Δ (left) and Λ (right) form of $[Eu(DPA)_3]^{3-}$.

Geometries of His^{2+} associates with Δ- and Λ-$[Eu(DPA)_3]^{3-}$ obtained by the molecular dynamic simulations are shown in Fig. 7. As also shown in Fig. 6, the formation of the first system is associated with the stabilization energy of 3.1 kcal mol^{-1}, while the second is stabilized by 0.7 kcal only. At this point, however, we wish to warn the reader that although the simulations provide a plausible explanation of the experiment, they may be hampered by limited configurational sampling, restriction to one protonation state, a 1 : 1 complex : amino acid ratio, and force field accuracy. Nevertheless, the results document the possibility of chiral discrimination and the energetics and stereochemistry involved. For the first structure in Fig. 7, for example, a π–π stacking between the histidine five-membered ring and the DPA ligand is possible, while the Δ configuration does not enable it. More realistic energy and geometry estimates are intended for future studies as the perturbation of the $\Delta \leftrightarrow \Lambda$ equilibrium is associated with rather small energy differences, which are especially difficult to reproduce for the weak chirality sensitive interactions.[47] For the $^5D_0 \rightarrow {}^7F_1$ transition in $[Eu(DPA)_3]^{3-}$ a maximal *g*-factor of ~0.1 was reported,[48] which again indicates a partial stereoselection in our experiments only.

$[Eu(DPA)_3]^{3-}$ complexes with human milk and hen egg-white lysozymes

Luminescence and induced CPL spectra of the lysozyme mixtures with $[Eu(DPA)_3]^{3-}$, together with the vibrational Raman and ROA spectra of pure proteins as obtained in ref. 49, are shown in Fig. 8. The two proteins are fairly similar, and the vibrational ROA/Raman spectra (upper part of the figure) differ in minor features only. Also the luminescence spectra are almost the same. However, the $[Eu(DPA)_3]^{3-}$ complex in the presence of the lysozymes provides induced CPL spectra of different chiralities. The strong $^5D_0 \rightarrow {}^7F_1$ 1976 and 1875 cm^{-1} bands of human milk lysozyme provide CID ratios of 1.46×10^{-2} and 7.14×10^{-3}, respectively, and the signal remains detectable at a much lower concentration (0.5 mg ml^{-1}) of the protein than needed for classical vibrational ROA measurements (>100 mg ml^{-1}). The hen egg-white lysozyme inverses the CPL sign of the strongest bands and provides in absolute magnitude bigger CIDs of -4.65×10^{-2} and -2.24×10^{-2}. A similar change occurs for the weaker bisignate $^5D_1 \rightarrow {}^7F_2$ CPL signal at 838 and 887 cm^{-1}.

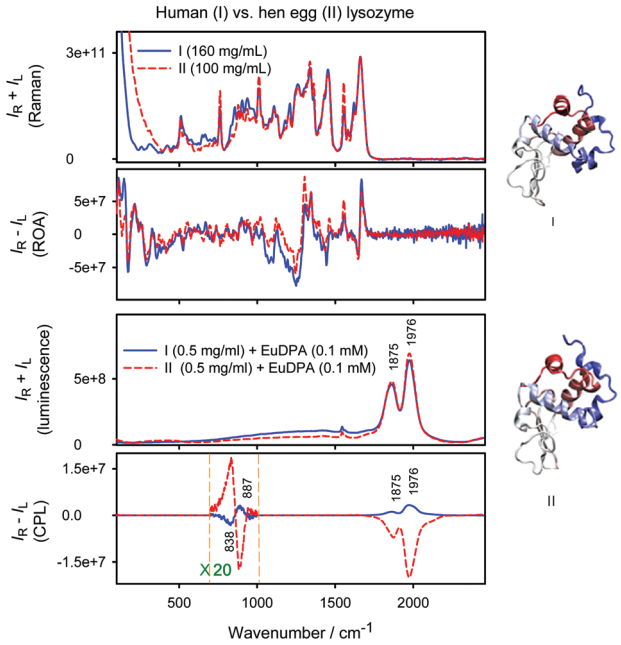


Fig. 8 Normalized Raman and ROA spectra of human milk and hen-egg lysozyme (from ref. 49) and luminescence and CPL spectra of the complexes with $[Eu(DPA)_3]^{3-}$ (this work).

Note that such an induced CPL signal takes full advantage of the ROA laser resource (1000 mW) which is not accessible by traditional CPL spectrometers. The observation is consistent with the sensitive response of the $\Delta \leftrightarrow \Lambda$ equilibrium to the individual amino acid chains, and indicates the potential of the method in biochemical analyses.

Experimental

Synthesis of the Eu^{III} complex

The $Na_3[Eu(DPA)_3]$ (DPA = dipicolinate = 2,6-pyridine-dicarboxylate) complex was obtained by a reaction of europium carbonate and pyridine-2,6-dicarboxylic acid in the 1 : 3 molar ratio in water, and pH was adjusted to 7 by adding 1 M sodium carbonate solution.

ROA and magnetic ROA experiments

Backscattering Raman and scattered circular polarization (SCP) ROA spectra were acquired on a BioTools ROA spectrometer operating with laser excitation at 532 nm at a resolution of 7 cm^{-1}, and the laser power at the sample was 200–1000 mW. Accumulation times were 1 hour for the ROA spectra of histidine–Eu^{III} complexes at pH = 2 and arginine at pH = 7; 4 hours for spectral panels of histidine–Eu^{III} at pH = 7 and Arginine–Eu^{III} at pH = 2, 10; and 20 hours for spectral panels of histidine–Eu^{III} at pH = 10 and alanine–Eu^{III} at pH = 2, 7 and 10, respectively. Magnetic ROA (MROA) spectra[41,50] were acquired under similar conditions, using a magnetic cell providing a field of about 1.5 tesla.[40] The collection time was 20 min for the MROA spectrum of $Na_3[Eu(DPA)_3]$ with a concentration of 2 mM in water. For ROA spectral measurements of pure amino acids,

the collection times were 10 h with a concentration of 0.5 M for L/D-alanine, 0.24 M for L/D-histidine and 0.34 M for L/D-arginine. The pH value was adjusted using 0.1 M HCl or NaOH aqueous solutions. Human milk and hen egg-white lysozymes (Sigma-Aldrich) were dissolved in water. For Raman, ROA and degree of circularity spectral measurements the protein concentration was 0.5 mg ml^{-1} and the EuDPA complex concentration was 0.1 mM; the measurement took 1 hour. All experiments were performed at room temperature (20 °C). Raman spectra are presented after subtraction of the water background signal. Note that Raman spectra also comprise the total luminescence and the ROA spectra comprise circularly polarized luminescence.

Theoretical calculations

$[Eu(DPA)_3]^{3-}$ geometry was optimized by energy minimization using the Gaussian 09[51] program, the B3LYP functional, the MWB28 pseudopotential and basis set for Eu, and the 6-31G(d,p) basis set for the other atoms. The solvent was modelled by a conductor-like polarizable continuum solvent model (CPCM).[52]

To assign spectral peaks and approximately simulate spectral intensities, the semi-empirical crystal theory was used.[3,53] The crystal (ligand) potential in $Na_3[Eu(DPA)_3]$ was calculated from ligand electronic density obtained using Gaussian at the B3LYP/6-31G(d,p)/PCM approximation level. The europium wavefunctions and energies were obtained using the Lanthanide program.[34] Then spectral intensities were generated using our AnalCFP program interfaced to Lanthanide, using the perturbational approach[54–56] where the Eu^{3+} free-ion wavefunction was perturbed by the same DPA ligand electrostatic potential. For example, the electric dipole moment matrix element for a transition between states A and B ($\alpha = x$, y, or z) was calculated as

$$\langle A'|\mu_\alpha|B'\rangle = \sum_{q,q'} C_q^{A*} C_{q'}^{B} \langle l^n q|\mu_\alpha|l^n q'\rangle_{\text{eff}}$$

where C_q^X are the expansion coefficients to individual electronic configurations $l^n q'$ ($n = 6$ electrons in l = f-shell). Effective dipole moments among the configurations were obtained as

$$\langle l^n q|\boldsymbol{\mu}|l^n q'\rangle_{\text{eff}} = K\left(\sum_u \langle l^n q|\boldsymbol{\mu}|u\rangle\langle u|V|l^n q'\rangle + \sum_u \langle l^n q|V^*|u\rangle\langle u|\boldsymbol{\mu}|l^n q'\rangle\right)$$

where K is a constant, V is the ligand electrostatic potential, and u denote configurations where one electron is transferred to the g or d-shell. Note that the plain dipole moments within the f-shell configurations $\langle f^6 q|\mu_\alpha|f^6 q'\rangle$ are zero; thus the spectrum is very sensitive to the lanthanide environment entering through the potential V. For magnetic CPL, the sum over states formula was used as developed previously for magnetic circular dichroism.[57] The potential was calculated from nuclear charges and electron density obtained by the Gaussian calculation.

Interaction of L-histidine with the Λ and Δ $[Eu(DPA)_3]^{3-}$ forms was also studied using molecular dynamics and the Amber14[58] software. Variously protonated histidine and alanine amino acids with the $[Eu(DPA)_3]^{3-}$ ion were placed inside a cubic box (30 Å a side) filled with water molecules. For a large distance of the two components, the energy of the system was minimized, and the geometry was equilibrated during a 1 ns *NVT* dynamics, using an integration step of 1 fs, temperature of 300 K, and the GAFF[59] (DPA ligands), Amber14SB[60] (alanine and histidine) and TIP3P[61] (water) force fields. Europium parameters were taken from ref. 62. Then the distance between carbonyl oxygen of histidine/alanine and europium was decreasing (14–4 Å) in 1 Å steps using 1 ns long constrained dynamics runs and a harmonic potential force constant of 2 kcal $Å^{-2}$ mol^{-1}. From resultant histograms the potential of mean force was obtained *via* the weighted histogram analysis method (WHAM).[35,36]

Conclusions

We synthesized the water-soluble $Na_3[Eu(DPA)_3]$ complex and explored its chirality induced by model amino acids. Due to the sensitive ROA instrumentation very weak CPL bands could be detected, invisible to conventional CPL experiments. The results confirm a tight link between measured CPL spectra patterns and the structure; in particular, the circularly polarized component was found to be extremely sensitive to the nature of the interacting ligand. Magnetic ROA spectra were measured as well, comprising mostly magnetically induced CPL, which enabled us to better discriminate and assign observed transitions. On a semi-quantitative basis, the relation of spectral intensities to fine structural changes could be modelled by the crystal field theory, which also confirmed the assignment of the transitions. For various alanine and histidine forms the molecular dynamics modelling confirmed the possibility of chiral discrimination, *e.g.* it indicated that the π–π stacking involving histidine side chain and electrostatic interactions may be equally important for the complex stabilization. Significant differences in induced CPL in the presence of human milk and hen egg-white lysozymes suggest that the sensitivity of the $[Eu(DPA)_3]^{3-}$ complex may facilitate its applications for sensitive protein detection in analytical chemistry and chemical imaging.

Acknowledgements

This work was supported by the Czech Science Foundation (16-08764Y, 16-05935 and S15-09072S).

Notes and references

1 M. C. Heffern, L. M. Matosziuk and T. J. Meade, *Chem. Rev.*, 2014, **114**, 4496–4539.
2 J. C. G. Bunzli and S. V. Eliseeva, *Chem. Sci.*, 2013, **4**, 1939–1949.
3 K. Binnemans, *Coord. Chem. Rev.*, 2015, **295**, 1–45.
4 F. Zinna and L. Di Bari, *Chirality*, 2015, **27**, 1–13.
5 L. Benda, J. Štěpánek, J. Kaminský and P. Bouř, in *Comprehensive Chirality*, ed. E. M. Carreira and H. Yamamoto, Elsevier, Amsterdam, 2012, vol. 8, pp. 520–544.
6 L. Benda, P. Bouř, N. Müller and V. Sychrovský, *J. Phys. Chem. B*, 2009, **113**, 5273–5281.

7 E. M. Sanchez-Carnerero, A. R. Agarrabeitia, F. Moreno, B. L. Maroto, G. Muller, M. J. Ortiz and S. de la Moya, *Chem. – Eur. J.*, 2015, **21**, 13488–13500.
8 V. Andrushchenko, D. Padula, E. Zhivotova, S. Yamamoto and P. Bouř, *Chirality*, 2014, **26**, 655–662.
9 H. G. Brittain, *Inorg. Chem.*, 1981, **20**, 3007–3013.
10 H. G. Brittain, *J. Coord. Chem.*, 1989, **20**, 331–347.
11 E. Huskowska and J. P. Riehl, *Inorg. Chem.*, 1995, **34**, 5615–5621.
12 G. Muller and J. P. Riehl, *J. Fluoresc.*, 2005, **15**, 553–558.
13 P. Gawryszewska, J. Legendziewicz, Z. Ciunik, N. Esfandiari, G. Muller, C. Piguet, M. Cantuel and J. P. Riehl, *Chirality*, 2006, **18**, 406–412.
14 A. Moussa, C. Pham, S. Bommireddy and G. Muller, *Chirality*, 2009, 497–506.
15 S. Kirschner, N. Ahmad, C. Munir and R. J. Pollock, *Pure Appl. Chem.*, 1979, **51**, 913–923.
16 J. P. Riehl and G. Muller, in *Comprehensive chiroptical spectroscopy, volume 1: Instrumentation, methodologies, and theoretical simulations*, ed. N. Berova, P. L. Polavarapu, K. Nakanishi and R. W. Woody, John Wiley & Sons, New Jersey, 2012, vol. 1, pp. 65–90.
17 B. T. Nguyen, A. J. Ingram and G. Muller, *Chirality*, 2016, **28**, 325–331.
18 W. Hug, *Appl. Spectrosc.*, 2003, **57**, 1–13.
19 T. Wu, J. Kapitán, V. Mašek and P. Bouř, *Angew. Chem., Int. Ed.*, 2015, **54**, 14933–14936.
20 L. Nafie, *Vibrational optical activity: Principles and applications*, Wiley, Chichester, 2011.
21 L. D. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, 2004.
22 J. Haesler, I. Schindelholz, E. Riguet, C. G. Bochet and W. Hug, *Nature*, 2007, **446**, 526–529.
23 S. Luber and M. Reiher, *ChemPhysChem*, 2010, **11**, 1876–1887.
24 E. W. Blanch, L. Hecht, C. D. Syme, V. Volpetti, G. P. Lomonossoff, K. Nielsen and L. D. Barron, *J. Gen. Virol.*, 2002, **83**, 2593–2600.
25 E. W. Blanch, I. H. McColl, L. Hecht, K. Nielsen and L. D. Barron, *Vib. Spectrosc.*, 2004, **35**, 87–92.
26 Z. Q. Wen, L. D. Barron and L. Hecht, *J. Am. Chem. Soc.*, 1993, **115**, 285–292.
27 M. Vargek, T. B. Freedman, E. Lee and L. A. Nafie, *Chem. Phys. Lett.*, 1998, **287**, 359–364.
28 J. A. Koningstein, *Russ. Chem. Rev.*, 1973, **42**, 834–850.
29 S. Yamamoto and P. Bouř, *Angew. Chem., Int. Ed.*, 2012, **51**, 11058–11061.
30 C. Merten, H. Li, X. Lu, A. Hartwiga and L. A. Nafie, *J. Raman Spectrosc.*, 2010, **41**, 1563–1565.
31 T. Wu, J. Hudecová, X. Z. You, M. Urbanová and P. Bouř, *Chem. – Eur. J.*, 2015, **21**, 5807–5813.
32 J. L. Lunkley, D. Shirotani, K. Yamanari, S. Kaizaki and G. Muller, *J. Am. Chem. Soc.*, 2008, **130**, 13814–13815.
33 J. L. Lunkley, D. Shirotani, K. Yamanari, S. Kaizaki and G. Muller, *Inorg. Chem.*, 2011, **50**, 12724–12732.
34 S. Edvardsson and D. Åberg, *Comput. Phys. Commun.*, 2001, **133**, 396–406.
35 S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
36 B. Roux, *Comput. Phys. Commun.*, 1995, **91**, 275–282.
37 K. Okutani, K. Nozaki and M. Iwamura, *Inorg. Chem.*, 2014, **53**, 5527–5537.
38 E. Deplazes, W. van Bronswijk, F. Zhu, L. D. Barron, S. Ma, L. A. Nafie and K. J. Jalkanen, *Theor. Chem. Acc.*, 2008, **119**, 155–176.
39 X. Li, K. H. Hopmann, J. Hudecová, W. Stensen, J. Novotná, M. Urbanová, J. S. Svendsen, P. Bouř and K. Ruud, *J. Phys. Chem. A*, 2012, **116**, 2554–2563.
40 J. Šebestik and P. Bouř, *Angew. Chem., Int. Ed.*, 2014, **53**, 9236–9239.
41 L. D. Barron and J. Vrbancich, *J. Raman Spectrosc.*, 1983, **14**, 118–125.
42 A. A. Kolomenskii and H. A. Schuessler, *Spectrochim. Acta, Part A*, 2005, **61**, 647–651.
43 L. B. Kong, P. Setlow and Y. Q. Li, *Analyst*, 2012, **137**, 3683–3689.
44 D. R. Foster and F. S. Richardson, *Inorg. Chem.*, 1983, **22**, 3996–4002.
45 J. R. G. Thorne, M. Jones, C. S. McCaw, K. M. Murdoch, R. G. Denning and N. M. Khaidukov, *J. Phys.: Condens. Matter*, 1999, **11**, 7851–7866.
46 A. Berthod, *Anal. Chem.*, 2006, **78**, 2093–2099.
47 J. Kessler, M. Dračínský and P. Bouř, *J. Phys. Chem. A*, 2015, **119**, 5260–5268.
48 P. Gawryszenska, J. Legendziewicz, Z. Ciunik, N. Esfandiari, G. Muller, C. Piguuet, M. Cantuel and J. P. Riehl, *Chirality*, 2006, **18**, 406–412.
49 J. Kessler, J. Kapitán and P. Bouř, *J. Phys. Chem. Lett.*, 2015, **6**, 3314–3319.
50 L. D. Barron and J. Vrbancich, *J. Raman Spectrosc.*, 1984, **15**, 47–50.
51 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, Jr., 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, *Revision D.01.*, 2013.
52 A. Klamt, *J. Phys. Chem.*, 1995, **99**, 2224–2235.
53 W. T. Carnall, G. L. Goodman, K. Rajnak and R. S. Rana, *J. Chem. Phys.*, 1989, **90**, 3443–3457.
54 G. S. Ofelt, *J. Chem. Phys.*, 1962, **37**, 511–520.
55 B. R. Judd, *Phys. Rev.*, 1962, **127**, 750–761.
56 F. S. Richardson and T. R. Faulkner, *J. Chem. Phys.*, 1982, **76**, 1595–1606.

57 P. Štěpánek and P. Bouř, *J. Comput. Chem.*, 2015, **36**, 723–730.
58 D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. M. Ferguson, G. Seibel and P. A. Kollman, *Comput. Phys. Commun.*, 1995, **91**, 1–41.
59 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2005, **25**, 1157–1174.
60 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950–1958.
61 W. L. Jorgensen, J. Chandrasekhar and J. D. Madura, *J. Chem. Phys.*, 1983, **79**, 926–935.
62 P. Li, L. F. Song and K. M. Merz, *J. Phys. Chem. B*, 2015, **119**, 883–895.

9.4 Appendix D – *J. Phys. Chem. B* 2013, 118, 6937.

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Arrangement of Fibril Side Chains Studied by Molecular Dynamics and Simulated Infrared and Vibrational Circular Dichroism Spectra

Jiří Kessler,[†,‡] Timothy A. Keiderling,[*,§] and Petr Bouř[*,†]

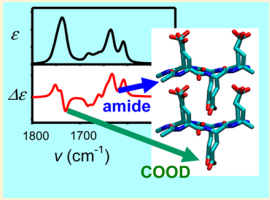
[†]Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Flemingovo náměstí 2, 166 10 Prague, Czech Republic

[‡]Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University, Hlavova 8, 128 40 Prague, Czech Republic

[§]Department of Chemistry, University of Illinois at Chicago, 845 West Taylor Street, Chicago, Illinois 60607-7061, United States

Supporting Information

ABSTRACT: Highly ordered assemblies of β-sheet-forming peptide and protein fibrils have been the focus of much attention because of their multiple and partially unknown biological functions, in particular as related to degenerative neuronal disorders. Recently, vibrational circular dichroism (VCD) spectra have been shown to provide a unique means of detection for such extended structures utilizing modes of the peptide main chain backbone. In the case of poly-glutamic acid, surprising VCD responses were also found for side chain modes. In this study, in an attempt to explain this latter observation and obtain a link between fibrillar structure and its optical spectral properties, molecular dynamics (MD) methods are used to model the geometry and dynamics of assemblies containing repeating β-strands of Glu_n. A crystal-like model was adopted for the MD structure simulations. Infrared and VCD spectra for segments of MD modeled fibrillar geometries were first calculated using density functional theory (DFT), and then, those parameters were applied to larger structures by means of Cartesian coordinate transfer (CCT) of atomic tensors from the segments. The computations suggest the side chains exhibit residual conformational constraints, resulting in local coupling giving rise to non-negligible VCD intensity, albeit with an overall broad distribution. Calculated spectral distributions are qualitatively consistent with the experimental results but do differ in magnitude. The possibility of realistic modeling of vibrational spectra significantly broadens the potential for application of optical spectroscopies in structural studies of these aggregated biopolymers.



INTRODUCTION

Structure and interactions of fibrils and similar protein aggregates are intensively studied because of the appearance of that morphology for many proteins and peptides associated with the symptoms of several neurodegenerative diseases (Alzheimer's, Parkinson's, Huntington's, etc.).[1] More fundamentally, such structures are of interest due to the general need to understand alternate aspects of peptide and protein folding (or mis-folding) and the associated molecular interactions that drive such structure formation.[2] However, insight into the structure and other physical properties of fibrils remains rather limited. For example, structural irregularities in solid phase samples can make conventional X-ray studies and thus spatial resolution at the atomic level impossible.[3] At the same time, difficulties in solubilizing such fibrils to a uniform distribution can cause difficulties for some spectroscopic studies, preventing reliable analyses.[4−8]

In spite of these problems, electronic[9] and vibrational spectra,[10] in particular coupled with circular dichroism (CD), have revealed unique properties of fibrillar structures. Traditional tests of fibril formation have involved fluorescence changes of adsorbed dyes, such as thioflavin T, and detection of induced CD in the ultraviolet (UV). More recently, vibrational circular dichroism (VCD, the differential absorption of left- and right-circularly polarized light in the infrared, IR) has been shown to detect the sense of supermolecular fibrillar twist and some degree of fibrillar assembly (and even morphology, e.g., thin ribbons vs helical aggregates), as confirmed by complementary atomic force microscopy (AFM) studies.[10−18] Such vibrational spectra can be computationally modeled, even for relatively large structures, permitting development of structure−spectra correlations on a fundamental level.[19−23]

In principle, the structural basis of such spectral responses can be obtained via quantum chemical simulations. However, large size molecules and molecular aggregates pose a challenge due to the nonlinear scaling of more reliable calculations that use, for example, density functional theory (DFT) methods. If the structures are regular, it is possible to compute spectral properties of a limited number of fragments and transfer those onto a larger structure. We have shown this method to work well for computations of vibrational spectra, particularly IR and VCD by use of a Cartesian coordinate transfer method (CCT).[24−26] The complexity of an aggregated peptide structure makes complete understanding and simulations of all spectral features rather difficult. The resultant spectral bandshapes and frequency distributions are dependent on both local conformation of the peptide chains (secondary structure)[27] and longer-range coupling through periodic

Received: March 3, 2014
Revised: May 22, 2014
Published: May 28, 2014

   dx.doi.org/10.1021/jp502178d | J. Phys. Chem. B 2014, 118, 6937−6945

structures enabling vibrational mode delocalization.[28] With formation of supermolecular structure, these polymeric interactions gain another dimension, becoming coherent (ordered) in a relatively large particle, which can alter the intensities of various underlying spectral components. Vibrational spectra, including IR and VCD intensities, can be reliably calculated with relatively efficient DFT computations for peptides of moderate size.[29,30] These computations can be extended to larger molecular assemblies with the CCT method, provided reasonable structures can be determined.[24−26] In this way, spectra of structures containing extended single or multiple sheets (each with many strands) can be simulated and related to experimental data.[31−33]

Poly-L-glutamic acid at low pH has low solubility and upon heating forms an aggregate with an underlying β-sheet secondary structure that has unusual characteristics, which are evidenced as a higher density structure correlated to a decrease in intersheet separation and an IR spectrum having an amide I (primarily C═O stretch) band with a maximum below 1600 cm^{-1}, which is much lower than normal for β-sheets. This has been termed a β_2 structure, whose unique low-frequency amide I character (<1600 cm^{-1}) and prominent (~1730 cm^{-1}) —COOD stretching band has been attributed to bifurcated H-bonding of the amide C═O both intrasheet, i.e., cross-strand to other amide N—H groups, and intersheet to side-chain carboxylic acid, —COOH, groups associated with a neighboring (stacked) sheet.[4,34,35] The VCD spectrum correlates to this IR pattern with an intense couplet shape centered on the most intense, lower frequency amide I IR component and a weaker couplet (or sometimes three-band pattern) associated with the —COOH side chain band.[4,36]

Simulating spectra for such poly-glutamic acid[4] structures poses new challenges. While spectra of both the side chain and backbone indicate ordering, the degree of order is probably not the same. Normally, side-chain modes result in zero or very weak VCD, in that the VCD spectrum of an otherwise achiral side-chain chromophore (e.g., −COOH) must arise primarily from through-space coupling or a systematic geometry perturbation coming from at least partially ordered repeating structures. Because of the significant conformational freedom of such side chains, modeling the fibril structure based on a complete conformer search is not reasonable.

Therefore, as a first approximation to the fibrillar structure, we used molecular dynamics (MD) techniques to simulate the dense fibril-like structure with highly ordered strands, and explored initial effects of backbone and side chain flexibility using pseudocrystal-like periodic boundary conditions. Once a stable assembly was established, snapshots of structures along the trajectory were used as a basis for DFT computations of IR and VCD spectral parameters for fragments of the sheet structures that contain several interacting amide and carboxylic acid groups. The CCT method was then adopted to transfer these DFT spectral parameters from fragment computations to effectively model spectra of quite large segments of fibril structures. The variations in structures obtained were then encompassed in our spectral simulations by obtaining an average for a sufficient number of structures derived from MD trajectory snapshots. The results show that, even for largely disordered fibril side chains, a net −COOH (or −COOD) VCD can develop, qualitatively reflecting some underlying order in the side chains leading to their coupling. While these results do not fit all the experimental details, they do suggest an origin for observed Glu_n IR and VCD results and demonstrate that a combined quantum chemical/MD simulation can provide a link between formation of supermolecular structure and its spectroscopic response.

METHODS

Molecular Dynamics. For two 15-amide protonated poly-glutamic acid (PLGA) strands, $[Ac\text{-}Glu_{14}\text{-}NH_2]_2$, with acetyl on the N-terminus and $-NH_2$ on the C-terminus as in Figure 1,

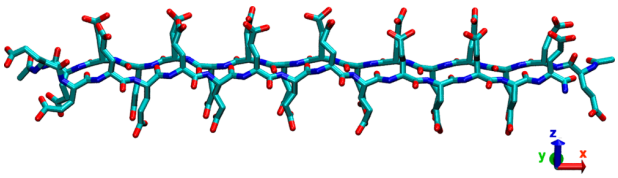


Figure 1. Two peptide strands ($[Ac\text{-}(Glu)_{15}\text{-}CONH_2]_2$) placed in antiparallel arrangement into the periodic box; hydrogen atoms and 15 water molecules terminating the box in the x-direction are not shown. This snapshot structure results after annealing the initial more twisted structure taken from experiment (see Figure S1, Supporting Information) under constraint of periodic boundary conditions.

the initial geometry (see Figure S1 in the Supporting Information) structure was placed in a monoclinic periodic box (66.00 Å × 9.77 Å × 8.07 Å, α = 105°). The remaining space in the x-direction was filled with 15 water molecules, with no waters coming between strands or sheets (i.e., unit cells in y- and z-direction). The initial geometry was based on an analysis of a previously published X-ray pattern of poly-L-glutamic acid (PLGA) powder (Figure S1, Supporting Information).[34,35] Using the Gromacs program[37] and Amber03 force field,[38] the system was equilibrated as an NVT ensemble by simulated annealing, from 1000 to 300 K. For the NVT ensemble, 19 independent annealing cycles were performed, with times ranging from 0.7 to 1.6 ns, to achieve the randomization. Each annealing was followed by a 10 ns equilibration and 100 ns production stage, at this stage with both with NpT and NVT conditions, using T = 300 K and p = 1 atm. For the NpT dynamics, the final box dimensions stabilized at geometries close to the X-ray geometry, with an average dimension (with RMS deviations) of 66.15 (0.19) Å × 9.79 (0.03) Å × 8.09 (0.02) Å.

Spectral Generation. To better represent important vibrational interactions within and between the β-sheet planes and to minimize end-effects, aggregate sheet structures were created for spectral modeling from multiple representations of the periodic box or unit cell (Figure 1). The assembly was expanded in the sheet stacking interaction direction, or z-direction, creating stacks of two or four two-stranded antiparallel sheets (e.g., systems "**z2**" and "**z4**"), and also in the y-direction (interstrand H-bonded direction) to allow consideration of larger, four-stranded sheets (providing systems as "**y2z2**" and "**y2z3**"). The largest **y2z3** assembly is composed of three four-stranded antiparallel sheets, stacked with the same (in-phase) relative alignment, and was used for the final spectral generation, while smaller structures were used for various tests.

Absorption and VCD spectra of these model systems were obtained by computing harmonic force field and atomic polar and axial tensor parameters for smaller fragments. The fragments comprised 8- and 12-amide-containing molecules (illustrated as the **F8** and **F12** structures in Figure 2). By default, the more extended **F12** model was used to provide a

dx.doi.org/10.1021/jp502178d | J. Phys. Chem. B 2014, 118, 6937−6945

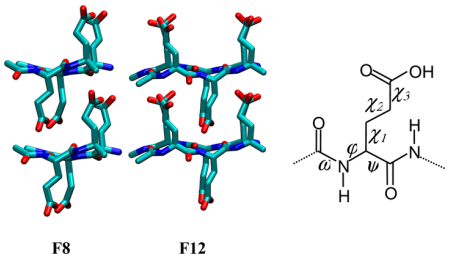


Figure 2. (left) Fragments [Ac-Glu-CONH-CH($CH_2CH_2CO_2H$)-Me·NH_2-Glu_2-CONHMe]$_2$ (**F8**) and [Ac-Glu_2-CONH-CH-($CH_2CH_2CO_2H$)-Me·NH_2-Glu_3-CONHMe]$_2$ (**F12**) used in the DFT computations and (right) principle torsion angles in the Glu residue.

comparison to experiment, whereas the smaller **F8** system enabled us to shorten the computational time and perform additional convergence and other theoretical tests, which are discussed at the end of the following section. To represent the variety of side chain conformations present in the MD runs, 16/24 (for **F8**/**F12**, respectively) fragment conformations were derived from four randomly chosen MD snapshots captured along one trajectory, separated by about 15 ns, both for the *NpT* and *NVT* conditions. Then, fragments had a variety of side-chain conformations and came from both the middle of the sequence and segments adjacent to but not on the termini.

The fragments were then partially geometry optimized in normal mode coordinates.[39−41] By constraining motion along normal coordinates whose modes had frequencies below 300 cm^{-1}, the MD geometry distribution could be largely conserved, while the higher energy coordinates important for the spectra (e.g., C═O stretching) could be completely relaxed. The Gaussian program[42] and the BPW91[43]/6-31G** level of approximation were used. The crystal/fibril environment was simulated using the COSMO dielectric solvent model[44] with a water dielectric constant (ε_r = 78), presumably close to that of the aggregated peptide. Although it is known that the BPW91 functional slightly underestimates harmonic vibrational frequencies,[45] it provides a reasonable and efficient basis for simulation of vibrational properties of the amide and carboxyl groups including their VCD spectra.[33]

The force field and tensors were calculated with the Gaussian programs and transferred[24−26] to the target sheet assemblies (**z2**, **z4**, **y2z2**, **y2z3**) selected for spectral modeling. To complete the transfer, parameters for each atom pair in the three target structures were selected on the basis of determining the best overlap obtained with the corresponding atoms in each of the 16 or 24 fragments (depending on the model, **F8** or **F12**, used). The RMS distance between the closest covalently bound atoms was taken as the criterion for overlap quality. For the assemblies, absorption and VCD spectral frequencies and intensities were generated by usual procedures.[46] Lorentzian profiles (10 cm^{-1} full width at half-maximum) were assigned to each mode, scaled to its intensity, and summed to obtain overall IR (ε) and VCD ($\Delta\varepsilon$) band shapes. The calculated intensities are given in the usual units of L mol^{-1} cm^{-1}, normalized to one amino acid residue.

The spectra were simulated for structural snapshots obtained 10 ns apart for trajectories starting from each of the 19 annealed structures. These were then averaged to provide representations of the final computed spectra for both the *NVT* and *NpT* trajectories. Since experiments are predominantly conducted in D_2O, all spectra were computed to represent deuteration (H/D exchange) of the NH and COOH groups.

As an alternative to the *ab initio*, tensor transfer approach, the transition dipole coupling (TDC) model[47,48] was used to generate the absorption and VCD spectra of **y2z2** as a test. The transition electric dipole moments and frequencies of the C═O stretching vibration were calculated for the CH_3CO—$NDCH_3$ and CH_3COOD (separate computations for the *cis*- and *trans*-OD conformation) molecules at the BPW91/6-31G**/COSMO(H_2O) level and transferred to the fibrillar geometry. For amide I simulations, dipoles were placed in the middle of the C═O bonds, and the TDC Hamiltonian comprised only of the dipole−dipole interactions was diagonalized. This procedure does not include the frequency dispersion due to the DFT force field, since all the amide or —COOD oscillators are initially degenerate and the resultant dispersion is only from their dipolar coupling. From the obtained coupled frequencies and intensities, the spectra were generated as described above.

RESULTS AND DISCUSSION

Molecular Dynamics. The *NpT* conditions could not be maintained during the annealing because at computed high temperatures the strands would not remain H-bonded and would effectively "evaporate". Consequently, we computed the 19 *NVT* annealed structures and then ran both *NVT* and *NpT* trajectories starting from those initial structures. The final *NpT* elementary cell size (periodic box) dimensions were very close to the original postulated *NVT* structure of 66.00 × 9.77 × 8.07 $Å^3$ based on the powder X-ray studies.[35] This indicates that the model geometry used as well as the Amber03 force field might be realistic for PLGA fibrils.

Our postulated conformation corresponds to the β_2 form of PLGA which forms fibrillar structures at low pH and higher temperatures.[4,35,49] This compact form with an intersheet spacing of 8.07 Å was initially designated as β_2 by Itoh and Fasman[35] to contrast with the more typical β_1 form. For PLGA, the β_1 form diffraction data indicates a larger intersheet spacing of 9.35 Å, and it condenses as a less dense gel from the soluble pH ∼ 4 (α-helical) PLGA solution immediately upon heating. By contrast, the β_2 form requires extended incubation at higher temperatures, and can be further affected in yield by lower pH. These forms are also possible to obtain with Glu oligopeptides, under somewhat different aggregation conditions, and have both been shown to have antiparallel sheet secondary structures with possibly different interstrand registries.[36] The registry was determined by isotopic labeling with Val and is likely to be a function of the substitution sequence used.

In Figure 3, equilibrium distributions for various torsional angles are plotted as averages of 19 annealing cycles (black traces), which are almost identical to those obtained using only 9 annealing cycles (red traces), indicating convergence. Additionally, the *NVT* and *NpT* trajectories result in very similar angular distributions (Figure 3, left and right) with only minor differences. The ω-torsion angle oscillates around the usual value of 180°, and the relatively large amplitude suggests unusual flexibility of the otherwise planar and conjugated π-electron amide system.[50,51] Only the terminal amides had any propensity for the *cis* conformation, resulting in the small population at $\omega \sim 0°$. Average φ and ψ angles (about −160 and 150°, respectively) correspond reasonably well to canonical antiparallel β-sheet values (−139 and 135°).[52] The side chain angles χ_1 and χ_2 also exhibit surprisingly narrow distributions,

dx.doi.org/10.1021/jp502178d | J. Phys. Chem. B 2014, 118, 6937−6945

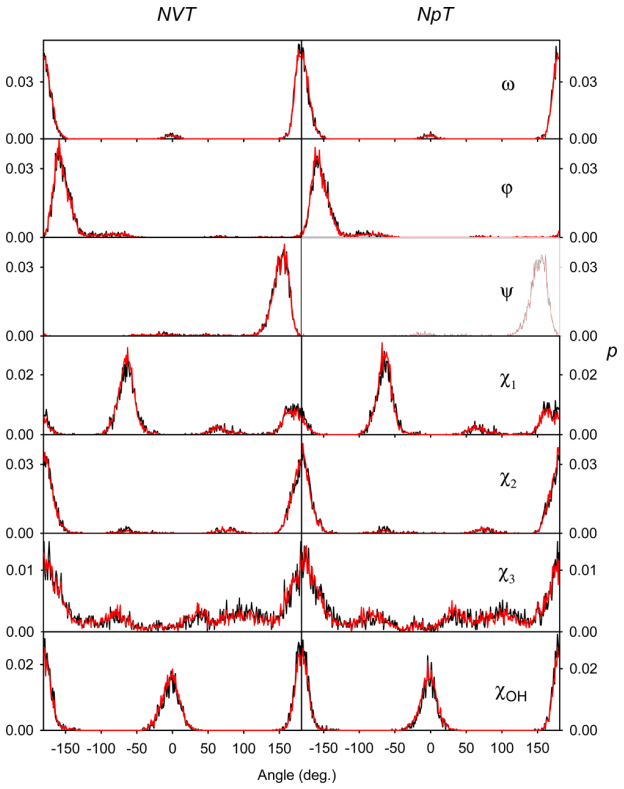


Figure 3. Equilibrium angular distributions calculated for the *NVT* and *NpT* ensembles. The average obtained from 19 annealing cycles is plotted in black, and an intermediate average from 9 cycles is in red.

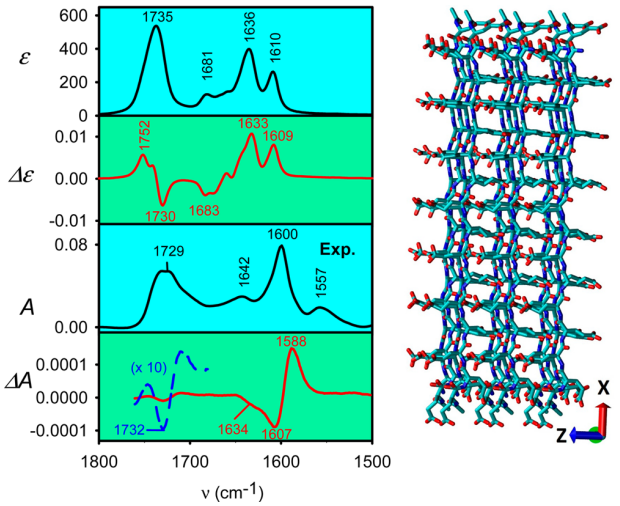


Figure 4. Simulated spectra (*NVT* ensemble) for the **y2z3** fibrillar structure (transfer from **F12**) and experimental spectra for **E10** fibrils formed at pH 1 in DCl/D_2O solvent and room temperature. (For the experimental data, the expanded —COOD part of the VCD spectrum is added as the blue dashed line; the residual experimental absorbance at ~1557 cm^{-1} is probably due to unexchanged amides, i.e., amide II mode; because of difficulties in determining the concentration, absorbance units are used, which suggests comparison is best done on the basis of $\Delta A/A$ values referenced to peak absorbance intensities.) Directions of the Cartesian axis are indicated; *y* and *z* approximately align with the *b* and *c* monoclinic periodic box axes ($\alpha = \angle bc = 105°$).

not much broader than those for φ and ψ, with maxima at about −70 and 180°, respectively, and minor populations for other rotamers (particularly χ_1 at ~70 and ~170°). By contrast, χ_3 apparently can adopt many values but favors an extended conformation ($\chi_3 = 180°$), as indicated by the probability plots. Angle χ_{OH} is again rather constrained, either 0 or 180°, lying in the COO plane, as expected. Although all of the angles strongly prefer one or two positions, the simulations do not indicate that the side chain geometries would adopt a regular crystal-like pattern. At the high temperatures of the annealing (under the *NVT* conditions), the side chains move freely but stay lined up within the gaps provided by the side chains on the opposing sheets, eventually becoming disordered due to χ_3 at the −COOH ends. However, the peptide backbone remains aligned and stacked at T = 300 K for the *NVT* and *NpT* runs, and this common aspect of the structure is maintained in the trajectories.

Comparison of Modeled Spectra to Experiment. The averaged simulated absorption and VCD spectra for the **y2z3** (deuterated) system are plotted in Figure 4, and compared there with the recently published experimental spectra for the Glu_{10} oligopeptide fibrils (**E10**).[36] The **E10** fibril data is given as a typical example; longer poly-glutamic acid chains as well as models with Val or Leu substitutions provide somewhat different experimental VCD spectra that are dependent on conditions of fibril preparation and methods of sampling.[4,36] The IR experimental pattern of PLGA fibrils in various systems is less affected by chain length.[4,49]

In Figure 4, we can see a reasonable correspondence between the observed and simulated IR patterns. The absorption peak calculated at 1735 cm^{-1} and observed at 1729 cm^{-1} corresponds to the C═O stretching vibration of the COOD group (as obtained in D_2O solution). In the experimental spectrum, this peak is slightly split for all Glu sequences but is one band with a lower frequency shoulder for mixed sequences. The splitting is less apparent in the averaged simulated spectrum with the 10 cm^{-1} bandwidth, where only a shoulder is seen. A more detailed analysis (dynamic normal mode displacement) of the simulations suggests that two types of modes contribute to this band, correlated to the *cis* and *trans* orientations of the OD group with respect to the vibrating carbonyl group. The *cis* orientation provides slightly lower frequencies for this vibration, i.e., closer to the amide I region.

The absorption maxima calculated at 1681, 1636 (with a weak feature at ~1670 cm^{-1}), and 1610 cm^{-1} are primarily composed of C═O stretching of the amide groups ("amide I" bands) and may be associated with the experimental maxima evident at 1642 and 1600 cm^{-1}. The computed amide I bands are all higher in frequency than the corresponding experimental values and have a greater apparent dispersion. The —COOD vibrations show closer agreement with experiment; nevertheless, both spectral regions can be considered well-simulated within expectations for the DFT method, which normally overestimates frequencies of C═O stretching motions. Anharmonicity of the C═O stretching potential and solvent−solute interactions not completely encompassed in the continuum solvent model[53,54] both provide sources for part of the error. The predicted C═O stretching absorption maximum intensity is larger for the —COOD than for the amide, whereas the opposite relative dipole strength is observed experimentally. This is probably due to the dispersion distributing intensity differently in the experiment (dominant

dx.doi.org/10.1021/jp502178d | J. Phys. Chem. B 2014, 118, 6937−6945

amide I band, split —COOD) and simulation (split amide I, both intense, and single —COOD) but also may be affected by the limited size of the fibril for which we could simulate spectra, since the experimental fibrils are very long.[4,36]

Overall, however, the simulations seem to provide a reasonable basis for interpreting the experimental spectra, especially for the splitting of the amide I signal into the three components and for the mode assignments. The modes calculated around the 1682 cm^{-1} band provide transition dipole moments predominantly polarized along the *x*-axis, while the 1636 and 1610 cm^{-1} bands are *y*-polarized (see also the decomposition of the spectra into the *x*, *y*, and *z* components for a randomly selected snapshot in Figure S2 (Supporting Information)). This reflects the orientation of the amide groups, more or less following the β-sheet plane.

Simulation of individual polarization components would be particularly important for oriented samples. In some experiments, the *y*-direction would correspond to the fibril growth direction, as the fibrils will predominantly lie in the plane of the sample cell window. However, without some mechanical orientation of the fibrils, it is not practical to take advantage of this internal alignment, since the *x* and *z* polarized modes would be randomly oriented. Consequently, in this study, we concentrate on the more general case of randomly oriented fibril precipitates.

The intense low frequency mode in β-sheets is *y*-polarized, and these two bands, 1636 and 1610 cm^{-1}, correspond to what is normally an intense single band, but are split here, perhaps due to our smaller structure or to edge effects in the simulated structure. The COOD modes are also predominantly polarized in the *y*- and *x*-directions, although comparatively large *z*-components occur as well, and the vibrations do not appear to be split according to the polarization.

The spectra in Figure 4 are averages of predictions for ∼190 structures obtained as snapshots from the MD runs. It is useful to ask how the spectra for the 19 annealed structures vary. There is a substantial intensity variation between them, but the basic spectral pattern and the frequencies are in fair agreement. In all, the −COOD feature at ∼1740 cm^{-1} is much more intense than other bands, but the modes associated with it all lie within a narrow range, having much less dispersion than the amide I, since it arises from only TDC coupling of the −COODs which occurs over larger distances. This lower dispersion leads to the overall high band intensity, while individual modes have similar intensities to those modes contributing to the amide I bands. The simulated amide I intensity for the snapshots is spread over four to five components, with the most intense being that at ∼1640 cm^{-1}. Most spectra have substantial bands at ∼1615 and ∼1655 cm^{-1}, although frequencies and intensities vary. Weaker bands occur at ∼1675−1685 and ∼1580−1590 cm^{-1} in some spectra. Different trajectories are qualitatively similar, but amide I intensities vary as do different snapshots in a trajectory.

The main simulated VCD features (Figure 4) are also qualitatively consistent with the experimental results for PLGA and oligomeric variants, as well as for some other fibril structures,[4] in terms of peak positions, spectral shapes, and the VCD/IR intensity ratios ($\Delta A/A = \Delta\varepsilon/\varepsilon$) if they are correlated to the computed IR absorbance spectra (thus accounting for differences in dispersion). Note that accurate determination of the fibril concentration is difficult, and thus, the experimental absorption/VCD spectra are given as dimensionless values of absorbance or differential absorbance.

The most intense experimental amide I peak in the IR corresponds to a positive "−,+" couplet. Each of the intense computed amide I features (although split) is associated with a couplet contribution of the same sign,[31] often within the broadened band shape used for comparison with experiment. As in the IR, the computed positive VCD at ∼1609 cm^{-1} does not have an experimental counterpart, perhaps due to an imperfect averaging or edge effects in the simulation. Despite the split amide I band, the overall couplet pattern in simulated VCD is qualitatively suggestive of the experimental results, with the splitting of the more intense amide I modes resulting in a more complex oscillating pattern that is exacerbated by our extensive conformational averaging.

The simulated VCD/IR ratio for the amide signal (e.g., $\Delta\varepsilon_{1633}/\varepsilon_{1636} \sim 3 \times 10^{-5}$) is much smaller than that for the PLGA experiment (e.g., $\Delta A_{1588}/A_{1600} \sim 2 \times 10^{-3}$). Such a low simulated VCD intensity is consistent with our previous DFT-based results for models of idealized structures that have no twisting in or between sheets.[19,20] Part of this difference can be explained by the flattening that modified the initial slightly twisted structure (Figure S1, Supporting Information) during the annealing in order to make the model structure better fit the periodic boundary conditions of the monoclinic box. Indeed, when we simulated the VCD spectra of the twisted alanine analogue, the $\Delta\varepsilon/\varepsilon$ ratio increased by about 10 times (see Figure S3, Supporting Information). The relative rigidity of the periodic structure is a drawback of the present model, which is otherwise well-suited to capture the compact geometry in the β_2 form of the PGA fibril, including the side chain dynamics and interactions.

As expected, VCD changes more dramatically than the absorption spectra for the different annealing cycles. In general, the net couplet nature, positive to low frequency, of the amide I VCD seen in the average VCD is detectable for each snapshot, but the shapes and positions of the bands encompass a wide variety of combinations. There is not a dominant interaction coming through in the computed spectra but rather an underlying consistent pattern that persists on the summing of many structures.

For the −COOD signal, the simulated VCD/IR intensity ratio (e.g., $\Delta\varepsilon_{1730}/\varepsilon_{1735} \sim 10^{-5}$) is again smaller than that for the PLGA experiment (e.g., $\Delta A_{1732}/A_{1729} \sim 2 \times 10^{-4}$) but has better agreement than that for the amide I. Its simulated VCD is thus relatively more intense than for experiment when compared to the amide I but is actually much weaker in absolute terms, and is composed of several overlapping alternating sign features, also as seen experimentally. This sort of pattern reflects coupling of the −COOD groups, but the detailed shape is difficult to compare to experiment due to the multiple underlying bands. The overall +/− (1752/1739 cm^{-1}) simulated couplet is negatively biased, centered around 1740 cm^{-1}, which would only approximately correspond to the dominant negative 1732 cm^{-1} PLGA experimental VCD. Note, however, that the experimental shape does vary significantly according to experimental conditions, substitution patterns in hetero-oligomers, and even the length of the poly-glutamic acid main chains. In many of these variants, this band exhibits a −/+ couplet (going to low frequency) which corresponds to the two lower components of the three-peak pattern for the experimental VCD of Glu_{10} shown in the bottom panel of Figure 4.[4,36] For our crystal model, reproducing the complete sign pattern for the solution phase fibrils might be a high goal.

dx.doi.org/10.1021/jp502178d | J. Phys. Chem. B 2014, 118, 6937−6945

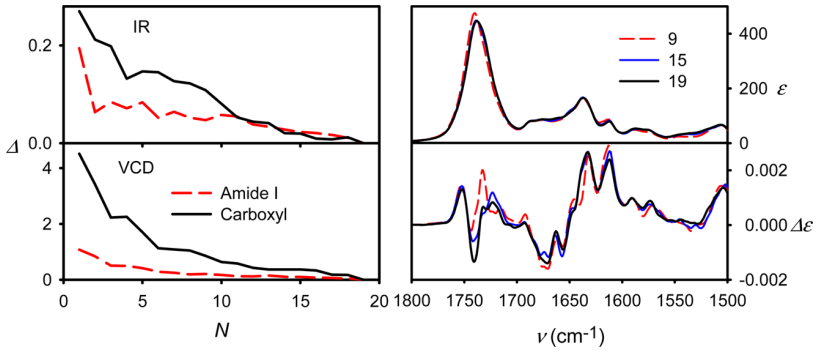


Figure 5. Absorption (top) and VCD (bottom) spectral convergence with the number of independent annealing cycles, for **y2z2**. On the left-hand side, relative integral errors ($\Delta = \int|S - S_f|\,d\nu/(\int|S|\,d\nu \int|S_f|\,d\nu)^{1/2}$, where S_f is the reference spectrum) are plotted for the carboxyl (1800−1700 cm^{-1}) and amide I (1700−1550 cm^{-1}) regions. On the right, spectra obtained by averaging of 9, 15, and 19 annealing cycles are plotted.

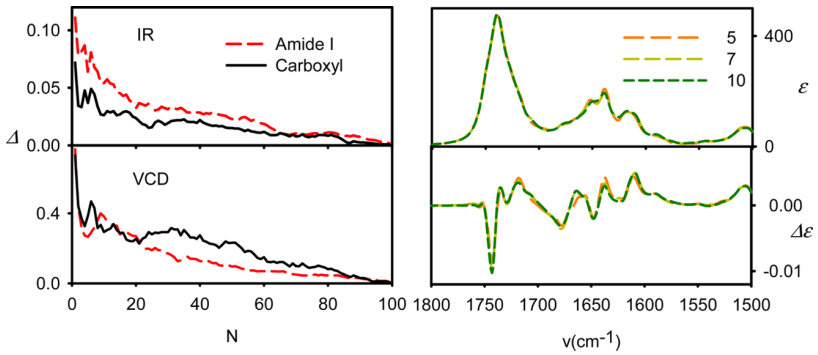


Figure 6. Absorption (top) and VCD (bottom) spectral convergence with the number of snapshots taken from a trajectory corresponding to one annealing cycle and the **y2z2** structure (as in Figure 5).

Nonetheless, the relative intensities suggest we are getting some residual local ordering of the COOD groups.

Probably the most important observation is that, in both the experiment and these MD structure-DFT computations, the −COOD band develops significant VCD even though the structures do not show an obvious systematic side-chain ordering beyond the preferred angles (Figure 3). Normally, such modes would have very little VCD due to their locally achiral structures and the low level (due to longer distance) and disordered coupling to other like modes. Certainly these oligomers and PLGA itself at higher pH have no VCD from the carboxylate, COO^-, bands (shifted below the amide I). However, the stacking of sheets constrains the side chains so that they sample a reduced configurational space, as seen in the distributions shown in Figure 3. Beyond this average distribution, the local structures presumably favor a distribution of near neighbor orientations so that the local dipolar coupling is not random. Indeed, our snapshot structures often have pairwise chiral coupling of multiple −COOD groups, which presumably do not cancel on averaging and result in a significant residual VCD.

Spectra for Fibrils of Different Sizes. The restricted size of the models used for both the fragment and extended fibril can possibly cause errors in the simulated spectral patterns. To estimate the effects of terminating the fibril in either direction (extension of H-bonds or stacking), we simulated the IR and VCD spectra for the smaller **z2**, **z4**, and **y2z2** models (as shown in Figure S4, Supporting Information) using the smaller **F8** fragment. All three structures provide similar spectral patterns, not much different from the larger **y2z3** model (Figure 4). This suggests that, as expected, the spectral shapes are mostly determined by short-range interactions between neighboring (in-strand) and direct cross-strand amide groups via covalent and dipolar interactions.

Indeed, if we compute only the spectral contributions of the inner, H-bonded amides and carboxyl groups, i.e., eliminate those from the surface groups in **y2z3** (by deleting the atomic polar and atomic axial tensors, see Figure S5, Supporting Information), this results in ∼30% loss of intensity in both the IR and VCD due to fewer oscillators contributing to the simulated spectra, but only relatively small changes occur in the correlated amide I IR and VCD patterns. However, the −COOD VCD signal changes more with respect to its absorbance. The intensity is reduced by ∼50%, and the original overall couplet is distorted to a different pattern, W-shaped (+/−/+, 1747/1735/1725 cm^{-1}). The latter shape might offer an improved agreement with experiment, but the sign is a problem and shape changes associated with this elimination of surface residues suggest that getting better agreement will require computing spectra for much larger structures, which currently is not possible for us. The loss of IR intensity is less dramatic and presumably corresponds to loss of those −COOH groups on the outside of the sheet which have no explicit acceptors for potential H-bonded interactions and are eliminated in the surface deleted computations. The variation in the VCD pattern for this mode is consistent with the experimental variations in its VCD,. as seen for glutamic acid fibrils prepared in different ways.[4,31]

The C═O vibrations in the −COOD groups most likely interact by dipolar coupling.[33] Changing the model fibril

dx.doi.org/10.1021/jp502178d | J. Phys. Chem. B 2014, 118, 6937−6945

structure to add potential interactions, i.e., between aligned sheet planes (the **z2** → **z4** elongation), and increasing the extent of cross-strand coupling in single β-sheet layers (**z2** → **y2z2**) result in relatively minor spectral changes for these side-chain modes (Figure S4, Supporting Information). That the increasing size does not increase VCD intensities or the VCD/absorption ratio (i.e., $\Delta A/A$, the g-factor) suggests that this model does not encompass the underlying interaction that leads to enhanced VCD intensity[28,55] which has been reported for some fibril structures, and is primarily seen in the amide I modes.[13] Similarly, the simulated spectra are stable with respect to the type of the dynamics: The NpT and NVT ensembles lead to very similar spectra for **y2z2** (see Figure S6, Supporting Information). In the amide I and —COOD regions, the VCD has the same overall profile after averaging over simulations for structures from both sets of trajectories, with only minor variation in detailed shapes.

Convergence of the MD Averaging. Given the partial stochastical character of the side chain conformations, stability of the simulated spectra with respect to MD parameters is of primary interest. As can be seen from Figure 5 for **y2z2** (simulated with **F8**), the absorption and VCD patterns converge with increasing number of independent annealing cycles that we sample to get initial structures for the trajectories, but VCD shows more variation, primarily in terms of spectral shapes. As expected, the −COOD IR and VCD converge more slowly than the amide I, since the side chains are more flexible. Similar convergence behavior is seen (Figure 6) with an increase in the number of snapshots averaged along a trajectory for a randomly selected annealing **y2z2** geometry, except that the amide I IR showed more variance than the −COOD IR for fewer snapshots. These MD simulations suggest that the residual VCD of the glutamic acid side chains originates in a net chiral relative orientation of the COOD groups.

Atomic Contributions to the Spectra and the TDC Model. To understand the fibrillar spectra and the relative contribution of other modes to the side-chain —COOD spectra in more detail, absorption and VCD spectra for a **y2z2** snapshot are simulated with and without contribution of the atomic intensity tensors localized on the amide (HN—C═O) atoms (Figure 7). These results show that the C═O stretching modes from both the amide and —COOD contribute almost independently to the spectra. There is only a minor contribution of the amide to the split VCD signal of —COOD at ∼1740 cm^{-1}, which is apparent as an intensity change. This suggests that interaction of the C═O transition dipoles, such as within the TDC mechanism,[50,51] is the primary source of observable VCD. For the —COOD VCD, the positive components of the "W"-like shape (the small and large maxima at 1760 and 1725 cm^{-1}) are related to predominantly y-polarized modes, whereas modes contributing to the negative lobe at 1740 cm^{-1} are x-polarized, which is consistent with the patterns seen in our larger, **y2z3** system simulation (Figure 4).

As a test of the applicability of the simpler dipole coupling (TDC) model for simulating these spectra, comparison calculations specifically for the amide and —COOD C═O stretching modes were made for the same **z2y2** structure using transition dipoles computed using N-methylacetamide and acetic acid as sources of the transition dipole moments, as described in the Methods section. Addition of transition dipoles corresponding to other modes, such as amide II (N—D bending and C—N stretching) and amide A (N—D stretching), changed the resultant TDC spectral pattern by less than 10%.

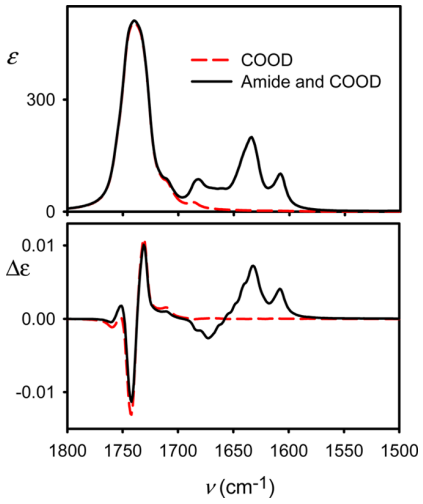


Figure 7. Absorption and VCD spectra of one **y2z2** snapshot simulated with atomic intensity tensors from just the −COOD and from both −COOD and amide groups.

One can immediately see that the TDC spectra (Figure 8) are simpler than the DFT ones (Figure 7), presumably due to the reduced set of interactions considered with the TDC method. The amide I IR is predicted as only an intense single band, while the −COOD yields a split band. Due to this change in mode dispersion, the VCD pattern that results is at least quantitatively different with the two models. The intense IR single band amide I results in an intense positive VCD couplet, which does reflect the **E10** experimental results[36] and provides only a qualitative match to the DFT results in terms of the VCD sign pattern and intensity.

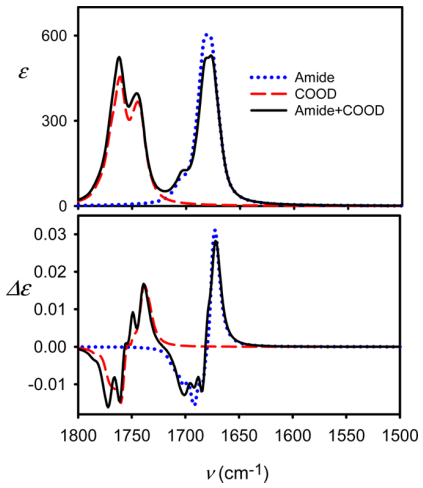


Figure 8. Absorption and VCD spectra of **y2z2** (same snapshot structure as in Figure 7) simulated using the TDC approximation, for transition dipoles on the carboxyl, the amide, and on both chromophores.

As an inherent property of the TDC mechanism, the VCD intensity is conservative (both the −COOD and amide groups give positive and negative bands of nearly the same areas). The TDC model potentially includes longer-range interactions, not

dx.doi.org/10.1021/jp502178d | J. Phys. Chem. B 2014, 118, 6937−6945

present in the DFT computations due to the transfer of parameters from limited size fragments, but these long-range couplings do not have a large effect, although they might explain the −COOD IR band splitting, reproduced only partially by DFT (cf. the discussion of Figure 4). The $\Delta\varepsilon/\varepsilon$ ratio obtained by TDC is about the same as that for DFT, probably because of the neglect of the fibril twist in both models. The −COOD VCD intensity in Figures 7 and 8 is enhanced, since these are single snapshots without averaging. Averaging over many snapshots causes more cancellation in the side-chain modes due their increased variation in conformation.

We can summarize what was learned from the modeling of poly-glutamic acid side chain VCD: It originates in a partial chiral orientation of the −COOD groups on the side chains and their dipole−dipole interaction. It is fairly uncoupled to the amide and other peptide vibrational modes, and is polarized within the side chain layer sandwiched between the β-sheet planes. Using these insights, the VCD of this or other side chains can possibly be used to probe the geometry of the intersheet space as well as to complement the information about fibril structure obtained from the amide I modes.

CONCLUSIONS

In order to explain the VCD spectra observed for the side chains of the poly-glutamic acid, which has been shown to form dense β_2 fibrils, we adopted a crystal-like geometry model. With several annealing cycles and a molecular dynamic equilibration, we verified that the structure was stable both at the *NVT* and *NpT* conditions. The angular distribution thus obtained confirmed the side chains maintained a limited local ordering, although not a periodic or crystalline structure. Using the Cartesian coordinate-based transfer of the force field and intensity tensors calculated for smaller fragments, absorption and VCD spectra of larger stacked sheet segments could be simulated that showed qualitative features in common with experimental data on Glu_{10} (**E10**) fibrillar systems. These simulations gave IR spectral patterns in reasonable agreement with experiment, although edge effects seem to distort the dispersion. We were not able to replicate the VCD enhancement seen experimentally for the amide I mode in fibrils. In spite of the statistical character, the averaged spectral patterns converged relatively quickly with the number of MD snapshots and annealing cycles. Analysis of the calculated side-chain VCD indicated that it reflects a residual chiral mutual −COOH group coupling, developed through population of a few favored conformations, but evidences only limited interaction with the amide group. VCD spectroscopy combined with the simulations is thus able to recognize very specific geometry features of fibrillar structures.

ASSOCIATED CONTENT

Supporting Information

Computational details and various computational tests. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: tak@uic.edu.

*E-mail: bour@uochb.cas.cz.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was supported by the Academy of Sciences (M200550902), Grant Agency of the Czech Republic (P208/11/0105), and Ministry of Education (LH11033, LM2010005, CZ.1.05/3.2.00/08.0144). The work was done in part while T.A.K. was a guest Professor at the University of Konstanz supported as a Research Awardee of the Alexander von Humboldt Foundation.

REFERENCES

(1) Harrison, P. M.; Bamborough, P.; Daggett, V.; Prusiner, S. B.; Cohen, F. E. The Prion Folding Problem. *Curr. Opin. Struct. Biol.* **1997**, *7*, 53−59.

(2) Miranker, A. D.; Dobson, C. M. Collapse and Cooperativity in Protein Folding. *Curr. Opin. Struct. Biol.* **1996**, *6*, 31−42.

(3) Morris, K. L.; Serpell, L. C. X-Ray Fibre Diffraction Studies of Amyloid Fibrils. *Methods Mol. Biol.* **2012**, *849*, 121−135.

(4) Fulara, A.; Lakhani, A.; Wójcik, S.; Nieznańska, H.; Keiderling, T. A.; Dzwolak, W. Spiral Superstructures of Amyloid-Like Fibrils of Polyglutamic Acid: An Infrared Absorption and Vibrational Circular Dichroism Study. *J. Phys. Chem. B* **2011**, *115*, 11010−11016.

(5) Benditt, E. P.; Eriksen, N.; Berglund, C. Congo Red Dichroism with Dispersed Amyloid Fibrils, an Extrinsic Cotton Effect. *Proc. Natl. Acad. Sci. U.S.A* **1970**, *66*, 1044−1051.

(6) Yamamoto, S.; Watarai, H. Raman Optical Activity Study on Insulin Amyloid and Prefibril Intermediate. *Chirality* **2012**, *24*, 97−103.

(7) Reichert, D.; Pascui, O.; deAzevedo, E. R.; Bonagamba, T. J.; Arnold, K.; Huster, D. A Solid-State NMR Study of the Fast and Slow Dynamics of Collagen Fibrils at Varying Hydration Levels. *Magn. Reson. Chem.* **2004**, *42*, 276−284.

(8) Kanaori, K.; Nosaka, A. Y. Characterization of Human Calcitonin Fibrillation in Aqueous Urea Solution by H-1 NMR Spectroscopy. *Biochemistry* **1996**, *35*, 12671−12676.

(9) Dzwolak, W.; Loksztejn, A.; Galinska-Rakoczy, A.; Adachi, R.; Goto, Y.; Rupnicki, L. Conformational Indeterminism in Protein Misfolding: Chiral Amplification on Amyloidogenic Pathway of Insulin. *J. Am. Chem. Soc.* **2007**, *129*, 7517−7522.

(10) Maa, S.; Cao, X.; Mak, M.; Sadik, A.; Walkner, C.; Freedman, T. B.; Lednev, I.; Dukor, R.; Nafie, L. Vibrational Circular Dichroism Shows Unusual Sensitivity to Protein Fibril Formation and Development in Solution. *J. Am. Chem. Soc.* **2007**, *129*, 12364−12365.

(11) Kurouski, D.; Lombardi, R. A.; Dukor, R. K.; Lednev, I. K.; Nafie, L. A. Direct Observation and pH Control of Reversed Supramolecular Chirality in Insulin Fibrils by Vibrational Circular Dichroism. *Chem. Commun.* **2010**, *46*, 7154−7156.

(12) Kurouski, D.; Dukor, R. K.; Lu, X.; Nafie, L. A.; Lednev, I. K. Spontaneous Inter-Conversion of Insulin Fibril Chirality. *Chem. Commun.* **2012**, *48*, 2837−2839.

(13) Kurouski, D.; Kar, K.; Wetzel, R.; Dukor, R. K.; Lednev, I. K.; Nafie, L. A. Levels of Supramolecular Chirality of Polyglutamine Aggregates Revealed by Vibrational Circular Dichroism. *FEBS Lett.* **2013**, *578*, 1638−1643.

(14) Nishijima, M.; Tanaka, H.; Yang, C.; Fukuhara, G.; Mori, T.; Babenko, V.; Dzwolak, W.; Inoue, Y. Supramolecular Photochirogenesis with Functional Amyloid Superstructures. *Chem. Commun.* **2013**, *49*, 8916−8918.

(15) Babenko, V.; Piejko, M.; Wójcik, S.; Mak, P.; Dzwolak, W. Vortex-Induced Amyloid Superstructures of Insulin and its Component A and B Chains. *Langmuir* **2013**, *29*, 5271−5278.

(16) Babenko, V.; Dzwolak, W. Amino Acid Sequence Determinants in Self-Assembly of Insulin Chiral Amyloid Superstructures: Role of C-Terminus of B-Chain in Association of Fibrils. *FEBS Lett.* **2013**, *587*, 625−630.

(17) Dzwolak, W.; Surmacz-Chwedoruk, W.; Babenko, V. Conformational Memory Effect Reverses Chirality of Vortex-Induced Insulin Amyloid Superstructures. *Langmuir* **2013**, *29*, 365−370.

(18) Wójcik, S.; Babenko, V.; Dzwolak, W. Insulin Amyloid Superstructures as Templates for Surface Enhanced Raman Scattering. *Langmuir* **2010**, *26*, 18303−18307.
(19) Welch, W. R. W.; Keiderling, T. A.; Kubelka, J. Structural Analyses of Experimental 13c Edited Amide I′ IR and VCD for Peptide β-Sheet Aggregates and Fibrils Using DFT-Based Spectral Simulations. *J. Phys. Chem. B* **2013**, *117*, 10359−10369.
(20) Welch, W. R. W.; Kubelka, J.; Keiderling, T. A. Infrared, Vibrational Circular Dichroism, and Raman Spectral Simulations for β-Sheet Structures with Various Isotopic Labels, Interstrand, and Stacking Arrangements Using Density Functional Theory. *J. Phys. Chem. B* **2013**, *117*, 10343−10358.
(21) Measey, T.; Schweitzer-Stenner, R. Vibrational Circular Dichroism as a Probe of Fibrillogenesis: The Origin of the Anomalous Intensity Enhancement of Amyloid-Like Fibrils. *J. Am. Chem. Soc.* **2011**, *133*, 1066−1076.
(22) Karjalainen, E. L.; Ravi, H. K.; Barth, A. Simulation of the Amide I Absorption of Stacked β-Sheets. *J. Phys. Chem. B* **2011**, *115*, 749−757.
(23) Mandal, P.; Eremina, N.; Barth, A. Formation of Two Different Types of Oligomers in the Early Phase of pH-Induced Aggregation of the Alzheimer Aβ(12−28) Peptide. *J. Phys. Chem. B* **2012**, *116*, 12389−12397.
(24) Bouř, P.; Sopková, J.; Bednárová, L.; Maloň, P.; Keiderling, T. A. Transfer of Molecular Property Tensors in Cartesian Coordinates: A New Algorithm for Simulation of Vibrational Spectra. *J. Comput. Chem.* **1997**, *18*, 646−659.
(25) Yamamoto, S.; Li, X.; Ruud, K.; Bouř, P. Transferability of Various Molecular Property Tensors in Vibrational Spectroscopy. *J. Chem. Theory Comput.* **2012**, *8*, 977−985.
(26) Bieler, N. S.; Haag, M. P.; Jacob, C. R.; Reiher, M. Analysis of the Cartesian Tensor Transfer Method for Calculating Vibrational Spectra of Polypeptides. *J. Chem. Theory Comput.* **2011**, *7*, 1867−1881.
(27) Bouř, P.; Keiderling, T. A. Ab Initio Simulation of the Vibrational Circular Dichroism of Coupled Peptides. *J. Am. Chem. Soc.* **1993**, *115*, 9602−9607.
(28) Andrushchenko, V.; Bouř, P. Circular Dichroism Enhancement in Large DNA Aggregates Simulated by a Generalized Oscillator Model. *J. Comput. Chem.* **2008**, *29*, 2693−2703.
(29) Bak, K. L.; Jorgensen, P.; Helgaker, T.; Ruud, K. Basis Set Convergence and Correlation Effects in Vibrational Circular Dichroism Calculations Using London Orbitals. *Faraday Discuss.* **1994**, *99*, 121−129.
(30) Cheeseman, J. R.; Frisch, M. J.; Devlin, F. J.; Stephens, P. J. Ab Initio Calculation of Atomic Axial Tensors and Vibrational Rotational Strengths Using Density Functional Theory. *Chem. Phys. Lett.* **1996**, *252*, 211−220.
(31) Setnička, V.; Huang, R.; Thomas, C. L.; Etienne, M. A.; Kubelka, J.; Hammer, R. P.; Keiderling, T. A. IR Study of Cross-Strand Coupling in a Beta-Hairpin Peptide Using Isotopic Labels. *J. Am. Chem. Soc.* **2005**, *127*, 4992−4993.
(32) Huang, R.; Setnička, V.; Etienne, M. A.; Kim, J.; Kubelka, J.; Hammer, R. P.; Keiderling, T. A. Cross-Strand Coupling of a β-Hairpin Peptide Stabilized with an Aib-Gly Turn Using Isotope-Edited IR Spectroscopy. *J. Am. Chem. Soc.* **2007**, *129*, 13592−13603.
(33) Kubelka, J.; Kim, J.; Bouř, P.; Keiderling, T. A. Contribution of Transition Dipole Coupling to Amide Coupling in IR Spectra of Peptide Secondary Structures. *Vib. Spectrosc.* **2006**, *42*, 63−73.
(34) Keith, H. D.; Giannoni, G.; Padden, F. J. Single Crystal of Poly(L-Glutamic Acid). *Biopolymers* **1969**, *7*, 775−792.
(35) Itoh, K.; Foxman, B. M.; Fasman, G. D. The Two β-Forms of Poly(L-Glutamic Acid). *Biopolymers* **1976**, *15*, 419−455.
(36) Chi, H.; Welch, W. R. W.; Kubelka, J.; Keiderling, T. A. Insight into the Packing Pattern of β2 Fibrils: A Model Study of Glutamic Acid Rich Oligomers with 13-C Isotopic Edited Vibrational Spectroscopy. *Biomacromolecules* **2013**, *14*, 3880−3891.
(37) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; vanderSpoel, D.; et al. Gromacs 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *27*, 845−854.
(38) Kamiya, N.; Watanabe, Y. S.; Ono, S.; Higo, J. Amber-Based Hybrid Force Field for Conformational Sampling of Polypeptides. *Chem. Phys. Lett.* **2005**, *401*, 312−317.
(39) Bouř, P.; Keiderling, T. A. Partial Optimization of Molecular Geometry in Normal Coordinates and Use as a Tool for Simulation of Vibrational Spectra. *J. Chem. Phys.* **2002**, *117*, 4126−4132.
(40) Bouř, P. Convergence Properties of the Normal Mode Optimization and its Combination with Molecular Geometry Constraints. *Collect. Czech. Chem. Commun.* **2005**, *70*, 1315−1340.
(41) Hudecová, J.; Hopmann, K. H.; Bouř, P. Correction of Vibrational Broadening in Molecular Dynamics Clusters with the Normal Mode Optimization Method. *J. Phys. Chem. B* **2012**, *116*, 336−342.
(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, revision B01; Gaussian, Inc.: Wallingford, CT, 2009.
(43) Becke, A. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(44) Klamt, A. Cosmo and Cosmo-Rs. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, U.K., 1998; Vol. *1*, pp 604−615.
(45) Hudecová, J.; Profant, V.; Novotná, P.; Baumruk, V.; Urbanová, M.; Bouř, P. Ch Stretching Region: Computational Modeling of Vibrational Optical Activity. *J. Chem. Theory Comput.* **2013**, *9*, 3096−3108.
(46) Nafie, L. *Vibrational Optical Activity: Principles and Applications*; Wiley: Chichester, U.K., 2011.
(47) Holzwarth, G.; Chabay, I. Optical Activity of Vibrational Transitions: A Coupled Oscillator Model. *J. Chem. Phys.* **1972**, *57*, 1632−1635.
(48) Zhong, W.; Gulotta, M.; Goss, D. J.; Diem, M. DNA Solution Conformation Via Infrared Circular Dichroism: Experimental and Theoretical Results for B-Family Polymers. *Biochemistry* **1990**, *29*, 7485−7491.
(49) Yamaoki, Y.; Imamura, H.; Fulara, A.; Wójcik, S.; Bożycki, Ł.; Kato, M.; Keiderling, T. A.; Dzwolak, W. An FT-IR Study on Packing Defects in Mixed β-Aggregates of Poly(L-glutamic Acid) and Poly(D-Glutamic Acid): A High-Pressure Rescue from a Kinetic Trap. *J. Phys. Chem. B* **2012**, *116*, 5172−5178.
(50) Rick, S. W.; Cachau, R. E. The Nonplanarity of the Peptide Group: Molecular Dynamics Simulations with a Polarizable Two-State Model for the Peptide Bond. *J. Chem. Phys.* **2000**, *112*, 5230−5241.
(51) Andrushchenko, V.; Matějka, P.; Anderson, D. T.; Kaminský, J.; Horníček, J.; Paulson, L. O.; Bouř, P. Solvent Dependence of the N-Methylacetamide Structure and Force Field. *J. Phys. Chem. A* **2009**, *113*, 9727−9736.
(52) Creighton, T. E. *Proteins: Structures and Molecular Properties*, 2nd ed.; W. H. Freeman and Co.: New York, 1993.
(53) Kubelka, J.; Keiderling, T. A. Ab Initio Calculation of Amide Carbonyl Stretch Vibrational Frequencies in Solution with Modified Basis Sets. 1. N-Methyl Acetamide. *J. Phys. Chem. A* **2001**, *105*, 10922−10928.
(54) Bouř, P. On the Influence of the Water Electrostatic Field on the Amide Group Vibrational Frequencies. *J. Chem. Phys.* **2004**, *121*, 7545−7548.
(55) Profant, V.; Baumruk, V.; Li, X.; Šafařík, M.; Bouř, P. Tracking of the Polyproline Folding by Density Functional Computations and Raman Optical Activity Spectra. *J. Phys. Chem. B* **2011**, *115*, 15079−11589.

9.5 Appendix E – *J. Phys. Chem. Lett.* 2015, 6, 3314.

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

First-Principles Predictions of Vibrational Raman Optical Activity of Globular Proteins

Jiří Kessler,[†,‡] Josef Kapitán,[§] and Petr Bouř*[,†]

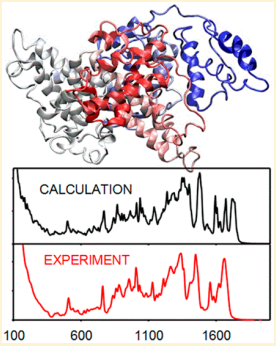
[†]Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Flemingovo náměstí 2, 16610 Prague, Czech Republic

[‡]Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University, Hlavova 8, 12840 Prague, Czech Republic

[§]Department of Optics, Palacký University, 17. listopadu 12, 77146 Olomouc, Czech Republic

Supporting Information

ABSTRACT: Computational methods based on the Schrödinger equation have been traditionally confined to rather small molecules. Using an automatic computational methodology, however, we obtained a stunning agreement between experimental and theoretical vibrational spectra of large globular proteins containing thousands of atoms as well. Principle atomic properties are obtained from small molecular fragments and combined with a minimal accuracy loss. This "first-principles" interpretation of the data reveals a wealth of information, such as nature of localized molecular motions as well as collective vibrational modes describing folding of larger protein parts. A new insight is provided to the origin of the chiroptical effects, and the theory lends the used spectroscopic techniques, unpolarized Raman scattering and vibrational Raman optical activity, immense potential to structural studies of biological systems.



Recent advances in Raman spectroscopy revolutionized many domains including structural chemistry, biomedical research, and material science.[1,2] Raman optical activity (ROA) spectroscopy appears as one of these emerging techniques; it is sensitive to fine details in molecular structure including isotopic isomer effects[3] and has been applied to a plethora of systems including small molecules as well as whole viruses.[4,5] Because of their enhanced sensitivity to structural variations, such chiral spectroscopy techniques may be thought of as versatile alternatives to high-resolution approaches, such as X-ray diffraction. For solutions, they are often indispensable.

The polarization phenomena were linked to molecular structure already by Louis Pasteur;[6] however, only the recent advances in quantum chemistry made it possible to use the information provided by chiroptical experiments in full. Optical activity in the vibrational realm including ROA appears particularly attractive because it provides many well-resolved bands and intensities are tightly linked to the local molecular structure.[7] Note that this is often not the case in the electronic spectra.[8]

Yet after the discovery of molecular ROA[9] suitable computers and simulation and interpretation techniques were not immediately available. The subsequent development in the theoretical field included many steps, such as incorporation of the ROA intensity formulas in the quantum-chemical framework,[10] the origin-independent formulation,[11] and coupling of ROA with the density functional theory (DFT).[12]

Even within the DFT domain, however, it is difficult to simulate vibrational spectra of sizable molecules due to the unfavorable dependence of computational time on the number of atoms (N). This dependence, often referred to as "scaling", can be as sharp as N^5 for energy calculation by advanced DFT methods[13,14] and even less favorable for more accurate simulations or higher-order energy derivatives needed for ROA.[15] Clearly, alternative approaches need to be developed to efficiently combine spectroscopic and theoretical methods in structural biology.

In the present study, we use an automatized variant of the Cartesian coordinate transfer (CCT) technique to extend the quantum-chemical methodology for accurate simulations and understanding of ROA spectra of five different globular proteins. The CCT concept[16] is based on the locality of the molecular property tensors[17] that determine intensities and positions of spectral bands and the spring-like interatomic interactions involved in molecular vibrations. In the language of practical computations, we split the target molecule to chemically meaningful fragments, for which the relevant properties—tensors—can be obtained at high precision, for example, by DFT, and transfer them back to the target "atom by atom". Working in the Cartesian coordinates rather than in the previously used intrinsic coordinates is convenient if automatic computer routines are employed; the transfer then consists of fairly simple operations such as rotation and origin-dependence corrections. Although the method may introduce some error,[18,19] its precision is controllable and mostly

Received: July 14, 2015
Accepted: August 4, 2015
Published: August 4, 2015

DOI: 10.1021/acs.jpclett.5b01500
J. Phys. Chem. Lett. 2015, 6, 3314−3319

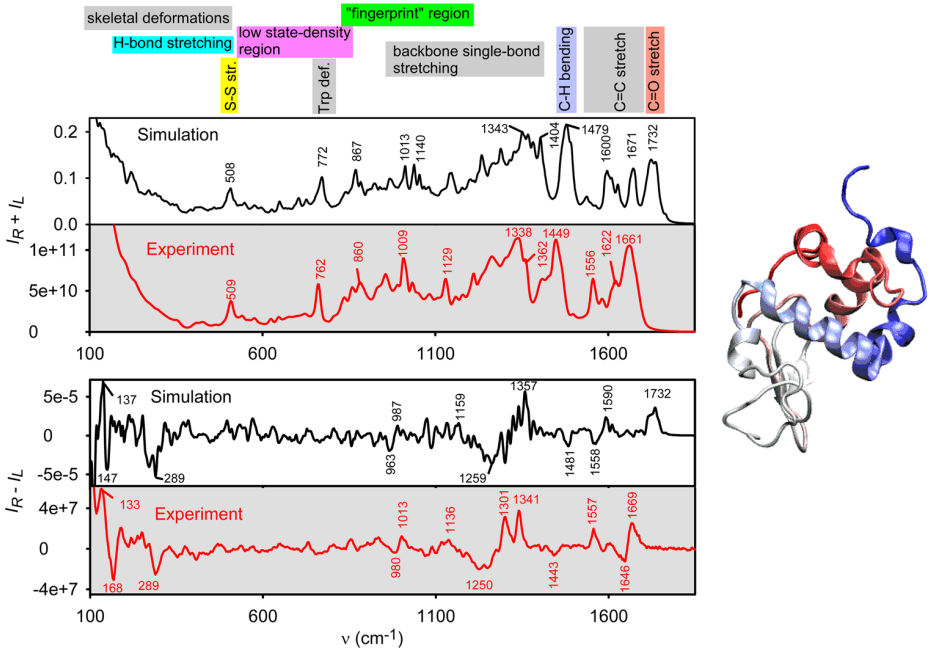


Figure 1. Comparison of calculated and experimental spectra. For bovine α-lactalbumin the Raman ($I_R + I_L$, top) and ROA ($I_R - I_L$, bottom) spectra obtained at the B3BPW91/6-31++G**/CPCM approximation level exhibit a surprisingly good correspondence with the experiment and allow for an accurate assignment of most vibrational bands. Selected vibrational classes are indicated at the top.

sufficient to explain experimental data.[20] For the globular proteins reported here, we achieved stunning accuracy in vibrational frequencies and spectral intensities so far achievable for comparable quantum chemical simulations in small molecules only.

Already early protein studies highlighted the sensitivity of ROA to backbone stereochemistry,[21] and the possibility to simulate and interpret the spectra from the first-principles significantly encouraged the development of the technique. Lately, a "brute force" DFT computation provided reasonable ROA spectra of the β domain of rat metallothionein.[22] Similarly, the CCT method helped us to understand the spectra and, more importantly, the solution structures of valinomycin[23,24] and insulin;[25] however, the much larger globular proteins explored in the present study could only be tackled after the fragmentation and transfer procedure was fully automated. To a certain extent, the transfer approach also makes it possible to explore the effect of conformational flexibility on the spectra, as many snapshots obtained from molecular dynamics can be averaged. The protein samples comprise bovine α-lactalbumin, concanavalin-A, hen egg-white lysozyme, human lysozyme, and human serum albumin. Concanavalin-A is an example of β-sheet-dominated protein; the others have a higher content of α-helix but exhibit finer structural differences. The human serum albumin is the largest molecule with the highest α-helical ratio.

Details of the protein structure and used computational methodology can be found in the Supporting Information. In the simplest case, X-ray protein geometry was divided into many overlapping four-amide residues, and these fragments were partially optimized using the vibrational normal mode coordinates[26] to preserve their conformation in the protein but relax principal vibrational motions. Then, vibrational spectroscopic parameters (force field, ROA tensors) were calculated at the B3PW91[27]/6-31++G**/CPCM[28,29] level using the Gaussian software[30] and transferred back to the protein. The Cartesian coordinate transfer procedure is described in detail elsewhere[16,20] and consists of the rotation of individual tensor components while taking into account their origin dependence. Note that absolute Raman intensities cannot be easily measured; the spectra are plotted in atomic and instrumental units for the simulation and experiment, respectively.

The reliability of the simulations is documented in Figure 1 for α-lactalbumin. The Raman intensity increases steeply and nearly monotonically toward the lowest-wavenumber region (300−100 cm^{-1}) because of the large signal and temperature excitations of the lowest-energy vibrational modes. The water background exhibiting similar trend has already been subtracted. Surprisingly, the ROA spectra (bottom part of Figure 1) provide quite distinct features also within this interval, and the experimental bands at 133, 168, and 289 cm^{-1} belong to the strongest ones in the entire ROA spectrum. The calculation reproduces them at 137, 147, and 289 cm^{-1} with the correct "+ − −" sign pattern. Visualization of the normal mode motion revealed that these bands are delocalized, mostly centered on the peptide main chain, but are difficult to assign to a particular chemical residue or local coordinate. Note that each visible band is composed of many individual transitions, as the density of vibrational states reaches ∼2.9 vibrational modes/cm^{-1} in this region.

Within the 300−900 cm^{-1} interval the density of vibrational states is relatively low, 2.1 modes/cm^{-1}. The simulation faithfully reproduces the most distinct spectral features such as the string S−S stretching (at 509/508 cm^{-1} in experiment/simulation) and tryptophan vibrations (762/772 and 860/867 cm^{-1}). The spectral intensities can thus be often related to local chemical coordinates. (Figure S5 and Tables S3 and S4 summarize these and other spectral features; see also refs 31−34 for vibrational assignment of protein ROA bands.) Typically, the strongest and relatively narrow Raman bands

DOI: 10.1021/acs.jpclett.5b01500
J. Phys. Chem. Lett. 2015, 6, 3314−3319

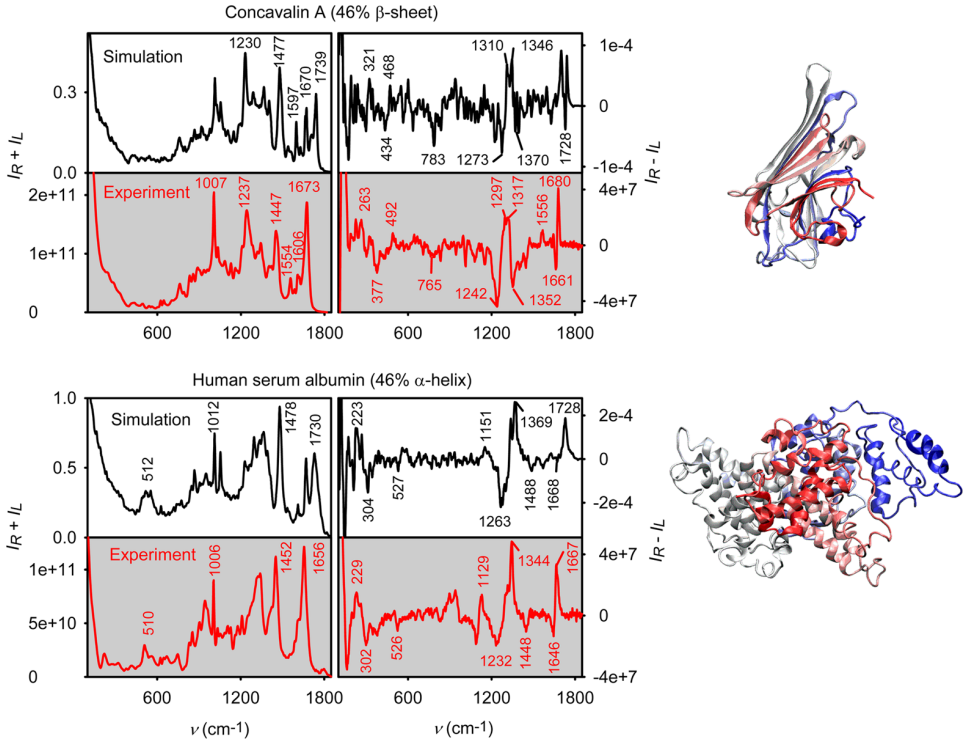


Figure 2. β sheet versus α-helical protein. Simulated and experimental Raman and ROA spectra of concavalin A and human serum albumin demonstrate sensitivity of both approaches in probing the peptide structure. For example, in concavalin, the 510 cm^{-1} Raman band of the disulfide bridges stabilizing the α-helical protein is missing, and the amide I frequency (1656 cm^{-1}) is very low due to the typical β-sheet interactions and band splitting. Even more differences may be found in the ROA spectra, as discussed in the main text.

originate in aromatic Trp, Phe, and Tyr amino acid residues; they exhibit a strong ROA signal as well. In ROA, however, the side chains' signals largely average out, and the resultant spectral shape is mainly due to the peptide main-chain contributions. In particular, in this wavenumber region the ROA intensities are rather weak, yet most of the experimental bands can be reproduced and assigned, as apparent from the expanded spectra (Figure S6). Within the 330−1130 cm^{-1} interval, for example, we counted about 23 Raman and ROA computed bands that probably match the experimental ones.

The 900−1400 cm^{-1} interval is a true "fingerprint" region with many overlapping vibrational bands and 3.3 vibrational modes/cm^{-1}; however, the overall character and the most distinct spectral features are reproduced as well, such as the Trp/Phe aromatic Raman signal around 1009 cm^{-1} and the Trp band at 1129 cm^{-1} (calculated at 1013 and 1140 cm^{-1}, respectively). This region comprises several distinct ROA bands, such as the negative signal at 980 cm^{-1} (largely peptide main chain deformation with a participation of C−C stretching reproduced at 963 cm^{-1}), the positive Trp band (exp./calc. 1013/987 cm^{-1}), and the broad negative lobe at 1250/1259 cm^{-1} of skeletal and amide vibrations. For vibrational spectroscopy of proteins, the so-called extended amide III spectral region (1280−1360 cm^{-1}) heavily involving $^{\alpha}$C−H bending is very important indeed.

The involvement of the peptide main chain groups in amide III vibrations and mode coupling makes ROA an excellent tool to monitor the secondary structure.[35] For example, α-helical structures exhibit two strong positive bands. The one at ∼1300 cm^{-1} was previously assigned to unhydrated canonical conformation and the 1340 cm^{-1} band attributed to a more open conformation stabilized by hydrogen bonding.[4] This assignment, based on empirical observation of α-helical peptides and proteins in polar and nonpolar solvents, did not explain all recorded data. In particular, poly(γ-benzyl-L-glutamate) in nonpolar $CHCl_3$ lacks the 1300 cm^{-1} "unhydrated" band. On the contrary, hydrated poly(L-lysine) or alanine-rich AK21 peptides do contain a strong 1300 cm^{-1} band.[4] Our computations suggest that these bands to a large extent originate in two $^{\alpha}$C−H bending modes in planes approximately parallel with and perpendicular to the $^{\alpha}$C−N bond (Figure S7). These are present in all conformers, and the intensity is varied through coupling to other vibrational modes.

Above 1400 cm^{-1} the vibrations become less dense again (2.3 modes/cm^{-1}) and can be assigned to local "chromophores" more easily, such as the C−H bending Raman signal at 1449/1479 cm^{-1} (exp./cal.), aromatic C═C stretching (1556/1600 cm^{-1} and 1622/1671 cm^{-1}), and, finally, the carbonyl "amide I" mode (largely C═O stretching, at 1661/1732 cm^{-1}). In the experimental ROA spectrum, the negative 1646 cm^{-1} peak is not well reproduced by the simulation, which in turn predicts a negative signal at 1558 cm^{-1}, with an experimental counterpart missing. The 1558 cm^{-1} band largely comes from induced optical activity of the tryptophan residues[36,37] and is thus very sensitive to conformational averaging. This and other minor inconsistencies can also be attributed to the anharmonicity and solvent effects that are only partially accounted for by the dielectric CPCM model and a limited precision of the DFT method causing improper mode ordering. In particular, the C═C aromatic stretching frequencies predicted by DFT are too high. The error of the CCT method, in particular the limited size of the fragments,[18,20] should also

DOI: 10.1021/acs.jpclett.5b01500
J. Phys. Chem. Lett. 2015, 6, 3314−3319

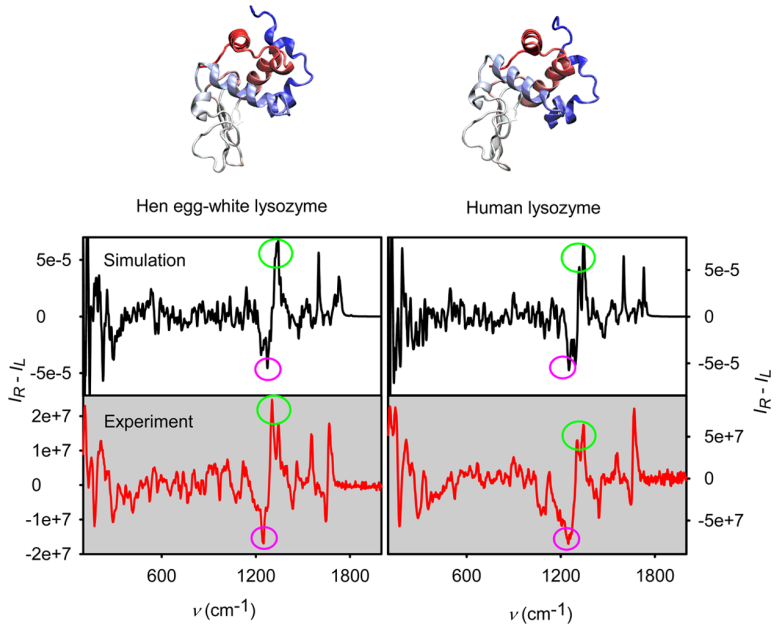


Figure 3. Reproducing fine structural differences. Calculated and experimental ROA spectra of hen and human lysozyme are compared. The two lysozymes have a very similar structure, and thus many differences are hidden in the experimental noise and computational error. However, some are detectable and can be reproduced. The encircled parts differ because of the different amounts of aromatic acids (Tyr, Phe) and their interaction with backbone and $^{\alpha}$C−H vibrations.

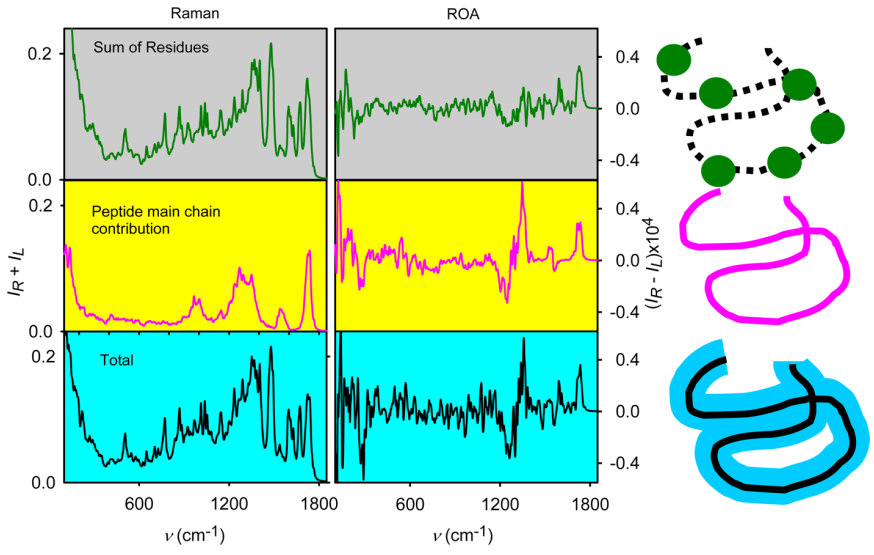


Figure 4. Contribution of protein parts to the Raman and ROA spectra. The Raman α-lactalbumin spectrum is approximately a sum of contribution of individual amino acids, whereas the ROA spectrum is dominated by the backbone signal sensitive to the secondary structure.

be considered, although it is supposed to be minor for irregular protein structures. Overall, however, the computation provides a solid ground for interpretation of the experimental data in this region. The ROA/Raman intensity ratio, often referred to as circular intensity difference (CID), is ~3 × 10^{-4}, and it is reproduced by the simulations as well.

Raman and ROA spectra of concavalin A and human serum albumin are presented in Figure 2, as examples of predominantly β-sheet and α-helical structures, respectively. The Raman concavalin spectrum, when compared with α-lactalbumin, primarily reflects a different amino acid composition. The most obvious example is the absence of the disulfide bridges of cysteine and their Raman signal at 510 cm^{-1}. The ROA spectrum is more sensitive to the secondary structure, and the amino acid sequence is not so important. Unlike for α-lactalbumin, concavalin provides a predominantly positive signal in the low-wavenumber region (below 300 cm^{-1}), and a new negative band appears at 377 cm^{-1}. In the extended amide III region, the −/+/+ α-lactalbumin feature at 1250/1301/1341 cm^{-1} becomes a "−/+/+/−" pattern at 1242/1297/1317/1352 cm^{-1}, which is well reproduced by the calculation. The negative CH bending band changes to a weak −/+ couplet (1447/1474 cm^{-1}), and amide I couplet becomes stronger (−/+ lobes at 1661/1680 cm^{-1}).

DOI: 10.1021/acs.jpclett.5b01500
J. Phys. Chem. Lett. 2015, 6, 3314−3319

Human serum albumin with 9161 atoms is the largest structure in the set. Unlike for concavalin, its Raman spectrum (lower part in Figure 2) does not differ that much from α-lactalbumin (Figure 1); however, the larger α-helical content is reflected in ROA. The "α-helical" features slightly change; for example, the lowest positive ROA signal (133 cm^{-1} for α-lactalbumin) moves to 107 cm^{-1}, the negative 168 and 289 cm^{-1} bands are mostly conserved (at 162 and 302 cm^{-1}), and the same applies to the broad negative lobe around 1250 cm^{-1}. A more dramatic change occurs in the extended amide III region, where the positive ROA signal at 1344 cm^{-1} sharpens and the intensity increases. As for α-lactalbumin, there is a negative band of the CH bending at 1448 cm^{-1} and the predominantly positive amide I signal (maximum at 1667 cm^{-1}) is accompanied by a negative lobe (1646 cm^{-1}), and these features are reproduced to a large extent by the simulation. The α-helix/β-sheet differences are obviously more substantial, as apparent from the direct comparison of the human serum albumin and concavalin in Figure 2.

Although the hen and human lysozymes have very similar structures, their ROA spectra differ (Figure 3). In general, they are similar to those of α-lactalbumin, for example, in the "+/−/−" bands at ∼111/172/289 cm^{-1} and the broad "−/+" couplet centered around 1270 cm^{-1}. However, the hen lysozyme exhibits much sharper negative ROA band at 1250 cm^{-1} and a positive one at 1304 cm^{-1} than the human one, which is at least qualitatively reproduced by the simulations. Corresponding Raman spectra are plotted in Figure S8. A dynamic visualization of the vibrational normal modes reveals that the intensity changes can be to a large extent attributed to the different amino acid content. For example, the human lysozyme has twice as many Tyr residues as hen lysozyme. The aromatic vibrational modes are coupled to the main chain $^{\alpha}C-H$ bending, which makes the band at 1341 cm^{-1} (experiment) stronger. Several differences are also apparent for bands around 150 and 250 cm^{-1} and within 1350−1800 cm^{-1}.

To investigate in detail the link between various protein parts and the resultant spectral response, we constructed arbitrary spectra where Raman and ROA intensity tensors[17] of selected atoms were considered only. This is illustrated in Figure 4, where α-lactalbumin spectra are simulated (i) as a sum of individual amino acid residues, (ii) considering the peptide backbone atoms only, and (iii) involving all atoms. Note that the result has to be interpreted carefully as the total spectrum is not a simple sum of atomic contribution. Nevertheless, this computational exercise well documents the different character of the Raman and ROA signal: the Raman spectrum is approximately a sum of individual amino acid contributions, whereas ROA is dominated by the backbone signal. The importance of the backbone for the ROA intensity was confirmed by an alternate model, an all-alanine peptide of the same secondary structure as α-lactalbumin (Figure S9). These computational models clearly showed that the tight link to the backbone structure lends the ROA spectroscopy increased sensitivity to the conformation. At the same time, the backbone vibrations are heavily mixed with those of the side chains, and only the full-molecule simulation can explain the experimental data.

During the reviewing stage of this work an interesting question appeared: To what extent is it possible to actually reconstruct complete protein 3D solution structure "backward" from the ROA spectrum? This clearly remains a difficult task for a low-resolution technique. Nevertheless, the results above indicate that a systematic improvement in this respect is possible. Perhaps future experimental and theoretical advances can make the process more routine, resembling the combination of molecular modeling and experimental data usual in NMR protein studies already. One can, for example, think about combining MD/ab initio computations with empirical techniques recognizing specific spectral patterns.[38]

So far we can conclude that due to the automatic fragmentation procedure an unprecedented precision in frequencies and spectral intensities of five globular was achieved, so far achievable for small molecules only. The quantum-chemical accuracy is critical for correct interpretation of the experimental Raman and Raman optical activity protein spectra. In particular, it enabled us to relate fine spectral patterns to amino acid composition, local structural features, and peptide secondary structure. Differences in ROA intensities well reflected the β-sheet/α-helical differences, and the double-degenerate $^{\alpha}C-H$ vibration was assigned to the important conformational marker band around 1320 cm^{-1}. The theoretical apparatus thus makes the Raman and ROA spectroscopic methods versatile tools to study protein structure, for example, enhancing applications in biology and medical imaging.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.5b01500.

Experimental and computational details documenting molecular properties and the spectral analysis. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bour@uochb.cas.cz.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The present study was supported by the Grant Agency (P208/11/0105 and 15-09072S) and Ministry of Education (LM2010005 and CZ.1.05/3.2.00/08.0144) of the Czech Republic. We thank Dr. R. Pelc for useful comments on the manuscript.

■ REFERENCES

(1) Freudiger, C. W.; Min, W.; Saar, B. G.; Lu, S.; Holtom, G. R.; He, C.; Tsai, J. C.; Kang, J. X.; Xie, X. S. Label-Free Biomedical Imaging with High Sensitivity by Stimulated Raman Scattering Microscopy. *Science* **2008**, *322*, 1857−1861.

(2) Palonpon, A. F.; Ando, J.; Yamakoshi, H.; Dodo, K.; Sodeoka, S.; Kawata, S.; Fujita, K. Raman and SERS Microscopy for Molecular Imaging of Live Cells. *Nat. Protoc.* **2013**, *8*, 677−692.

(3) Haesler, J.; Schindelholz, I.; Riguet, E.; Bochet, C. G.; Hug, W. Absolute Configuration of Chirally Deuterated Neopentane. *Nature* **2007**, *446*, 526−529.

(4) McColl, I. H.; Blanch, E. W.; Hecht, L.; Barron, L. D. A Study of α-Helix Hydration in Polypeptides, Proteins, and Viruses Using Vibrational Raman Optical Activity. *J. Am. Chem. Soc.* **2004**, *126*, 8181−8188.

(5) Blanch, E. W.; McColl, I. H.; Hecht, L.; Nielsen, K.; Barron, L. D. Structural Characterization of Proteins and Viruses Using Raman Optical Activity. *Vib. Spectrosc.* **2004**, *35*, 87−92.

DOI: 10.1021/acs.jpclett.5b01500
J. Phys. Chem. Lett. 2015, 6, 3314−3319

(6) Pasteur, L. *Théses De Chimie Et De Physique*; Bachelier: Paris, 1847.
(7) Bouř, P.; Keiderling, T. A. Ab Initio Simulation of the Vibrational Circular Dichroism of Coupled Peptides. *J. Am. Chem. Soc.* **1993**, *115*, 9602−9607.
(8) Kaminský, J.; Kubelka, J.; Bouř, P. Theoretical Modeling of Peptide a-Helical Circular Dichroism in Aqueous Solution. *J. Phys. Chem. A.* **2011**, *115*, 1724−1742.
(9) Barron, L. D.; Buckingham, A. D. Rayleigh and Raman Scattering from Optically Active Molecules. *Mol. Phys.* **1971**, *20*, 1111−1119.
(10) Polavarapu, P. L.; Nafie, L. Vibrational Optical Activity: Comparison of Theoretical and Experimental Results for (+)-(3R)-Methylcyclohexanone. *J. Chem. Phys.* **1980**, *73*, 1567−1575.
(11) Helgaker, T.; Ruud, K.; Bak, K. L.; Jorgensen, P.; Olsen, J. Vibrational Raman Optical Activity Calculations Using London Atomic Orbitals. *Faraday Discuss.* **1994**, *99*, 165−180.
(12) Ruud, K.; Helgaker, T.; Bouř, P. Gauge-Origin Independent Density-Functional Theory Calculations of Vibrational Raman Optical Activity. *J. Phys. Chem. A* **2002**, *106*, 7448−7455.
(13) Schwabe, T.; Grimme, S. Double-Hybrid Density Functionals with Long-Range Dispersion Corrections: Higher Accuracy and Extended Applicability. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397−3406.
(14) Hudecová, J.; Profant, V.; Novotná, P.; Baumruk, V.; Urbanová, M.; Bouř, P. CF Stretching Region: Computational Modeling of Vibrational Optical Activity. *J. Chem. Theory Comput.* **2013**, *9*, 3096−3108.
(15) Ruud, K.; Thorvaldsen, J. Theoretical Approaches to the Calculation of Raman Optical Activity Spectra. *Chirality* **2009**, *21*, E54−E67.
(16) Bouř, P.; Sopková, J.; Bednárová, L.; Maloň, P.; Keiderling, T. A. Transfer of Molecular Property Tensors in Cartesian Coordinates: A New Algorithm for Simulation of Vibrational Spectra. *J. Comput. Chem.* **1997**, *18*, 646−659.
(17) Barron, L. D. *Molecular Light Scattering and Optical Activity*; Cambridge University Press: Cambridge, U.K., 2004.
(18) Bieler, N. S.; Haag, M. P.; Jacob, C. R.; Reiher, M. Analysis of the Cartesian Tensor Transfer Method for Calculating Vibrational Spectra of Polypeptides. *J. Chem. Theory Comput.* **2011**, *7*, 1867−1881.
(19) Yamamoto, S.; Bouř, P. On the Limited Precision of Transfer of Molecular Optical Activity Tensors. *Collect. Czech. Chem. Commun.* **2011**, *76*, 567−583.
(20) Yamamoto, S.; Li, X.; Ruud, K.; Bouř, P. Transferability of Various Molecular Property Tensors in Vibrational Spectroscopy. *J. Chem. Theory Comput.* **2012**, *8*, 977−985.
(21) Barron, L. D.; Hecht, L.; Blanch, E. W.; Bell, A. F. Solution Structure and Dynamics of Biomolecules from Raman Optical Activity. *Prog. Biophys. Mol. Biol.* **2000**, *73*, 1−49.
(22) Luber, S.; Reiher, M. Theoretical Raman Optical Activity Study of the B Domain of Rat Metallothionein. *J. Phys. Chem. B* **2010**, *114*, 1057−1063.
(23) Yamamoto, S.; Straka, M.; Watarai, H.; Bouř, P. Formation and Structure of the Potassium Complex of Valinomycin in Solution Studied by Raman Optical Activity Spectroscopy. *Phys. Chem. Chem. Phys.* **2010**, *12*, 11021−11032.
(24) Yamamoto, S.; Watarai, H.; Bouř, P. Monitoring the Backbone Conformation of Valinomycin by Raman Optical Activity. *ChemPhysChem* **2011**, *12*, 1509−1518.
(25) Yamamoto, S.; Kaminský, J.; Bouř, P. Structure and Vibrational Motion of Insulin from Raman Optical Activity Spectra. *Anal. Chem.* **2012**, *84*, 2440−2451.
(26) Bouř, P.; Keiderling, T. A. Partial Optimization of Molecular Geometry in Normal Coordinates and Use as a Tool for Simulation of Vibrational Spectra. *J. Chem. Phys.* **2002**, *117*, 4126−4132.
(27) Perdew, J. P.; Burke, K.; Wang, Y. Generalized Gradient Approximation for the Exchange-Correlation Hole of a Many-Electron System. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 16533−16539.
(28) Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. C. W. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074−5085.
(29) Mennucci, B.; Cappelli, C.; Cammi, R.; Tomasi, J. Modeling Solvent Effects on Chiroptical Properties. *Chirality* **2011**, *23*, 717−729.
(30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*; Revision D01 ed.; Gaussian, Inc.: Wallingford, CT, 2009.
(31) Jacob, C. R.; Luber, S.; Reiher, M. Understanding the Signatures of Secondary-Structure Elements in Proteins with Raman Optical Activity Spectroscopy. *Chem. - Eur. J.* **2009**, *15*, 13491−13508.
(32) Jacob, C. R.; Luber, S.; Reiher, M. Analysis of Secondary Structure Effects on the IR and Raman Spectra of Polypeptides in Terms of Localized Vibrations. *J. Phys. Chem. B* **2009**, *113*, 6558−6573.
(33) Weymuth, T.; Jacob, C. R.; Reiher, M. Identifying Protein β-Turns with Vibrational Raman Optical Activity. *ChemPhysChem* **2011**, *12*, 1165−1175.
(34) Weymuth, T.; Reiher, M. Characteristic Raman Optical Activity Signatures of Protein β-Sheets. *J. Phys. Chem. B* **2013**, *117*, 11943−11953.
(35) Weymuth, T.; Jacob, C. R.; Reiher, M. A Local-Mode Model for Understanding the Dependence of the Extended Amide III Vibrations on Protein Secondary Structure. *J. Phys. Chem. B* **2010**, *114*, 10649−10660.
(36) Jacob, C. R.; Luber, S.; Reiher, M. Calculated Raman Optical Activity Signatures of Tryptophan Side Chains. *ChemPhysChem* **2008**, *9*, 2177−2180.
(37) Hudecová, J.; Horníček, J.; Buděšínský, M.; Šebestík, J.; Šafařík, M.; Zhang, G.; Keiderling, T. A.; Bouř, P. Three Types of Induced Tryptophan Optical Activity Compared in Model Dipeptides: Theory and Experiment. *ChemPhysChem* **2012**, *13*, 2748−2760.
(38) Zhu, F.; Tranter, G. E.; Isaacs, N. W.; Hecht, L.; Barron, L. D. Delineation of Protein Structure Classes from Multivariate Analysis of Protein Raman Optical Activity Data. *J. Mol. Biol.* **2006**, *363*, 19−26.

DOI: 10.1021/acs.jpclett.5b01500
J. Phys. Chem. Lett. 2015, 6, 3314−3319

Supporting Information

The Journal of Physical Chemistry Letters

First Principles Predictions of Vibrational Raman Optical Activity of Globular Proteins

Jiří Kessler, Josef Kapitán, and Petr Bouř

Contents

Methods

Experimental Spectra. Raman and ROA spectra of five proteins (concanavalin-A, human lysozyme, hen egg-white lysozyme, bovine α-lactalbumin and human serum albumin) were measured with a backscattering scattered circular polarization (SCP) BioTools μ-ChiralRAMAN-2X instrument based on the design of W. Hug[1,2] and equipped with a diode-pumped solid-state laser operating at 532 nm. Proteins were dissolved to concentrations of 50-160 mg/mL in water, except for concanavalin-A dissolved in water solution of NaCl (2.5 mol/kg). Laser power at the sample was 330-560 mW, and acquisition time was 12-33 hours. All samples were measured in 4×3 mm fused silica rectangular cell at room temperature (293 K). For measurement in the low-frequency region the original holographic notch filter (Kaiser, cutoff at 200 cm^{-1}) was replaced by an edge filter (Semrock) working reliably down to 80 cm^{-1} for Raman and to 90 cm^{-1} for ROA spectra. Raman spectrum of the cell filled with the solvent was subtracted from all Raman spectra. Baseline in ROA spectra was corrected only for human lysozyme and human serum albumin that were weakly absorbing and contained some fluorescent impurities. The frequency was calibrated using a neon lamp. Intensity was corrected using a standard fluorescence reference material, (SRM 2242, obtained from National Institute of Standards and Technology, USA) and tungsten halogen radiometric calibration source (HL-3plus-CAL, OceanOptics), equipped with a cosine corrector (CC-3) and a multimode fiber (diameter 50 μm and numerical aperture 0.22; Thorlabs, M16L01). Two sources had to be used to achieve sufficient (better than 10%) accuracy as the fluorescence calibration standard is not certified for wavenumbers below 150 cm^{-1} and the numerical aperture of the tungsten halogen source was smaller than the one actually used in the ROA spectrometer (about 0.4). The intensity calibration proved to be essential as a significant intensity drop is observed in both low and high wavenumber spectral regions, due to a lower efficiency of holographic diffraction grating. In addition, non-uniform quasi-periodic transmission variability of the edge filter was observed, especially in the low wavenumber region.

Computations. X-ray data from the protein data bank database were used as starting geometries to generate the theoretical spectra. The raw geometries were refined using the *tleap* module of the Amber program suite,[3] e.g. missing heavy and hydrogen atoms were added. The peptide chains of the proteins were split into overlapping fragments, each containing four amide (three amino-acid) residues and capped with methyl groups as CH_3-NHCO- and –NHCO-CH_3 at the N- and C- peptide chain, respectively (**Figure S1**). Additional fragments were created for the disulfide bridges. For the lysozymes, "contact fragments" were created for all pairs of close (< 2.4 Å) but not covalently bonded aminoacid side chains. These fragments contained the two aminacid residues, and the N- and C-amide groups were capped in the same way as for the four-aminoacid fragments. The contact fragments had only a minor effect on the spectra (**Figure S2**) and thus they were not used in other simulations. Our own software ("pdb_dc.f") was employed for the automatic fragmentation based on the covalent bond table.

The geometry of each fragment was partially optimized by energy-minimization at the B3PW91[4]/6-31++G** level; the protein and water environment was simulated using the polarized continuum model (CPCM),[5,6] with water solvent parameters set to values implemented in the Gaussian 09 (version D01) program.[7] The B3PW91 functional previously provided good results for similar systems.[8] The normal mode optimization (NMO)[9] was used for constrained optimization fixing modes bellow a certain limit (ω_0, typically 100-300 cm^{-1}). The NMO optimization introduced only minor perturbations of fragment geometries with respect to the protein target, while the most important vibrational modes were fully relaxed. Then the atomic polar, atomic axial, Raman ROA, and harmonic force field tensors[10,11] were transferred from all fragments to the original protein using the Cartesian coordinate transfer (CCT) procedure in detail described elsewhere.[12,13] Harmonic vibrational frequencies and Raman and ROA intensities of the proteins were generated using the usual procedure.[11,14,15] To generate realistic spectral shapes (S) the backscattered intensities (I_{180}) were convoluted with Lorentzian functions and corrected to temperature as

$$S(\omega) = \frac{I_{180}}{\omega_i\{1-\exp[-\omega_i/(kT)]\}\{4[(\omega-\omega_i)/\Delta]^2+1\}},$$

where T = 293 K, k is the Boltzmann constant, ω_i is vibrational frequency, and the bandwidth Δ = 10 cm^{-1}. As absolute Raman scattering intensities cannot be easily measured arbitrary ("atomic") units were used for the simulation, and the instrumental counts (proportional to the number of electrons detected by the charged-coupled device detector) for the experiment.

Alternatively to the X-ray geometries, molecular dynamics (MD) was run using the Amber program[3] and Amber 2004 force field,[16] to account for the effects of relaxation and dynamics. The proteins were surrounded by water in a cubic cage so that at least 7 Å gap remained between the protein surface and the cage walls. This resulted in the following box lengths: α-lactalbumin 64 Å, human lysozyme 58 Å, hen egg-white lysozyme 58Å, concanavalin A 68 Å, and human serum albumin 90 Å. Na^+ or Cl^- counterions were added to neutralize the protein charges. Concavalin A contains metal ions (Mn^{2+} and Ca^{2+}) in the native structure. They were kept in the MD modeling, but neglected for the spectral generation. The dynamics was performed using the *nVT* ensemble, 293 K, and 1 fs integration time. The *npT* ensemble-based simulation (1 atm) was also tried but no significant difference in geometries was encountered. After an equilibration run (1 ns) geometry snapshots taken every 10 ps were saved for 10 ns. The average solute density was created and the snapshot most closely matching it was selected using an algorithm described elsewhere.[17] This "average" MD geometry was then used to generate the Raman and ROA spectra in the same way as for the X-ray structures. Finally, the force fields and ROA tensors thus obtained were transferred to 1000 snapshots and the resultant spectra were averaged. Human lysozyme Raman and ROA spectra obtained by the three geometry models are compared in **Figure S3**. Although the direct snapshot averaging is conceptually the most advanced method, the improvement in the spectra is not very convincing, and the X-ray based modeling was thus used by default.

The limited relaxation of the fragments had a greater effect on the spectra, which is shown in **Figure S4**, where the frequency cutoff (ω_0, cf. Ref. 18) of 100 cm^{-1} produced a slightly more realistic human lysozyme Raman and ROA shapes than ω_0 set to 300 cm^{-1}. The normal mode vibrational assignment and analysis was enhanced by computation of the potential energy distribution (PED)[19] in intrinsic molecular coordinates. As it was not feasible to manually define thousands of coordinates for a non-redundant set, all possible coordinates (bond length, bond and torsional angles) were defined using an automated procedure, and a simplified diagonal internal-coordinate force field F_I was used for the analysis, as described in Ref. 20. The force field was obtained by minimizing $(B^t F_I B - F_C)^2$, where F_C is the Cartesian force field obtained by the transfer, B is the Cartesian-internal transformation matrix, and index ^{t} denotes transposed matrix. Note that while this procedure leads to realistic frequencies and assignments, the internal force field is not good enough for spectra generation. Six examples of PED thus obtained are plotted in **Figure S5**.

(1) Hug, W.; Hangartner, G. A Novel High-Throughput Raman Spectrometer for Polarization Difference Measurements. *J. Raman Spectrosc.* **1999**, *30*, 841-852.

(2) Hug, W. Virtual Enantiomers as the Solutuion of Optical Activity's Deterministic Offset Problem. *Appl. Spectrosc.* **2003**, *57*, 1-13.

(3) Case, D. A.; Cheatham, I. T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, J. K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber Biomolecular Simulation Programs *J. Comput. Chem.* **2005**, *26*, 1668-1688.

(4) Perdew, J. P.; Burke, K.; Wang, Y. Generalized Gradient Approximation for the Exchange-Correlation Hole of a Many-Electron System. *Phys. Rev. B* **1996**, *54*, 16533-16539.

(5) Klamt, A.; Jonas, V.; Burger, T.; Lohrentz, J. C. W. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074-5085.

(6) Mennucci, B.; Cappelli, C.; Cammi, R.; Tomasi, J. Modeling Solvent Effects on Chiroptical Properties. *Chirality* **2011**, *23*, 717-729.

(7) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09; Revision D01 ed.; Gaussian, Inc.: Wallingford CT, 2009.

(8) Yamamoto, S.; Kaminský, J.; Bouř, P. Structure and Vibrational Motion of Insulin from Raman Optical Activity Spectra. *Anal. Chem.* **2012**, *84*, 2440-2451.

(9) Bouř, P.; Keiderling, T. A. Partial Optimization of Molecular Geometry in Normal Coordinates and Use as a Tool for Simulation of Vibrational Spectra. *J. Chem. Phys.* **2002**, *117*, 4126-4132.

(10) Devlin, F. J.; Stephens, P. J.; Cheeseman, J. R.; Frisch, M. J. Ab Initio Prediction of Vibrational Absorption and Circular Dichroism Spectra of Chiral Natural Products Using Density Functional Theory - Alpha-Pinene. *J. Phys. Chem.* **1997**, *101*, 9912-9924.

(11) Barron, L. D. *Molecular Light Scattering and Optical Activity*; Cambridge University Press: Cambridge, 2004.

(12) Bouř, P.; Sopková, J.; Bednárová, L.; Maloň, P.; Keiderling, T. A. Transfer of Molecular Property Tensors in Cartesian Coordinates: A New Algorithm for Simulation of Vibrational Spectra. *J. Comput. Chem.* **1997**, *18*, 646-659.

(13) Yamamoto, S.; Li, X.; Ruud, K.; Bouř, P. Transferability of Various Molecular Property Tensors in Vibrational Spectroscopy. *J. Chem. Theory Comput.* **2012**, *8*, 977-985.

(14) Polavarapu, P. L. Ab Initio Vibrational Raman and Raman Optical Activity Spectra. *J. Phys. Chem.* **1990**, *94*, 8106-8112.

(15) Nafie, L. *Vibrational Optical Activity: Principles and Applications*; Wiley: Chichester, 2011.

(16) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(17) Kessler, J.; Dračínský, M.; Bouř, P. Parallel Variable Selection of Molecular Dynamics Clusters as a Tool for Calculation of Spectroscopic Properties. *J. Comput. Chem.* **2013**, *34*, 366-371.

(18) Hudecová, J.; Hopmann, K. H.; Bouř, P. Correction of Vibrational Broadening in Molecular Dynamics Clusters with the Normal Mode Optimization Method. *J. Phys. Chem. B* **2012**, *116*, 336-342.

(19) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; Dover: New York, 1980.

(20) Benda, L.; Straka, M.; Sychrovský, V.; Bouř, P.; Tanaka, Y. Detection of Mercury−Tpt Dinucleotide Binding by Raman Spectra: A Computational Study. *J. Phys. Chem. A.* **2012**, *116*, 8313-8320.

(21) Pike, A. C.; Brew, K.; Acharya, K. R. Crystal Structures of Guinea-Pig, Goat and Bovine Alpha-Lactalbumin Highlight the Enhanced Conformational Flexibility of Regions That Are Significant for Its Action in Lactose Synthase. *Structure* **1996**, *4*, 691-703.

(22) Deacon, A.; Gleichmann, T.; Kalb, A. J.; Price, H.; Raftery, J.; Bradbrook, G.; Yarivc, J.; Helliwella, J. R. The Structure of Concanavalin a and Its Bound Solvent Determined

with Small-Molecule Accuracy at 0.94 Å Resolution. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 4305-4312.

(23) Mizutani, R.; Shimizu, Y.; Saiga, R.; Ueno, G.; Nakamura, Y.; Takeuchi, A.; Uesugi, K.; Suzuki, Y. Spatiotemporal Development of Soaked Protein Crystal. *Sci. Rep.* **2014**, *4*, 5731.

(24) Artymiuk, P. J.; Blake, C. C. Refinement of Human Lysozyme at 1.5 Å Resolution Analysis of Non-Bonded and Hydrogen-Bond Interactions. *J. Mol. Biol.* **1981**, *152*, 737-762.

(25) Min, X.; Carter, D. C. Atomic Structure and Chemistry of Human Serum Albumin. *Nature* **1992**, *358*, 209-215.

Table S1. Selected data on the five proteins subjected to ROA experiment and computations.[a]

Structure	ID	Ref.	Resolution (Å)	N_{res}	N_{at}
bovine α-lactalbumin	1HFZ	[21]	2.30 Å	124	1960
concanavalin-A	1NLS	[22]	0.94 A	237	3566
hen egg-white lysozyme	3WPJ	[23]	2.00 Å	129	1960
human lysozyme	1LZ1	[24]	1.50 Å	130	2021
human serum albumin	1UOR	[25]	2.80 Å	585	9161

Structure	Sample	*P* (mW)	*t* (hours)	*c* (mg/mL)	Solvent
bovine α-lactalbumin	L5385	500	13.3	100	water
concanavalin-A	C2010	560	32.2	50	2.5 mol/kg NaCl in water
hen egg-white lysozyme	L6876	330	11.9	100	water
human lysozyme	L1667	350	27.9	160	water
human serum albumin	A3782	350	12.1	50	water

[a] The identification number (ID) refers to the protein data bank database (PDB, http://www.rcsb.org); this geometry taken as a starting point in the modeling. The number of aminoacid residues (N_{res}) and the number of atoms (N_{at}) are related to the computational model, without solvent (H_2O) molecules. Sigma-Aldrich catalogue number ("Sample"), laser power at the sample (*P*), spectra accumulation time (*t*), concentration (*c*) and the solvent are specified in the bottom half of the table.

Table S2. Secondary structure contents (%) in the five proteins, according to the Protein Data Bank http://www.rcsb.org.

Structure	ID	α-helix	β-sheet
bovine α-lactalbumin	1HFZ	43	9
concanavalin-A	1NLS	3	46
hen egg-white lysozyme	3WPJ	41	10
human lysozyme	1LZ1	39	12
human serum albumin	1UOR	46	0

Table S3. Frequencies (cm^{-1}) of Characteristic Raman Bands

Assignment	Bovine α-lactalbumin	Concanavalin-A	Hen egg-white lysozyme	Human lysozyme	Human serum albumin	Calc[a]
amide I (ν(C=O))	1661	1673	1661	1661	1656	1732
Tyr ν(C=C)	1622	1617	1622	1620	1620	1671
Trp ν(C=C)	1556	1554	1555	1557	-	1600
ν(C-H)	1449	1453	1451	1453	1451	1479
Trp δ(C-H)	1362	1361[b]	1361	1359	-	1404
Tyr δ(C-H)	1338	1343	1338	1338	1339	1343
Tyr δ(C-H)	1211	1212	1209[b]	1208	1210	1234
δ(C-H)	1129	1129	1127	1127	1127	1140
Trp deformation	1034	1034	1031[b]	1031[b]	1033	1038
Trp deformation	1013[b]	-	1011	1011	-	1013
Phe deformation	1009	1007	1011	1011	1005	1012
Tyr deformation	860	858	855	856	853	867
Trp oop[c]	838	833	838	832	833	844
Trp+backbone def.	762	760	763	759	-	772
ν(S-S)	509	-	509	512	509	508

[a] for bovine α-lactalbumin
[b] shoulder in the spectrum
[c] out of plane

Table S4. Frequencies (cm^{-1}), Signs and Assignments of Selected ROA Bands

Assignment	Bovine α-lactalbumin	Concanavalin-A	Hen egg-white lysozyme	Human lysozyme	Human serum albumin	Calc[a]
amide I (ν(C=O))	1669+	1680+	1669+	1666+	1667+	1732
arom. ν(C=C)	1646-	1616-	1646-	1648-	1646-	
Trp ν(C=C)	1557+	1556+	1555+	1558+	-	1590
Phe δ(C-H)	1443-	1447-	1436-	1443-	1449-	1481
	-	-	-	1414-	1421+	
δ($^{\alpha}$C-H)	1341+	1352-	1345+	1346+	1345+	1357
δ($^{\alpha}$C-H), Tyr δ(C-H)	1301+	1296+	1304+	1305+	1311+	1311
δ($^{\alpha}$C -H)	1250-	1242-	1250-	1250-	-	1259
lysin δ(C-H)	1227-				1231-	1241
	1136+	1152-	1137+	-	1126+	1151
	1118+	-	1120+	1128+		
	1099-	-	1100-	1103-	1088-	
	1069-	1081-	1075-	1075-	1066-	
	1013+	1013-	-	-	1015+	
	980-	-	970-	966-	980-	963
	-	66+	679+	670+	679+	
Pro and main chain deformation	289-	302-	289-	289-	-	290
NH oop, CH_3 rotation	168-	-	172-	172-	168-	150
	133+	133+	135-	-	-	
	101+	105-	111+	112+	97+	

[a] for bovine α-lactalbumin

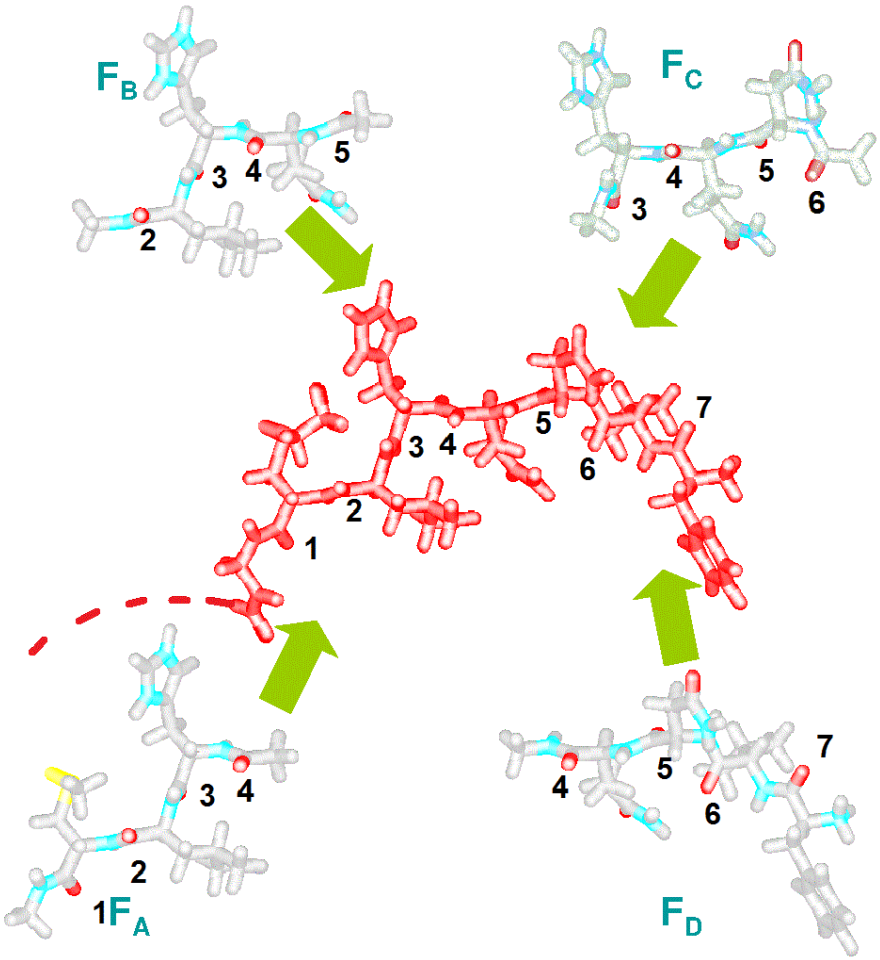


Figure S1. Example of fragmentation and transfer of vibrational spectroscopic properties. The peptide main chain (red, amide groups are temporarily numbered 1…7) is divided into fragments (F_A … F_D), containing four amide residues, shifted relative to each other by a one-amide increment along the chain. Up to *N* … *N*+3 amide-amide interactions (force field elements) can be reconstructed from the fragments. Additionally, contact fragments between close, but not covalently bonded aminoacid residues were included (not shown). Note that the force field elements were transferred (rotated) for each atom pair separately, to better account for possible small geometry differences between the peptide and the fragments.

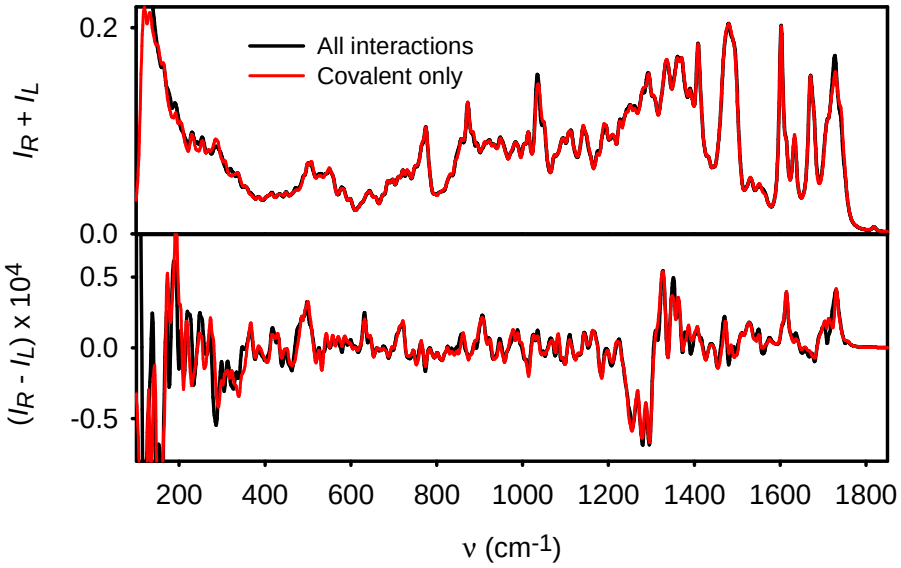


Figure S2. Effect of non-covalent interactions on the spectra. Simulated Raman and ROA spectra of α-lactalbumin. When fragments comprising non-covalent interactions are added to the fragments obtained by splitting the main peptide chain the spectra exhibit only minor changes.

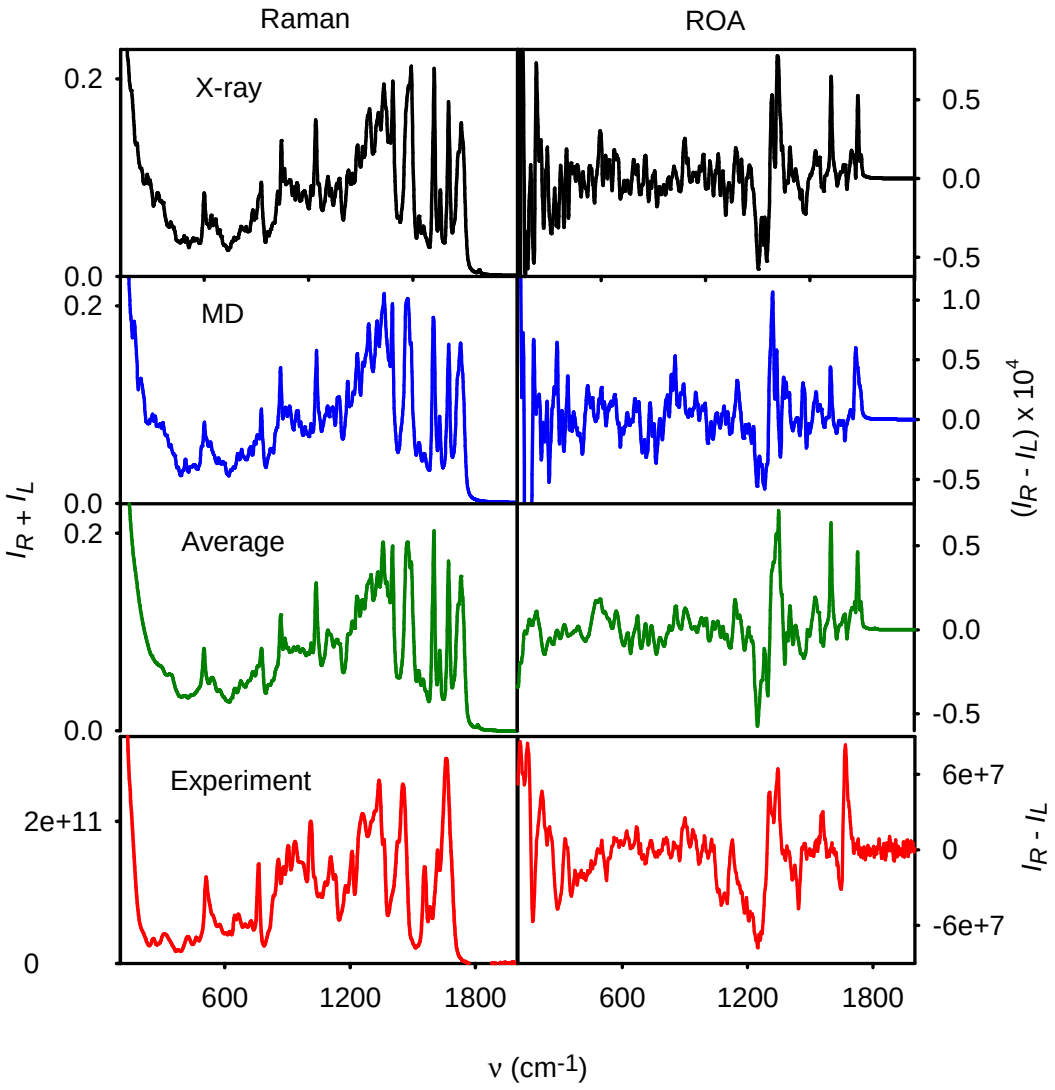


Figure S3. Variations of the spectra according to geometry. For human lysozyme the Raman and ROA spectra were generated using X-ray geometry, the average MD geometry, and averaging of 1000 MD geometries.

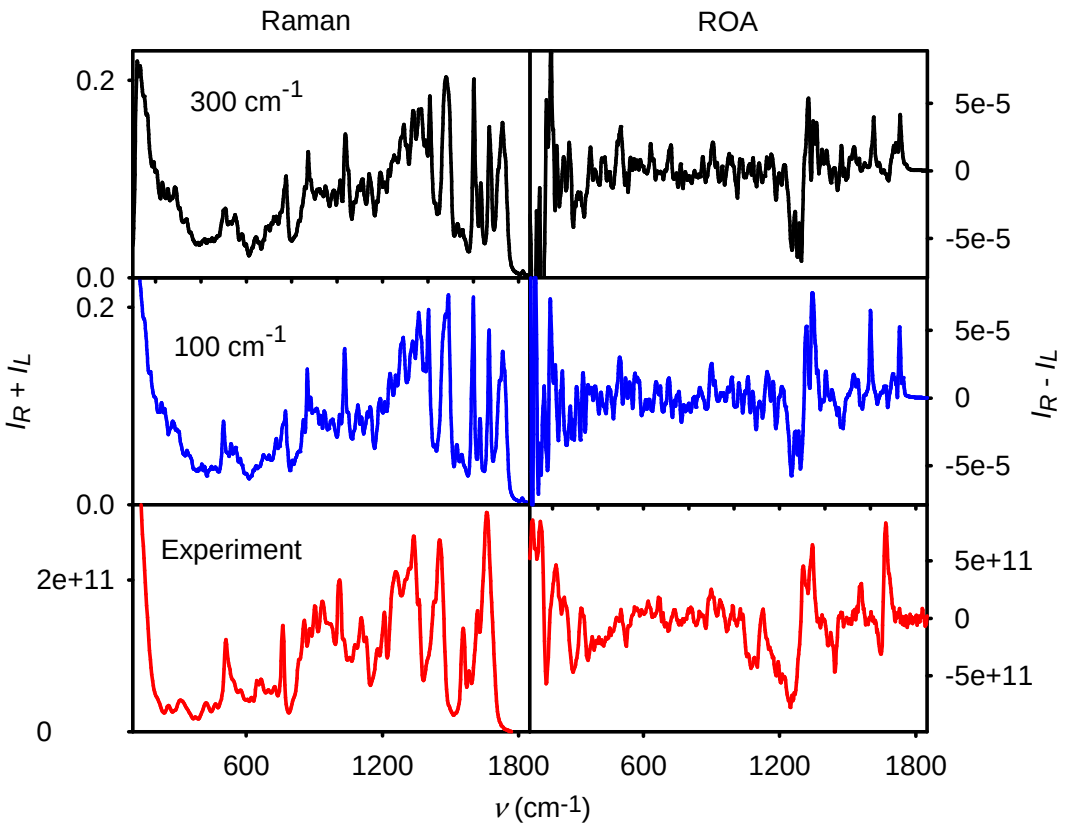


Figure S4. Effect of fragment relaxation on the spectra. Raman and ROA human lysozyme spectra were simulated from partially optimized fragments, using the 300 cm^{-1} and 100 cm^{-1} cutoffs for the fixed normal mode frequencies (cf. the parameter ω_0 in Ref. [18]). Compared to the experiment, the 100 cm^{-1} cutoff allowing for a more extensive relaxation of the X-ray initial fragment geometries seems to provide better spectral shapes.

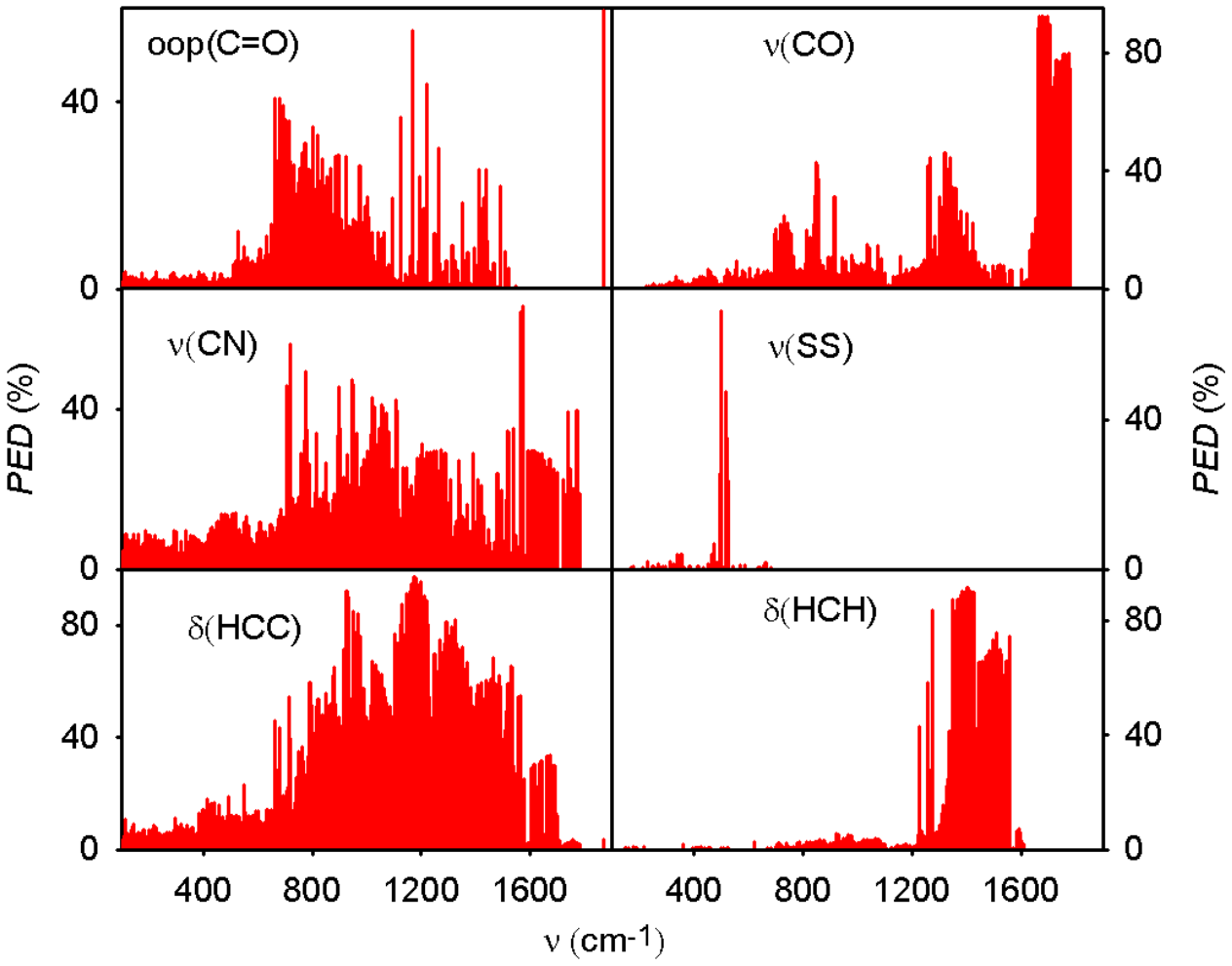


Figure S5. Potential energy distribution (*PED*) of selected internal coordinates. The *PEDs* provide an approximate link of the computed vibrational transitions to local "chemical" coordinates. The out of plane (oop) C=O wagging is mostly spread within the 700-1000 cm^{-1} interval, CO stretching (ν(CO)) involves rather narrow amide I C=O vibrations (calculated around 1700 cm^{-1}) as well as a more spread-out C-O (single bond) stretching at low wavenumbers, CN stretching (ν(CN)) is omnipresent, SS stretching (ν(SS)) is the most localized one on the frequency scale. The bending (δ) coordinates can be either spread (HCC) or localized (HCH) on the frequency scale.

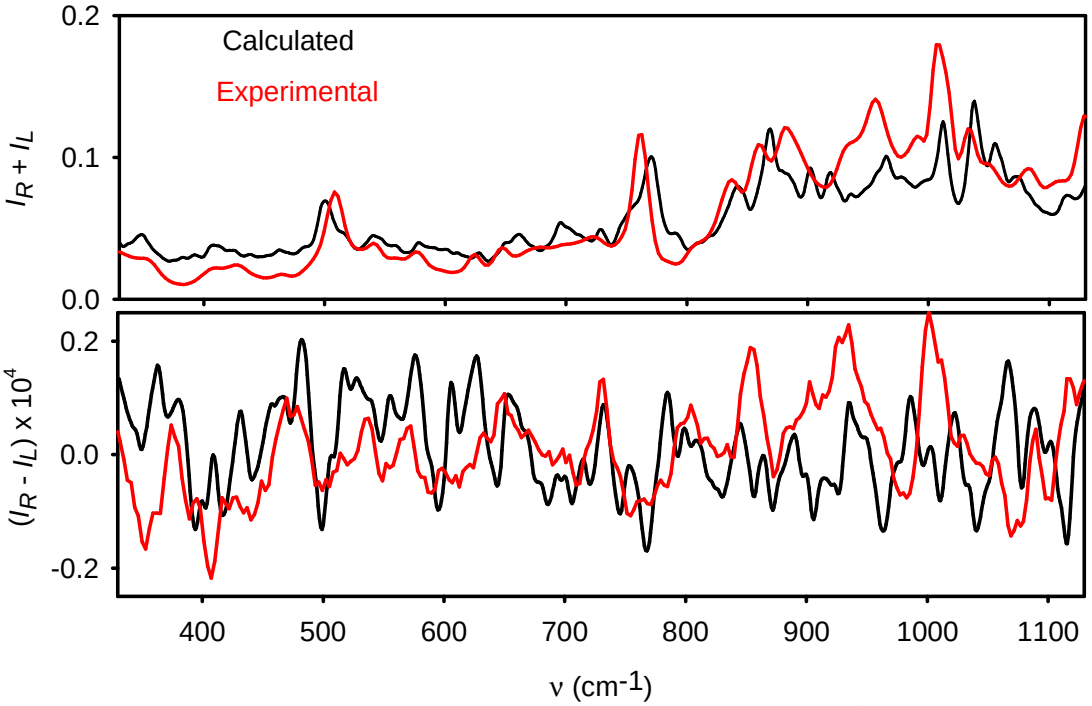


Figure S6. Fine Raman and ROA structure. Zoomed 330-1130 cm^{-1} region of the Raman and ROA bovine α-lactalbumin spectra reveals similarities and discrepancies between the experiment and calculations, not apparent from **Figure 1**.

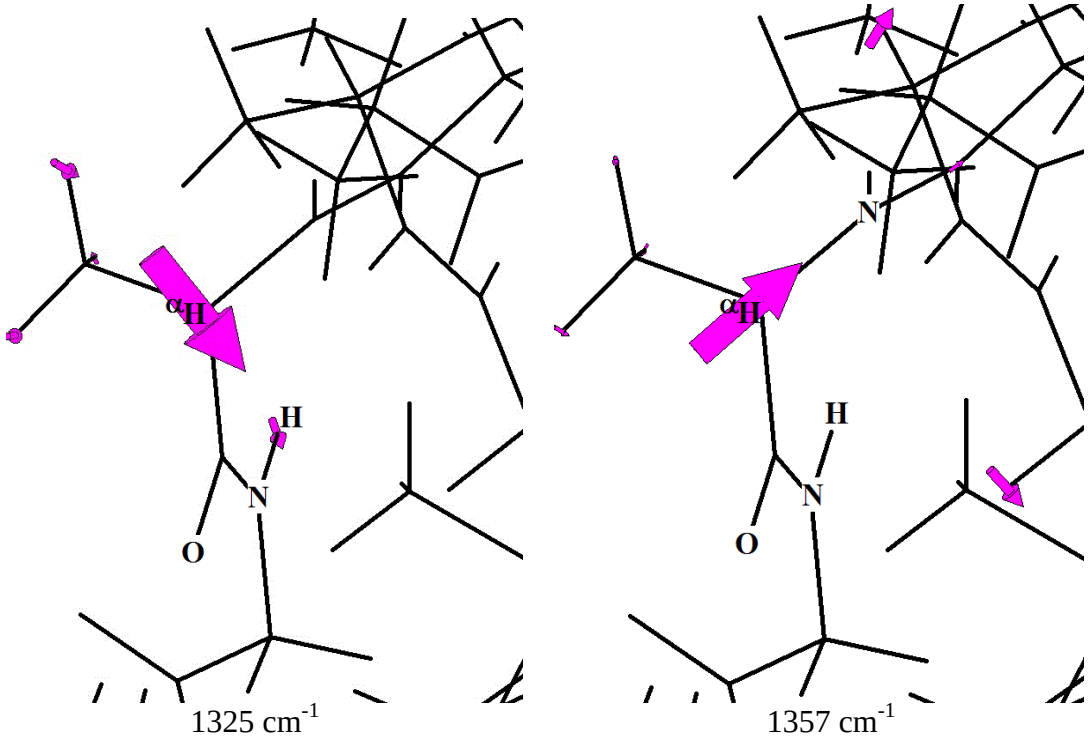


Figure S7. Example of two H-$^{\alpha}$C bending modes calculated in an all-alanine α-lactalbumin model. These modes are responsible for the two strongest ROA bands (experimentally at 1301 and 1341 cm^{-1}) in the extended amide III region.

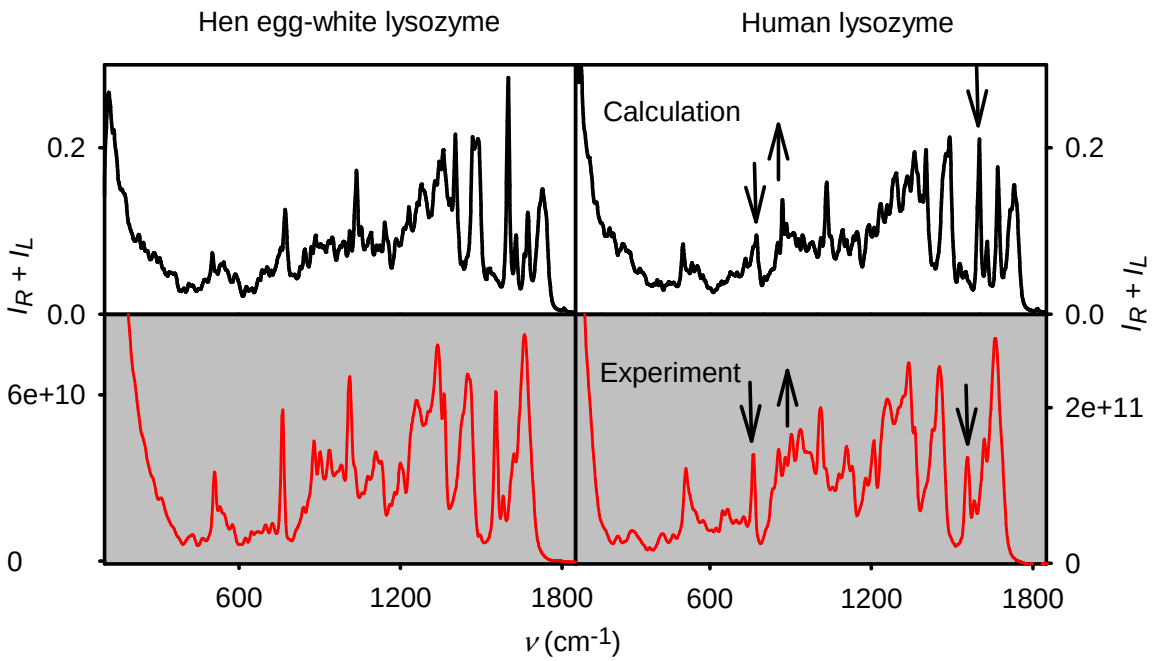


Figure S8. Calculated and experimental Raman spectra of hen and human lysozyme. The differences between the two lysozymes are in general at the edge of computational precision, but some variations of the Raman intensities are clearly reproduced. These are mostly due to the different contents of aromatic aminoacids.

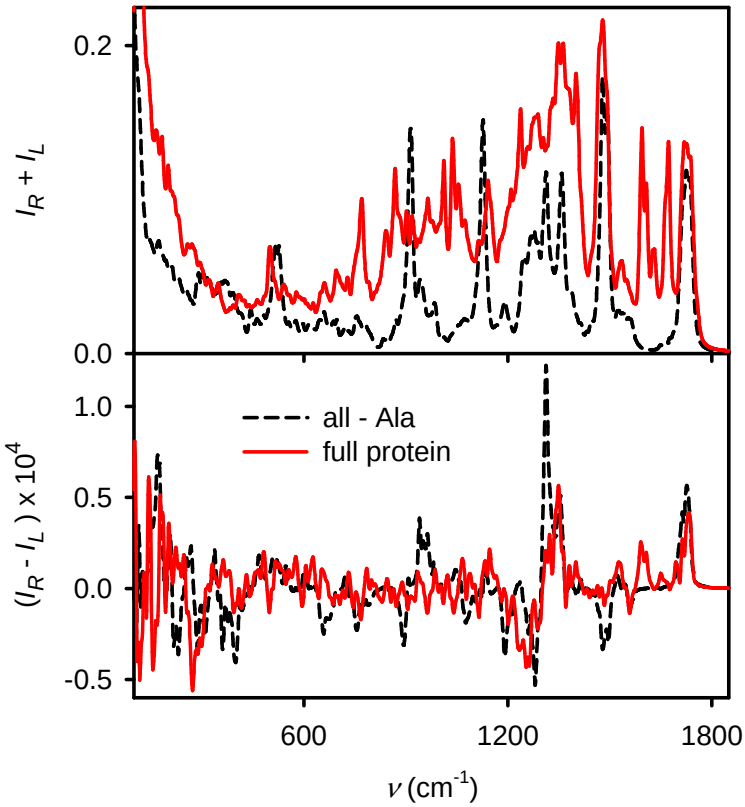


Figure S9. Contribution of the backbone to Raman and ROA intensities. The bovine α-lactalbumin spectra were generated for the full protein, and for the same secondary structure, where all aminoacid side chains were replaced by methyl groups. Apparently, the side chains contribute a lot to the Raman intensities, whereas ROA is dominated by the backbone vibrations.

9.6 Appendix F – *J. Comput. Chem.* 2014, 35, 1552.

Molecular Dynamics with Helical Periodic Boundary Conditions

Jiří Kessler[a,b] and Petr Bouř*[a]

Helical symmetry is often encountered in nature and thus also in molecular dynamics (MD) simulations. In many cases, an approximation based on infinite helical periodicity can save a significant amount of computer time. However, standard simulations with the usual periodic boundary conditions (PBC) are not easily compatible with it. In the present study, we propose and investigate an algorithm comprising infinitely propagated helicity, which is compatible with commonly used MD software. The helical twist is introduced as a parametric geometry constraint, and the translational PBC are modified to allow for the helical symmetry via a transitional solvent volume. The algorithm including a parallel code was implemented within the Tinker software. The viability of the helical periodic boundary conditions (HPBC) was verified in test simulations including α-helical and polyproline II like peptide structures. For an insulin-based model, the HPBC dynamics made it possible to simulate a fibrillar structure, otherwise not stable within PBC.

DOI: 10.1002/jcc.23653

Introduction

Being a commonplace in living systems, helical symmetry is often encountered also in their molecular dynamics (MD) simulations. Typical objects involve not only simple helical biopolymers (peptide helices, nucleic acids), but also larger and complex molecular assemblies, such as multithread muscle filaments,[1] amyloid-β fibrils forming Alzheimer's plaques,[2] or tubular networks supporting the cytoskeleton.[3]

Possible ways of tackling such nonperiodic systems include extensive solvation shells,[4] active site solvation,[5] or a hybrid explicit/implicit solvation model.[6] The shell models, such as the primary hydration shell method,[7] are often used when the standard periodic boundary conditions (PBC) are computationally very demanding. The translational periodicity is often desirable, for example, when the lattice–sum interactions are used and the periodicity elegantly increases the effective solvent volume seen by the molecule under investigation. It is compatible with parallelization of the computer code for short-range potentials,[8] for example, using distributions on graphical cards,[9] or potentially makes it amenable to analytical Ewald-like or other pairwise electrostatic summation techniques.[10]

As we are not aware of any MD software directly allowing for the helical periodicity, we investigate here a new algorithm suited to both infinite covalent systems and periodic molecular complexes. It is compatible with commonly used MD procedures and force fields. We implemented and tested it within the Tinker MD package[11] and Amber99[12] force field. The helicity is introduced for a part of the simulated system, which requires minor adjustments of the MD propagation algorithm only.

In this article, we define the helical periodic conditions, discuss the necessary modification of the propagation algorithm, and details of the implementation. For the example of α- and polyproline II helices, we compare the results with the usual PBC computations not using the helical symmetry. As the most applied example, simulation of a fibrillar system based on the insulin molecule is presented. The possibility to determine the overall twist on the run during the MD propagation is discussed as a possible future extension of the method. The results show that the algorithm is stable within a reasonable range of parameters and appears universally applicable as documented on the model systems.

The Helical Periodic Boundary Conditions (HPBC)

Within the standard rectangular PBC, the elementary cell is replicated by a translation. For an atom with a position vector **r,** the replicated positions are $\mathbf{r}' = \mathbf{a} + \mathbf{r}$, where the translational vector **a** is obtained by integer multiplications of the box dimensions b_x, b_y, b_z, ($a_\alpha = n_\alpha \times b_\alpha$, $n_\alpha = 0, 1, 2, \ldots$). Within HPBC, the helical periodicity (say in direction z) is introduced for a central part of the box, within a cylinder of radius r_1 (Fig. 1). This part would typically contain the molecule of interest and part of the solvent environment. During the replication, the cylinder is translated, that is, shifted by b_z, and rotated about its (helical) axis, by an angle φ_0. The rotation is symbolized by the matrix $\mathbf{U}(\varphi_0)$; the replicated positions can be

[a] Jiří Kessler, Petr Bouř
Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Flemingovo náměstí 2, 166 10, Prague, Czech Republic
E-mail: bour@uochb.cas.cz
[b] Jiří Kessler
Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University, Hlavova 8, 128 40, Prague, Czech Republic
Contract grant sponsor: Academy of Sciences; Contract grant number: M200551205; Contract grant sponsor: Grant Agency of the Czech Republic; Contract grant number: P208/11/0105; Contract grant sponsor: Ministry of Education; Contract grant number: LH11033 and LM2010005

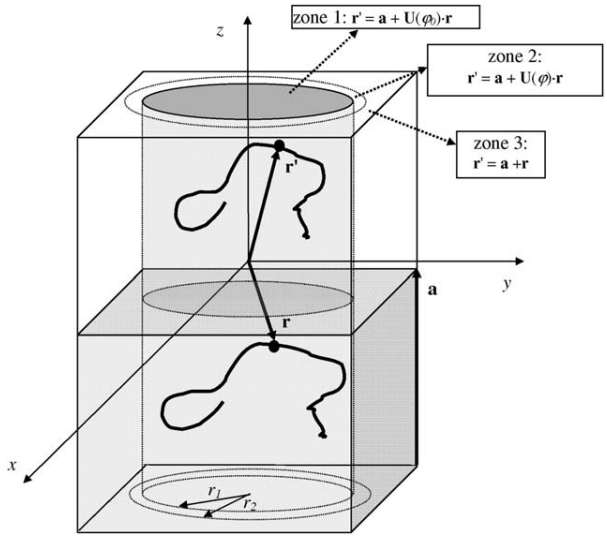


Figure 1. The helical periodic boundary conditions: within the inner cylinder (zone 1) molecules are translated and rotated by the angle φ_0; In the transition zone 2, a partial rotation by $\varphi = f\ \varphi_0$ is applied, and in zone 3 vertical translation by **a** takes place. **U** is the rotational matrix, **r** and **r'** are vectors pointing to an identical atom in neighboring boxes, *f* is a damping function, $f = 1$ and $f = 0$ at the walls of the inner and outer cylinder, respectively, see Figure 2.

expressed as $\mathbf{r}' = \mathbf{a} + \mathbf{U}(\varphi_0) \cdot \mathbf{r}$. As this is incompatible with PBC, a transition zone is defined by an annular space between the above mentioned cylinder (radius r_1) and another cylinder of radius r_2. Within the zone, an intermediate transformation takes place, as defined by $\mathbf{U}(\varphi)$, where $\varphi = f\ \varphi_0$, and *f* is a transition function.

The transition function "damps' the helicity, that is, $f = 1$ and 0 at the wall of the inner and outer cylinder, respectively. Similarly as other potential-adjusting functions, for example, van der Waals or Coulomb interaction cutoffs[4,13] it is desirable that *f* is simple, smooth, and perturbs this system as little as possible, which restricts its choice to a sigmoidal dependence. We chose

$$f(p) = \frac{1}{2}\left(1-(2p-1)\sqrt{\frac{d+1}{d+(2p-1)^2}}\right)$$

where $p = (r_{xy} - r_1)/(r_2 - r_1)$, r_{xy} being the distance from the *z*-axis, and the steepness parameter $d = 0.1$ (Fig. 2, left). Trial computations indicated that the model is not too sensitive to variation of the *d* parameter, although extreme values may lead to dynamic instabilities.

The introduction of the transition zone has little effect on the properties of the periodic simulated system if it is sufficiently separated from the central part of interest (see below). However, it introduces transition regions at the bottom and top of the cylinder, which may lead to serious instabilities during the MD simulation. The source of the problem is depicted at the right hand side of Figure 2, where we display a replication of two atoms (*a* and *b*) within the transitional zone along the helical axis. After a rotation and translation to the neighboring cell the interatomic distance r_{ab} may change, because the atoms *a* and *b* may not be rotated by the same angle. Note that this "deformation" does not affect the actual geometry within the box but only the "virtual" replicas used in the interaction with the environment. Only a small fraction of atoms close to the walls parallel with the *xy* plane and close to zone 2 is affected. Still, if untreated, the rotation causes instabilities in the annular regions close to the wall (*xy* plane) of the box. Most seriously, a solvent molecule passing through the wall may suddenly change its geometry, or a replicated atomic image may move much faster than the atom in the original box.

Fortunately, the problems can be overcome by modification of the propagation algorithm. Complete solvent molecules were rotated while producing the replicas, as based on the

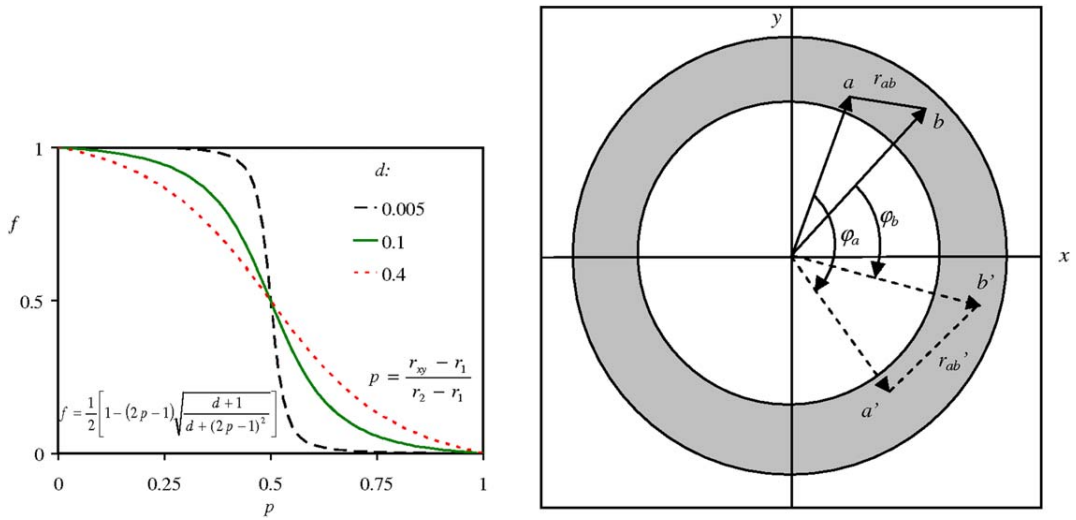


Figure 2. Examples of the damping function (left) and the deformation problem in the transition zone (right): after a rotation and translation about the helical axis (*z*, being perpendicular to the diagram) the distance r_{ab} between two atoms (*a* and *b*, prime denoting the rotated system) may not be conserved, as the rotation angles generally differ, $\varphi_a \neq \varphi_b$. See text for further explanation.

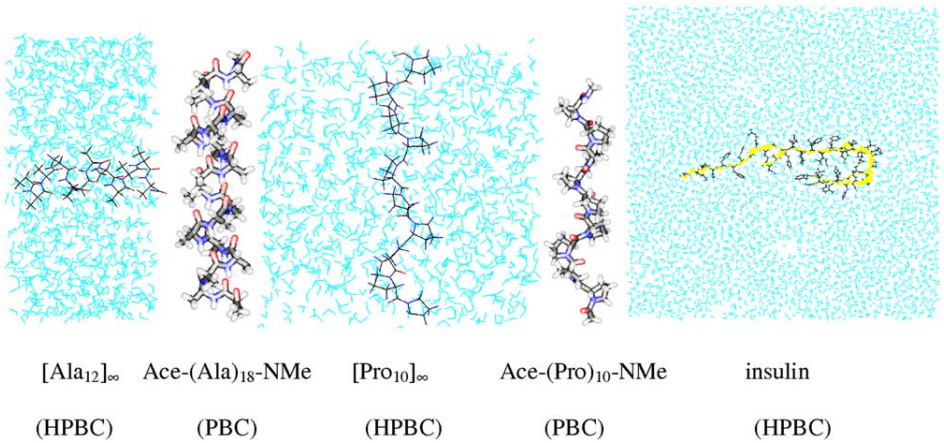


Figure 3. The model systems: HPBC simulation of polyalanine α-helix ($[Ala_{12}]_\infty$), Ace-$(Ala)_{18}$-NMe α-helix with PBC, HPBC simulation of polyproline II polyproline ($[Pro_{10}]_\infty$), the Ace-$(Pro)_{10}$-NMe peptide simulated with PBC, and insulin HPBC simulation. Water is not displayed in the PBC cases, and the plots are not to scale.

damping function determined for their mass centers. This prevented deformations of the geometry of individual molecules. Additionally, passing of molecules through the wall along the helical axis (z in Fig. 1) within the transitional zone was prevented during MD runs. This was achieved by applying a half-harmonic potential (V) for any molecule outside the box, which left it through zone 2, $V = k \times d^{[2]}$, where d is the distance from the box xy wall, and the force constant k was chosen to be 1 kcal/Å$^{[2]}$ ($V = 0$ inside the box). Values of k within 1–10 kcal/Å$^{[2]}$ yielded very similar results and did not significantly influence the geometry of interest in simulations not involving transport processes. Finally, for every dynamic integration step, each coordinate increment (Δr_a) was compared to a corresponding increment in the neighboring replicated box ($\Delta r_a'$). In cases when the latter shift was considerably larger ($|\Delta r_a'| > 1.1|\Delta r_a|$) the former shift was scaled down by a factor of $\Delta r_a/\Delta r_a'$.

Implementation

The HPBC were implemented within the Tinker[11] program. Full parallelization of the code was achieved under the OMP environment (see http://openmp.org/, a link to the adapted Tinker program can be found at http://hanicka.uochb.cas.cz/~bour/programs/list.html). The Amber99[12] force field (comprising the TIP3P[14] force field for water) was used without bond length constraints. The helical periodicity was applied to all the energy and gradient terms, that is, those associated with bond lengths, bond and torsional angles, improper torsions, van der Waals (Lennard–Jones), and point charge electrostatic interactions. All simulations were performed in *NVT* conditions and default Tinker parameters, using the default 9 Å cutoff distance both for van der Waals and Coulomb interactions. The cutoff was combined with potential switching and shifting,[15] using fifth-degree multiplicative and seventh-degree additive (Coulomb only) switching functions as implemented in the Tinker program[11] version 6.2. The Beeman[16] propagation scheme, 1 fs integration time step, temperature of 298 K, and the Berendsen[17] thermostat with the coupling parameter of 0.01 ps were used.

Examined Systems

The geometries of the principal studied systems are shown in Figure 3. As the first test system referred here to as $[Ala_{12}]_\infty$, we used an infinite polyalanine chain in α-helical conformation. To allow for a realistic motion and fluctuations of the peptide, 12 alanine residues and 813 water molecules were kept in the 37.2 × 37.2 × 18.56 Å$^{[3]}$ (helical axis) box, using $r_1 = 8$ Å and $r_2 = 17$ Å. Initial peptide backbone torsion angles ($\varphi = -60°$, $\psi = -45°$, and $\omega = 180°$) corresponded to a standard α-helical geometry.[18] After minimization and an equilibrium dynamics, the production run lasted 2 ns.

The twist angles φ_0 for all systems were determined on the basis of canonical values (X-ray data) or standard PBC simulations, and not varied during the dynamics. For ideal α-helix, its ideal value is about 99° per residue[19]; the equilibrium value supported by the Amber99 force field for polyalanine is only slightly larger (100.4° per residue), that is, 126° for $[Ala_{12}]_\infty$. To determine the effect of φ_0 variation on the dynamics, its magnitude for $[Ala_{12}]_\infty$ was additionally varied within 110°–140°. However, we emphasize that this variation goes beyond the implementation of the HPBC scheme (see above), where the twist is to be considered as a constant, similarly as the box dimensions.

For a comparison, standard PBC computation was set up for the Ace-$(Ala)_{18}$-NMe peptide and 2083 water molecules in a 40 × 40 × 40 Å$^{[3]}$ box, using the same standard initial torsion angles of the peptide backbone. To stabilize the helix, the terminal backbone torsion angles were fixed in three aminoacid residues at each (C— and N—) peptide end. Other MD parameters were the same as for the periodic system.

The second test system $[Pro_{10}]_\infty$ involved the polyproline chain in the polyproline II conformation. The system was created as for the α-helix, that is, 10 proline residues and 856 water molecules per a box dimensioned 30 × 30 × 29.9 Å$^{[3]}$. The twist angle φ_0 was changed within ~100°–120°, and the

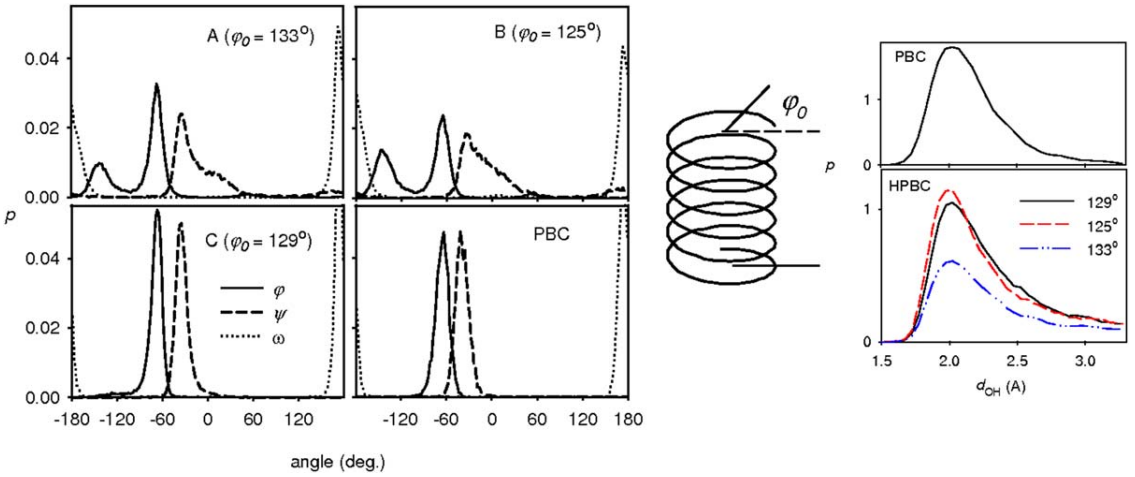


Figure 4. Probability distributions of the peptide main chain angles φ, ψ, and ω (left) and intrahelical hydrogen bond distances (for all NH···O atoms of i, $i+4$ residues, right) in the $[(Ala)_{12}]_\infty$ α-helical polypeptide, as obtained for three twist φ_0 angles, and distributions obtained from the standard PBC simulation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

initial standard polyproline II peptide backbone torsion angles[18] were set to $\varphi = -75°$, $\psi = 150°$, and $\omega = 180°$. By default, $r_1 = 9$ Å and $r_2 = 13.5$ Å although other dimensions of the transitional zone were also tested as specified below. To estimate a variation along the z(helical)-axis, we simulated also shorter ($[Pro_9]_\infty$, 30 × 30 × 26.91 Å$^{[3]}$) and longer ($[Pro_{12}]_\infty$, 30 × 30 × 35.88 Å$^{[3]}{}_\infty$) systems under analogous conditions as for $[Pro_{10}]_\infty$, and compared the results to a reference PBC computation performed for the Ace-$(Pro)_{10}$-NMe peptide and 2097 water molecules in the 40 × 40 × 40 Å$^{[3]}$ box. As for the α-helix, the total time of each simulation was 2 ns.

Finally, the largest system involved protonated insulin molecule (785 atoms), six Cl^- ions, and 4593 water molecules in 110 × 110 × 12 (helical axis) Å$^{[3]}$ box, with $\varphi_0 = 0°$, 3° and 6°, and $r_1 = 43$ Å and $r_2 = 53$ Å. The initial geometry mimicking the likely conformation of insulin in a fibrillar form was based on the β-roll protein (identification code 1VH4 in the protein data bank, http://www.rcsb.org/pdb/). Insulin torsion angles were set to those adopted by residues A256–A306 in the 1VH4 crystal structure.[20] The fibrillation of insulin is a well documented process.[21,22] The system was also partially inspired by the latest spectroscopic experiments using Raman optical activity that provided a new insight into the free insulin structure[23] and fibril formation.[24] On energy minimization and subsequent equilibration the system was let to develop for 300 ps under the default conditions of *NVT* ensemble, 298 K and 1 fs integration step.

For statistical analyses, we used in part our own scripts; resultant probability distributions of angles and distances are normalized ($\int p(x)dx = 1$).

Results and Discussion

Modeling of the α-helix

For the $[Ala_{12}]_\infty$ α-helical model we found a relatively tight dependence of the geometry on the overall twist φ_0. This is documented in Figure 4 where the calculated probability distributions of the peptide main chain torsion angles (φ, ψ, and ω) and the O···HN distances relevant for the intramolecular hydrogen bonds are plotted and compared to those obtained for the standard PBC Ace-$(Ala)_{18}$-NH-Me helical model. All distances measured between amide oxygen and hydrogen atoms of ith and $i+4$th aminoacid residues, respectively, were included in the statistics. For $\varphi_0 = 129°$, the average torsion angles ($\varphi = -66°$, $\psi = -35°$, and $\omega = 173°$) and their equilibrium distributions are very close to those obtained by PBC ($\varphi = -64°$, $\psi = -39°$, and $\omega = 173°$). Note that for the PBC model, the helix had to be stabilized by freezing the terminal torsion angles, which are, therefore, not included in the statistics. Nevertheless, both the HPBC and PBC results indicate the Amber99 force field enables standard α-helical polyalanine conformation only when arbitrarily stabilized, that is, by the infinite helicity in the former and by the terminal constraints in the latter simulation.

For the φ_0 twist angles of 133° and 125°, the α-helix is significantly perturbed in the HPBC simulations. The ω-angle is not significantly affected, which corresponds to the relative rigidity of the nearly planar amide group, but the φ and ψ angles are more dispersed, indicating a significant disruption of the regular helical geometry. Similarly, for $\varphi_0 = 129°$, the probability distribution of the hydrogen bond distances reasonably well correspond to the PBC simulation; interestingly, for $\varphi_0 = 133°$ the $i,\ldots,i+4$ hydrogen bonds are much more perturbed than for $\varphi_0 = 125°$.

The HPBC twist also perturbs motion of the solvent molecules. This is difficult to evaluate together with the peptide simulation, where many of the water molecules are hydrogen bonded to the solute. Instead, we applied the helical twist φ_0 to the box of the same size (37.2 × 37.2 × 18.56 Å$^{[3]}$) entirely filled with water. Because we were interested in the translational motion, presumably most affected by the twist, we defined an average translational atomic velocity as $<|r_i(t) - r_i(t-\Delta)|/\Delta>$, with a relatively large integration step of

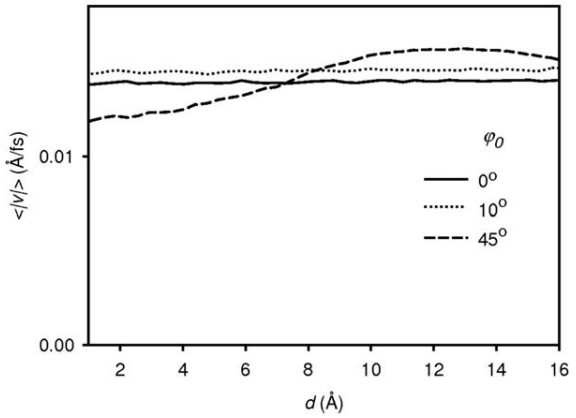


Figure 5. Average translational water atom velocities (see the text for definition) in the 37.2 × 37.2 × 18.56 Å[3] box filled with water only as perturbed by the helical twist, for $r_1 = 8$ Å, $r_2 = 17$ Å, d is the distance from the helical axis.

$\Delta = 10$ fs. Thus the translational motion (and part of the rotational one) could be separated from the vibrational degrees of freedom. The velocity as dependent on the distance from the helical axis is plotted in Figure 5. For φ_0 of 0° and 10°, the velocity is relatively independent of the distance, although for the latter case the solvent already appears to move slightly faster. For the angle of 45°, a larger deformation of the velocity profile appears. For radii approximately smaller than r_1, the velocity decreases, whereas it is larger than in the PBC simulation otherwise. It is, thus, certainly desirable to prefer smaller helical twists, which perturb the solvent motion less.

The HPBC dynamics of $[Pro_{10}]_\infty$

The polyproline II helix, known to well represent the so-called random conformation of general peptides,[25] is more flexible than α-helix. This corresponds to the wider distributions of the φ and ψ torsional angles (Fig. 6). The most probable angles ($\varphi = -69°$ and $\psi \sim 163°$) are almost the same for the HPBC and PBC dynamics. The maximal ω-value of 175° is only somewhat larger than in α-helix (173°) due to the absence of intramolecular hydrogen bonds in the polyproline. Compared to the α-helix, the polyproline structure appears to be more resistant to the overall twist. Nevertheless, for the φ_0 angles of 100° and 105°, the probability distribution already becomes markedly wider than for the most stiff geometry at $\varphi_0 = 110°$. This also corresponds well to the average value of 106.4° estimated from the CH_3CO-$(Pro)_9$-$CONHCH_3$ PBC simulation. The variation of the number of the proline residues ($[Pro_9]_\infty$ vs.

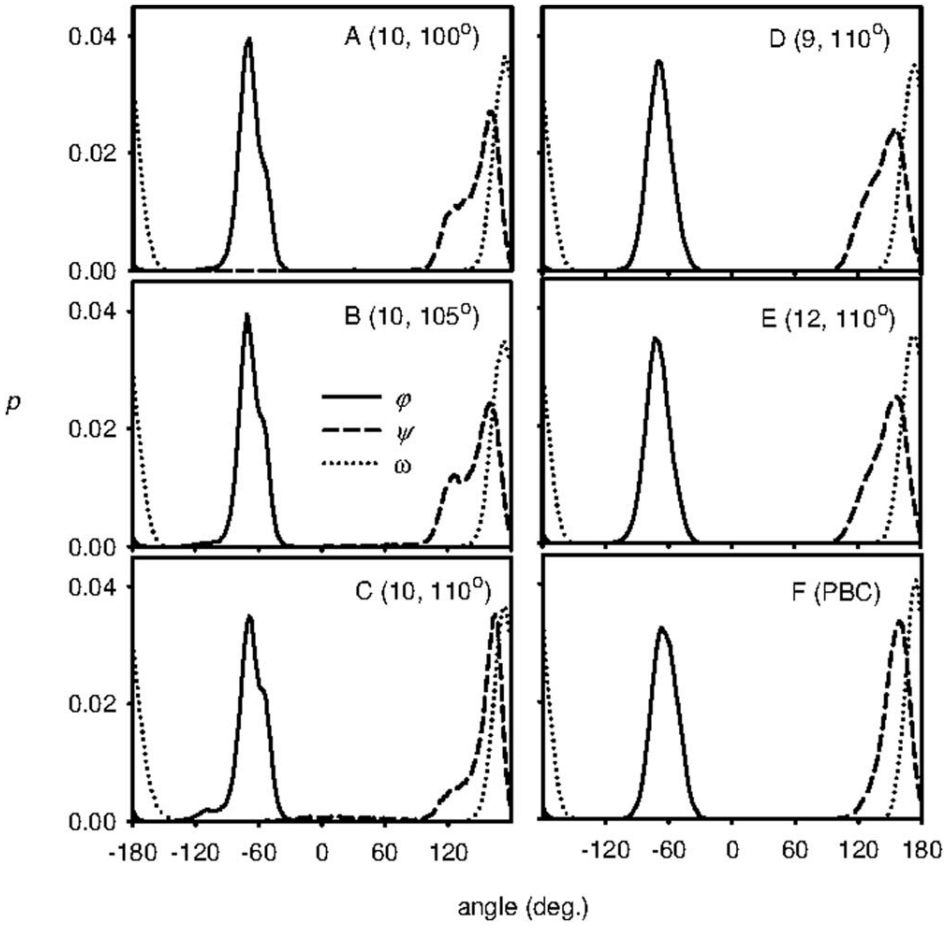


Figure 6. Probability distributions of the main chain torsion angles obtained using HPBC for various $[Pro_N]_\infty$ polyproline II models (panels A–E, the number N of the proline units and the twist φ_0 recalculated to $N = 10$ are indicated) and the PBC simulation of Ac-$(Pro)_9$-NMe (F).

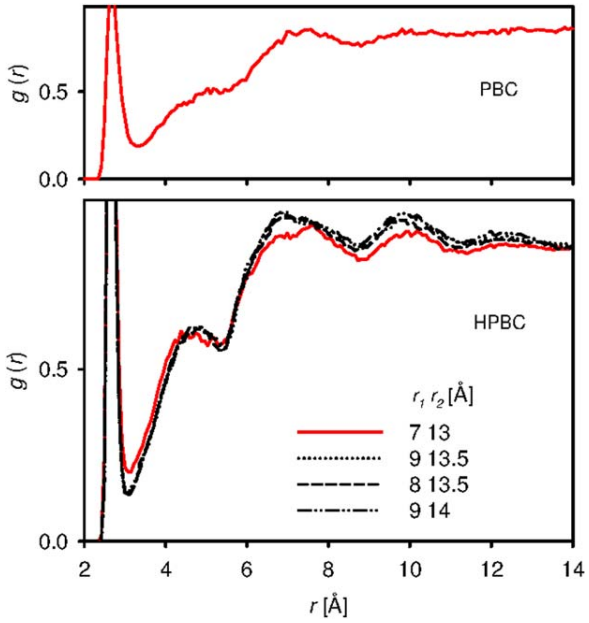


Figure 7. Radial distribution function (amide O···water mass center, bottom) in the $[Pro_{10}]_\infty$ system, as calculated for four different sizes of the transition zone; at the top, radial distribution from the standard PDB simulation of Ace-$(Pro)_{10}$-NMe is plotted. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$[Pro_{10}]_\infty$ vs. $[Pro_{12}]_\infty$) causes minor changes, mostly affecting the ψ-angle distribution. However, results of $[Pro_9]_\infty$ and $[Pro_{12}]_\infty$ provide a more symmetric distribution of φ than for $[Pro_{10}]_\infty$, due to a fine interplay between the box length and the helical twist.

We are naturally keen to find out if the transitional zone geometry affects the important simulated variables. Therefore, for example, the radial distribution function between an amide oxygen and mass centers of water molecules was calculated as shown in Figure 7 for four sizes of the zone, compared with the standard PBC result. Indeed, as apparent from the figure, if the zone is too close to the polyproline chain ($r_1 = 7$ Å) it perturbs the distribution in a wide range 2–14 Å of the radial distances between amide oxygen and H_2O mass center. For $r_1 = 8$ Å and larger the radial distribution function stabilizes and variations of the transition zone geometry have a negligible influence on it. Note that at the distance of $r_1 = 8$ Å from the helical axis some polyproline atoms are still as close as ~1.5 Å; the effect of the transitional zone on the geometry is thus quite minor and restricted to a short range only. The HPBC radial distribution function well reproduces the main features of the PBC result (Fig. 7, top) calculated for a center amide oxygen in Ace-$(Pro)_{10}$-NMe.

Similarly, for $r_1 = 7$ Å, the distribution of the torsional angles is perturbed in the HPBC dynamics (Fig. 8) but this is readily improved for the radii of 8 Å and greater. For $r_1 = 7$, the distribution of φ becomes wider, and ψ may even adopt a value around $-30°$, very far from the ideal polyproline II structure. Variation of the outer radius r_2 does not seem to have an effect at all, at least with respect to the polyproline geometry and the radial distribution function.

Insulin fibril vs. β-roll geometry

For the insulin model, we want to show that the geometry is compatible with that observed for the β-roll protein, as this would enable a long-range arrangement of the insulin monomeric units in the fibrils. Indeed, owing to the interaction to the box replicas the insulin geometry is quite stable during the HPBC dynamics (Fig. 9), and a fibril similar to the β-roll structure is formed.

The simulation time (300 ps) is rather short; nevertheless the simulations seem to exhibit converged potential energy profiles, and for the zero twist, the average geometry is very close to that obtained by the longer (8 ns) PBC computation (Fig. 10). The largest deviations appear outside the insulin loop for the semifree part of the peptide chain. For PBC, the Amber program[26] enabling a more extensive parallelization than Tinker was used, and the insulin molecule was stabilized by sandwiching it between two other molecules fixed in space in a 110 × 110 × 36 Å^3 box, otherwise the simulation parameters were the same.

In all cases, the conformation required for the fibril formation, thus, appears supported by the Amber99 force field. Additionally, for $\varphi_0 = 0°$ (standard PBCs), the structure is more flexible than for $\varphi_0 = 6°$ (cf. Fig. 9), which, as for the smaller systems' computations documented in Figures 4 and 6, most likely indicates that the latter "twisted" fibrillar geometry is favored by the Amber99 force field, in accord with the experimental observations of the β-roll protein[20] in other fibril studies.[27–30]

We recall that for each system the twist angle φ_0 is treated as a fixed parameter in the MD runs. Still, one can ask, if it can be somehow varied and optimized for each system. This, however, appears difficult. Indirectly, its equilibrium value could be

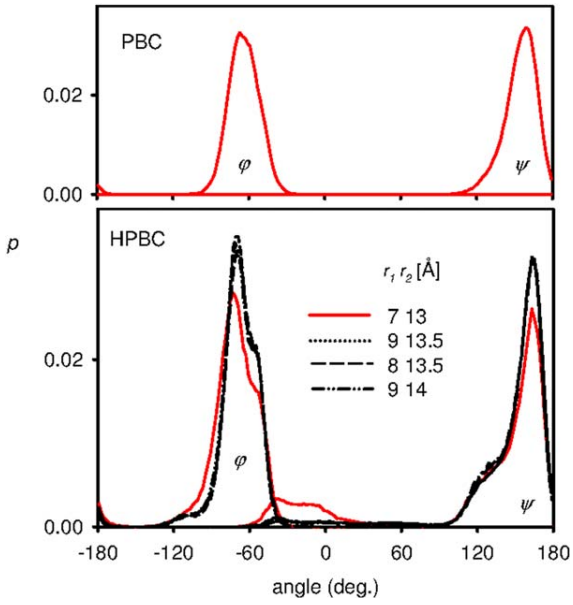


Figure 8. Probabilities of the φ and ψ angles in $[Pro_{10}]_\infty$ calculated for four different sizes of the transitional zone (bottom), and the result for standard PBC (top). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

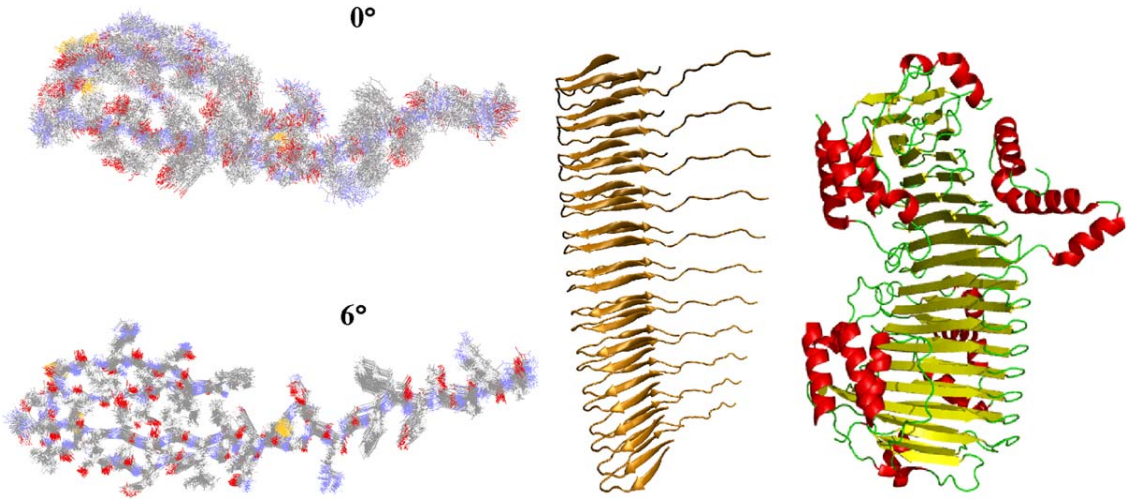


Figure 9. Snapshot overlaps from the HPBC dynamics of the insulin fibril for the twists φ_0 of 0° and 6° (left), the propagated fibril geometry (middle), and the β-roll 1VH4 protein X-ray structure (right). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

estimated from the geometric features of the system, such as those investigated in Figures 4, 6, and 9. We also attempted an alternate evaluation, from the average atomic torque. For each force term (bonds, torsions, bond angles, and pair wise interactions) in each MD step, we calculated the force $\mathbf{f}_i$ acting on an atom i at a position $\mathbf{r}_i$, and the moment $m_{i,z}$ corresponding to the helical axis, $m_{i,z} = f_{i,x}r_{i,y} - f_{i,y}r_{i,x}$. Then, we investigated if the negative moment components $<m_z^->$, averaged over a number of MD steps and atomic contributions, reaches a maximum at an equilibrium value of φ_0. For the $[Pro_{10}]_\infty$ polypeptide in empty and water-filled box, the dependencies of $<m_z^->$ on φ_0 are plotted in Figure 11. The dependence for polyproline in vacuum indicates a relatively sharp maximum for $\varphi_0 \sim 101.5°$. In water, the dependence indicates a maximum at $\varphi_0 \sim 106.0°$, well corresponding to the value of 106.4° from the unconstrained dynamics. Nevertheless, different values of $<m_z^->$ were obtained with independent MD runs for the same φ_0, and the overall precision remained low even for long simulation times, and such method may, thus, not be universally applicable.

Conclusions

We implemented the infinite helicity into the Tinker molecular dynamics program, based on a virtual transitional solvent region in the PBC replicas, where the helical twist angle φ_0 was treated as an additional constant parameter. The transitional region brought about some instabilities into the MD, which, however, could be effectively rectified by additional restraints, and which do not seem to significantly perturb the

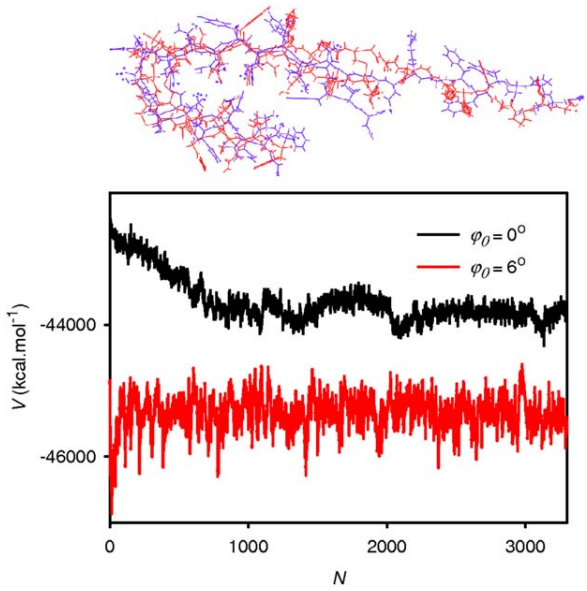


Figure 10. (Top) overlap of the average structures of the insulin fibril model obtained by HPBC (Tinker, $\varphi_0 = 0°$, 0.3 ns) and PBC (Amber99, 8 ns) dynamics, and (bottom) the potential energy from the HPBC simulations for the two twist angles, N is number of the snapshot taken in 100 fs intervals. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

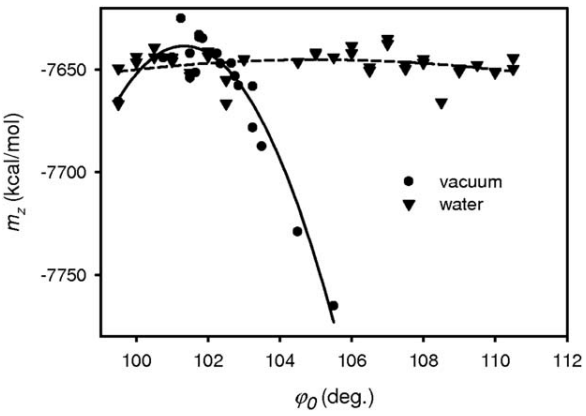


Figure 11. Average m_z torque moment dependent on the twist, as calculated with HPBC for $[Pro_{10}]_\infty$ in vacuum and in the water-filled box. The y-scale for the water was scaled down according to the vacuum values; quadratic fits are indicated by the solid and dashed lines. Note that some MD runs were repeated with different initial velocities, providing different m_z for the same twist angle.

geometry of the studied system. The HPBC dynamics was successfully tested on model systems, appeared to be compatible with common MD algorithms and force fields, and can, thus, be further used in theoretical investigations of the helical structures. For the most complex example, the insulin fibril, the HPBC dynamics made it possible to verify the stability of the fibril-forming conformation, which document the possibilities of this approach for structure and dynamics modeling of larger biologically relevant molecular systems.

Acknowledgments

The authors thank Prof. Shigeki Yamamoto (Osaka University) for the discussion on the insulin structure and Dr. Radek Pelc (Stentor Institute) for helpful comments to the manuscript.

Keywords: periodic boundary conditions · helical symmetry · molecular dynamics · protein structure · amyloid fibrils

How to cite this article: J. ř Kessler, P. Bouř. *J. Comput. Chem.* **2014**, *35*, 1552–1559. DOI: 10.1002/jcc.23653

[1] S. L. Hooper, K. H. Hobbs, J. B. Thuma, *Prog. Neurobiol.* **2008**, *86*, 72.
[2] J. C. Stroud, C. Liu, P. K. Teng, D. Eisenberg, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 7717.
[3] W. F. Boron, E. L. Boulpaep. Medical Physiology: A Cellular and Molecular Approach; Elsevier/Saunders: Philadelphia, **2003**.
[4] M. P. Allen, D. J. Tildesley. Computer simulation of liquids; Oxford University Press: New York, **1987**.
[5] P. L. Cummins, J. E. Gready, *J. Comput. Chem.* **1996**, *17*, 1598.
[6] G. Brancato, N. Rega, V. Barone, *J. Chem. Phys.* **2008**, *128*, 144501.
[7] M. B. Hamaneh, M. Buck, *J. Comput. Chem.* **2009**, *30*, 2635.
[8] J. Li. In Handbook of Materials Modeling; S. Yip, Ed.; Springer: Amsterdam, **2005**, p 565.
[9] W. Liu, B. Schmidt, G. Voss, W. Müller-Wittig, *Comput. Phys. Commun.* **2008**, *179*, 634.
[10] C. J. Fennell, J. D. Gezelter, J. Chem. *Phys.* **2006**, *124*, 234104.
[11] J. W. Ponder, *Tinker 6.2*, Washington University School of Medicine: Saint Louis, **2000**.
[12] N. Kamiya, Y. S. Watanabe, S. Ono, J. Higo, *Chem. Phys. Lett.* **2005**, *401*, 312.
[13] T. Schlick. Molecular Modeling and Simulation; Springer: Berlin, **2002**.
[14] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, *J. Chem. Phys.* **1983**, *79*, 926.
[15] P. J. Steibach, B. R. Brooks, *J. Comput. Chem.* **1994**, *15*, 667.
[16] D. Beeman, *J. Comput. Phys.* **1976**, *20*, 130.
[17] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, J. R. Haak, J. *Chem. Phys.* **1984**, *81*, 3684.
[18] T. E. Creighton. Proteins: Structures and Molecular Properties; W. H. Freeman and Co.: New York, **1993**.
[19] S. Kumar, M. Bansal, *Biophys. J.* **1998**, *75*, 1935.
[20] J. Badger, J. M. Sauder, J. M. Adams, S. Antonysamy, K. Bain, M. G. Bergseid, S. G. Buchanan, M. D. Buchanan, Y. Batiyenko, J. A. Christopher, S. Emtage, A. Eroshkina, I. Feil, E. B. Furlong, K. S. Gajiwala, X. Gao, D. He, J. Hendle, A. Huber, K. Hoda, P. Kearins, C. Kissinger, B. Laubert, H. A. Lewis, J. Lin, K. Loomis, D. Lorimer, G. Louie, M. Maletic, C. D. Marsh, I. Miller, J. Molinari, H. J. Muller-Dieckmann, J. M. Newman, B. W. Noland, B. Pagarigan, F. Park, T. S. Peat, K. W. Post, S. Radojicic, A. Ramos, R. Romero, M. E. Rutter, W. E. Sanderson, K. D. Schwinn, J. Tresser, J. Winhoven, T. A. Wright, L. Wu, J. Xu, T. J. R. Harris, *Proteins* **2005**, *60*, 787.
[21] L. Nielsen, S. Frokjaer, J. Brange, V. N. Uversky, A. L. Fink, *Biochemistry* **2001**, *40*, 8397.
[22] J. Brange, L. Anderson, E. D. Laursen, G. Meyn, E. Rasmussen, *J. Pharma. Sci.* **1997**, *86*, 517.
[23] S. Yamamoto, J. Kaminský, P. Bouř, *Anal. Chem.* **2012**, *84*, 2440.
[24] S. Yamamoto, H. Watarai, *Chirality* **2012**, *24*, 97.
[25] R. K. Dukor, T. A. Keiderling, *Biopolymers* **1991**, *31*, 1747.
[26] D. A. Case, I. T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, J. K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. Woods, *J. Comput. Chem.* **2005**, *26*, 1668.
[27] H. Chi, W. R. W. Welch, J. Kubelka, T. A. Keiderling, *Biomacromolecules* **2013**, *14*, 3880.
[28] S. Wójcik, V. Babenko, W. Dzwolak, *Langmuir* **2010**, *26*, 18303.
[29] W. Dzwolak, W. Surmacz-Chwedoruk, V. Babenko, *Langmuir* **2013**, *29*, 365.
[30] D. Kurouski, R. K. Dukor, X. Lu, L. A. Nafie, I. K. Lednev, *Chem. Commun.* **2012**, *48*, 2837.

Received: 25 March 2014
Revised: 20 May 2014
Accepted: 26 May 2014
Published online on 10 June 2014

9.7 Appendix G – *J. Chem. Theory Comput.* 2015, 11, 2210.

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

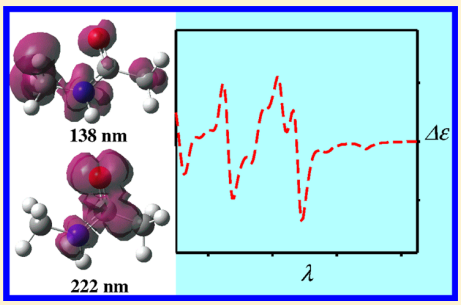
Transfer of Frequency-Dependent Polarizabilities: A Tool To Simulate Absorption and Circular Dichroism Molecular Spectra

Jiří Kessler[†,‡] and Petr Bouř*[,†]

[†]Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Flemingovo náměstí 2, 166 10 Prague, Czech Republic

[‡]Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University, Hlavova 8, 128 40 Prague, Czech Republic

ABSTRACT: Absorption and circular dichroism spectra reveal important information about molecular geometry and electronic structure. For large molecules, however, spectral shapes cannot be computed directly. In the past, transition dipole coupling (TDC) and related theories were proposed as simplified ways of calculating the spectral responses of large systems. In the present study, an alternative approach better reflecting the chemical structure is explored. It is based on the transfer of complex frequency-dependent polarizabilities (TFDP) of molecular fragments. The electric dipole−electric dipole, electric dipole−electric quadrupole, and electric dipole−magnetic dipole polarizabilities are obtained separately for individual chromophores and then transferred to a larger system composed of them. Time-dependent density functional theory and the sum over states methodology were employed to obtain the polarizability tensors of *N*-methylacetamide, and porphyrin molecules were chosen for a numerical test. The TFDP fails for charge-transfer states and close chromophores; otherwise, the results suggest that this method is capable of reproducing the spectra of large systems of biochemical relevance. At the same time, it is sufficiently flexible to account for a wide range of transition energies and environmental factors instrumental in the modeling of chromophore properties. The TFDP approach also removes the need for diagonalization in TDC, making computations of larger molecular systems more time-efficient.



INTRODUCTION

The transfer of spectroscopic properties from fragments to chemically similar residues in a bigger system is an efficient tool for understanding the optical properties of large molecules. For example, the Cartesian coordinates-based transfer (CCT)[1−3] of molecular property tensors[4] and similar approaches[5,6] have made it possible to simulate and interpret vibrational spectra of many large molecules of biological relevance.[2,7−9]

The situation for electronic spectra is quite different. Changes of electronic states are often associated with an extensive rearrangement of the electronic cloud, which can hardly be attributed to individual atoms. Several semiempirical schemes were proposed, such as the coupled oscillator model[10] and the transition matrix theory,[11] where chromophores and their interactions are treated via adjustable parameters. In the simplest transition dipole coupling (TDC) approach, chromophores are replaced by dipoles, and only the electrostatic interaction between them is considered. Encouraging results were obtained, and this was particularly true for biopolymers, including peptides, proteins, and nucleic acids.[12−17] Nevertheless, relatively complicated and mostly ad hoc corrections are necessary to make those approaches more precise and better reflect the chemistry of the system being studied.[18]

For small and medium sized molecular systems, the time-dependent density functional theory (TDDFT)[19] most likely represents the most practical way of obtaining absorption and circular dichroism spectra.[20−23] However, for biopolymers, direct TDDFT applications are limited not only by molecular size but also by the sensitivity of the spectra to environmental and dynamic effects.[24−29] A simplification using rigid Kohn−Sham orbitals has been proposed but has not been found to be sufficiently accurate.[30]

In the present study, we explore the transfer of frequency-dependent polarizabilities (TFDP) as a way of extending accurate ab initio (e.g., TDDFT) approaches to larger systems. The polarizability can be associated with an atom, bond, or a large fragment (chromophore), and it is transferred to a similar unit in the modeled system. Previously, analogous bond polarizability theory[31−33] or the atomic dipole interaction model,[34] breaking molecules down to their components, have already been successfully used to model Raman scattering and vibrational optical activity. The TFDP presented here includes both real and imaginary (dispersion) polarizability components; therefore, it is suitable to provide the electronic spectra as a dispersion part of the total polarizability. Being inherently linked to the chemical structure, the polarizability tensors provide the necessary flexibility to reflect the local chemical environment with the accuracy of TDDFT or other quantum-chemical computations. TFDP is thus better suited to reflect nuances of molecular geometry than previous semiempirical approaches.

Received: February 11, 2015
Published: April 20, 2015

DOI: 10.1021/acs.jctc.5b00136
J. Chem. Theory Comput. 2015, 11, 2210−2220

This study was also partially inspired by the modern "polarizable embedding" approaches that allow for accurate description of a chromophore entity, whereas longer range interactions are treated in a simpler way.[35,36] These combined quantum-mechanics and molecular-mechanics approaches allow for modeling of a range of effects, including system dynamics, protein or solvent environment, spectroscopic response, and chemical reactivity.[37−39]

We must emphasize, however, that TFDP and other transfer schemes are currently only suitable for weakly interacting (distant) chromophores. For charge transfer and other effects, the transferability concept is not valid, and a rigorous quantum mechanical (at least TDDFT) approach is necessary. Even at the limit of a large separation, the quantum-mechanical and classical (transfer-based) treatment of the chromophores may not converge; however, they are usually close enough to provide a solid basis for prediction and understanding of experimental data.

A solid theoretical basis for interpretation of the absorption and electronic circular dichroism (CD, ECD) spectra is important because of their vast potential in monitoring biomolecular structure and interactions. In particular, ECD (i.e., the differential absorption of the left- and right-circularly polarized visible or ultraviolet light) is very sensitive to molecular structure, conformation, and interactions with the environment.[40−45]

Unlike for previous models,[32,33] the polarizability including the real and dispersive (imaginary) parts is transferred as a function of the frequency. The introduction of a bandwidth parameter (Γ) preventing divergence at the resonant frequencies allows for modeling of light absorption. The theory is formulated in such a way that chromophore polarizabilities can be broken down to atomic ones. We also introduce an arbitrary frequency-dependent dipole polarization density (ρ) useful in the visualization of molecular excitation phenomena and localization of chromophore transitions. Fundamental TFDP properties and comparison with the TDC model are explored for a model *N*-methylacetamide (NMA) dimer. Larger porphyrin and photosynthetic reaction center models are used to document a typical application and as well as limitations of the TFDP approach.

FREQUENCY-DEPENDENT POLARIZABILITIES

Light scattering and absorption on molecules, including chiral phenomena such as differential absorption or scattering of left- and right-circularly polarized light, are conveniently described with the aid of molecular property tensors. The most important ones are the electric dipole−electric dipole (α), electric dipole−magnetic dipole (G'), and electric dipole−electric quadrupole (A) polarizabilities, defined in atomic units as[4,46]

$$\alpha_{\alpha\beta}(\omega) = 2\sum_{j\neq n}\omega_{jn} f_{jn}(\omega)\mathrm{Re}\mu_{nj,\alpha}\mu_{jn,\beta} \tag{1a}$$

$$G'_{\alpha\beta}(\omega) = -2\omega\sum_{j\neq n} f_{jn}(\omega)\mathrm{Im}\,\mu_{nj,\alpha} m_{jn,\beta} \tag{1b}$$

$$A_{\alpha,\beta\gamma}(\omega) = 2\sum_{j\neq n}\omega_{jn} f_{jn}(\omega)\mathrm{Im}\,\mu_{nj,\alpha}\Theta_{jn,\beta\gamma} \tag{1c}$$

$$\alpha^{\nabla}_{\alpha\beta}(\omega) = 2\sum_{j\neq n} f_{jn}(\omega)\mu_{nj,\alpha}\nabla_{jn,\beta} \tag{1d}$$

where the indexes j and n denote excited and ground states, respectively. For exact wave functions, the last expression (1d) is equal to 1a; the gradient form (α^{∇}) of the polarizability is based on dipole-velocity transformation, $\nabla_{jn,\beta} = \omega_{jn}\mu_{jn,\beta}$. In eqs 1a−1d, the sums run over the excited electronic states, n denotes the ground state, $\mu_{nj,\alpha} = \langle n|\mu_\alpha|j\rangle$ corresponds to the electric dipole moment, and an analogous notation is used for the gradient, magnetic dipole **m**, and electric quadrupole Θ. The Greek indices (α, β, and γ) are reserved for the Cartesian components (x, y, and z), $\omega_{jn} = \omega_j - \omega_n$ is the difference of angular frequencies for each state, and ω is the frequency of illuminating light.

With the transfer in mind, it is important to realize that while α is independent of the choice of coordinate origin, for an origin shift from **O** to **O** + **T**, tensors G' and A change to[4,47]

$$G'_{\alpha\beta} \rightarrow G'_{\alpha\beta} + \frac{1}{2}\omega\varepsilon_{\beta\gamma\delta}T_\gamma\alpha^{\nabla}_{\alpha\delta} \tag{2a}$$

$$A_{\alpha,\beta\gamma} \rightarrow A_{\alpha,\beta\gamma} - \frac{3}{2}(T_\beta\alpha_{\gamma\alpha} + T_\gamma\alpha_{\beta\alpha}) + \sum_{\delta=1}^{3} T_\delta\alpha_{\delta\alpha}\delta_{\beta\gamma} \tag{2b}$$

The frequency function f_{jn} can be written as[4]

$$\begin{aligned} f_{jn}(\omega) &= \frac{{\omega_{jn}}^2 - \omega^2}{({\omega_{jn}}^2 - \omega^2)^2 + \Gamma^2{\omega_{jn}}^2} \\ &\quad - i\frac{\Gamma\omega_{jn}}{({\omega_{jn}}^2 - \omega^2)^2 + \Gamma^2{\omega_{jn}}^2} \\ &= f_{r,jn}(\omega) + if_{i,jn}(\omega) \end{aligned} \tag{3}$$

where the frequency uncertainty Γ reflects the finite lifetimes of the excited electronic levels in case of resonance ($\omega \approx \omega_{jn}$). For the purposes of spectral modeling, Γ can also be associated with inhomogeneous line broadening. Then, the imaginary ("dispersion") part in eq 3 is responsible for light absorption, and traces of the tensors α and G' directly correspond to the absorption and CD spectra, respectively. For isotropic samples considered in the present study, the contribution of the A tensor to CD intensities vanishes.[4] In particular, when integrating over a selected transition (for $\Gamma \ll \omega_{jn}$), eqs 1a, 1b, and 3 yield

$$\int_{jn}\sum_{\alpha=1}^{3}\alpha_{\alpha\alpha}(\omega)\,\mathrm{d}\omega \cong -i\pi D_{nj} \tag{4a}$$

$$\int_{jn}\sum_{\alpha}G'_{\alpha\alpha}(\omega)\,\mathrm{d}\omega \cong i\pi R_{nj} \tag{4b}$$

where $D_{nj} = \boldsymbol{\mu}_{nj}\cdot\boldsymbol{\mu}_{jn}$ and $R_{nj} = \boldsymbol{\mu}_{nj}\cdot\mathbf{m}_{jn}$ are the usual dipole and rotational strengths, respectively.[48]

We can therefore relate the absorption (ε) and differential ($\Delta\varepsilon$) coefficients, both measured in L mol^{-1} cm^{-1}, to the polarizabilities and angular frequency ω computed in atomic units as follows[48]

$$\varepsilon(\omega) = -\frac{108.7}{\pi}\mathrm{debye}^2\omega^2\sum_{\alpha=1}^{3}\mathrm{Im}\alpha_{\alpha\alpha}(\omega) \tag{5a}$$

$$\Delta\varepsilon(\omega) = 4\times\frac{108.7}{\pi c}\mathrm{debye}^2\omega^2\sum_{\alpha=1}^{3}\mathrm{Im}G'_{\alpha\alpha}(\omega) \tag{5b}$$

where debye = 2.542 and c = 137.5.

DOI: 10.1021/acs.jctc.5b00136
J. Chem. Theory Comput. 2015, 11, 2210−2220

DIPOLE POLARIZATION DENSITY

We introduce the dipole polarization density as an arbitrary tool useful in locating chromophores as spectroscopically active molecular parts. Within a simplified sum over states theory where the excited states were substituted by the TDDFT expansions,[49,50] we calculate the dipole matrix element as

$$\mu_{nj,\alpha} = -\sqrt{2}\sum_{ab} c_{ab}^{j}\int \varphi_a(r) r_\alpha \varphi_b(r)\,\mathrm{d}\mathbf{r} \tag{6}$$

where c_{ab}^{j} are the TDDFT expansion coefficients of state j into single-excited Slater determinants (from an occupied molecular orbital a to a virtual orbital b). By inserting 6 into 1a, we get

$$\alpha_{\alpha\beta}(\omega) = \int \rho_\beta(\mathbf{r},\omega) r_\alpha\,\mathrm{d}\mathbf{r} \tag{7}$$

where we define

$$\rho_\beta(\mathbf{r},\omega) = -2\sqrt{2}\sum_{j\neq n}\omega_{jn} f_{jn}(\omega)\mu_{jn,\beta}\sum_{ab} c_{ab}^{j}\varphi_a(\mathbf{r})\varphi_b(\mathbf{r})$$

as the electric dipole polarization density. The dipole density $\rho_\beta(\mathbf{r},\omega)$ has a real and imaginary part because of the frequency function (eq 3). Most naturally, the absolute value of the density, $\rho = (|\rho_x|^2 + |\rho_y|^2 + |\rho_z|^2)^{1/2}$, can be used as an indication as to which molecular part is the observable transition localized.

DISTRIBUTION OF POLARIZABILITIES OVER MOLECULAR PARTS

In principle, the dipole polarization density defined in the previous section can be used as a tool to refine the distribution of chromophore polarizabilities over individual atoms. However, as the resultant spectra for separated chromophores are rather insensitive to fine distribution changes, this option is not investigated further. Here, we present results where (i) the chromophore is considered to be a point object, and the polarizability is placed to its center of mass, and (ii) the total polarizability calculated for a chromophore is assigned to its atoms with equal weights. The latter option enables estimation of the error associated with the point chromophore approximation.

TRANSITION DIPOLE MODEL

TDC is a well-established method of generating approximate biomolecular spectra[10] and is important as a reference because it is nearly exact for distant chromophores.[51] As detailed implementations may vary, we briefly describe the equations we used. The transition energies (e_i) and electric transition dipole moments ($\boldsymbol{\mu}_i$) obtained by TDDFT for a monomer/chromophore were transferred to each component in a dimer or a larger system. The energies, dipole ($D = \boldsymbol{\mu}_{gk}\cdot\boldsymbol{\mu}_{kg}$) and rotational ($R = \mathrm{Im}\boldsymbol{\mu}_{gk}\cdot\mathbf{m}_{kg}$) strengths for a transition $g \rightarrow k$ in the whole system are obtained via diagonalization of the interaction Hamiltonian.[10,51−54] The diagonal Hamiltonian elements are equal to the transition energies, $H_{ii} = e_i$, and off diagonal elements correspond to the dipole−dipole interaction, $V_{ij} = (r_{ij}^2\boldsymbol{\mu}_i\cdot\boldsymbol{\mu}_j - 3\boldsymbol{\mu}_i\cdot\mathbf{r}_i\boldsymbol{\mu}_j\cdot\mathbf{r}_{ij})/r_{ij}^5$. Then $\boldsymbol{\mu}_{gk} = \sum_j c_j^k\boldsymbol{\mu}_j$ and $\mathbf{m}_{gk} = (i/2)\sum_j c_j^k e_j\mathbf{r}_j \times \boldsymbol{\mu}_j$, where c_j^k are elements of the eigenvectors, and $\mathbf{r}_j$ are electric dipole moment positions. In our computations, we positioned the dipoles in the geometrical center of the HNCO groups (*N*-methylformamide) or in the mass center (porphyrin).

TRANSFER OF FREQUENCY-DEPENDENT POLARIZABILITIES (TFDP)

To obtain spectral properties of a larger system, we calculated the frequency-dependent polarizabilities $\boldsymbol{\alpha}$ and G' defined above for each chromophore by TDDFT and positioned in its mass center. Then, total (complex) polarizabilities of the entire system were obtained as[3,4,55]

$$\alpha_{\alpha\beta} = \sum_i \alpha_{i,\alpha\beta} + \sum_{j\neq i}\Big[\alpha_{i,\alpha\chi}T_{ij,\chi\delta}\alpha_{j\delta\beta} + \frac{1}{3}\big(A_{i\alpha,\chi\gamma}t_{ij,\chi\gamma\delta}\alpha_{j,\delta\beta} - \alpha_{i,\alpha\chi}t_{ij,\chi\delta\gamma}A_{j\beta,\delta\gamma}\big) + \frac{1}{c^2}G'_{i\alpha\chi}T_{ij,\chi\gamma}G'_{j\beta,\gamma}\Big] \tag{8a}$$

$$G'_{\alpha\beta} = \sum_i\Big(G'_{i\alpha,\beta} - \frac{\omega}{2}\varepsilon_{\beta\gamma\delta}r_{i,\gamma}\alpha_{i,\delta\alpha}\Big) - \frac{\omega}{2}\sum_{i\neq j}\Big[\varepsilon_{\beta\gamma\delta}r_{i,\gamma}\alpha_{i,\delta a}T_{ij,ab}\alpha_{j,b\alpha} + \frac{1}{3}\varepsilon_{\beta\varepsilon\delta}r_{i,\varepsilon}\big(\alpha_{j,\alpha a}t_{ij,\lambda\gamma\alpha}A_{i\delta,\lambda\gamma} - \alpha_{i,\delta a}t_{ij,a\lambda\gamma}A_{j\alpha,\lambda\gamma}\big) - \frac{2}{\omega}G'_{i\gamma,\beta}T_{ij,\gamma\delta}\alpha_{j,\delta\alpha}\Big] \tag{8b}$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ is the vector linking the centers of chromophores i and j, and the "distance tensors" are defined as $T_{ij,\alpha\beta} = (3r_{ij,\alpha}r_{ij,\beta} - \delta_{\alpha\beta}r_{ij}^2)(4\pi\varepsilon_0 r_{ij}^5)^{-1}$ and $t_{ij,\alpha\beta\gamma} = \nabla_{i\gamma}T_{ij,\alpha\beta}$. For brevity, we omit the frequency dependence in 8a and 8b. Note that the total polarizabilities are not mere sums of the individual components but also contain terms accounting for the mutual chromophore polarization. For G', an additional term is present due to the origin-dependence of this tensor (see eqs 2a and 8b).

Alternatively to this simpler TFDP approach, chromophore polarizabilities were distributed to individual atoms within the chromophore. For the models (porphyrin, NMA), we considered all atoms in the distribution. Then, the summations in 8a and 8b were run over the atomic indices. We refer to this variant as $\mathrm{TFDP_d}$.

Finally, monomer/chromophore polarizabilities were calculated in the presence of atomic partial charges mimicking the rest of the simulated system/dimer. This "embedded" variant is referred to as $\mathrm{TFDP_e}$. Obviously, the method also enables the "$\mathrm{TFDP_{de}}$" combination (i.e., to estimate chromophore polarizabilities in the presence of partial charges and distribute them to chromophore atoms). The TDC and TFDP models are diagrammatically summarized in Figure 1 as applied to an *N*-methylacetamide dimer.

COMPUTATIONAL DETAILS

Model systems included *N*-methylacetamide (NMA), porphyrin dimers, and a porphyrin cluster derived from the photosystem reaction center.

The geometry of one NMA molecule was optimized by energy-minimization using the Gaussian[56] program suite. The B3LYP[57]/6-311++G** approximation level was chosen as a default, as it previously provided reasonably accurate results in similar studies.[58,59] Other methods (CIS,[60] CAM-B3LYP,[61] LC-wPBE,[62] and TDHF[60]) with the same basis set were performed for control calculations as reported below. Electrostatic charges of NMA for the $\mathrm{TFDP_e}$ method were obtained by the Mulliken (MU) population analysis[63] and by the Merz−Singh−Kollman (MS),[64] CHelp (CHelp),[65] and Hu, Lu, and Yang (HLY)[66] electrostatic field fitting schemes at the B3LYP/6-311++G** level. The dimers were formed by rotation of one NMA molecule by 20° about an axis perpendicular to the NMA plane and shifting it by 4.5, 6.5, and 8.5 Å in a direction perpendicular to that plane (see Figure 2). For the dimers,

DOI: 10.1021/acs.jctc.5b00136
J. Chem. Theory Comput. 2015, 11, 2210−2220

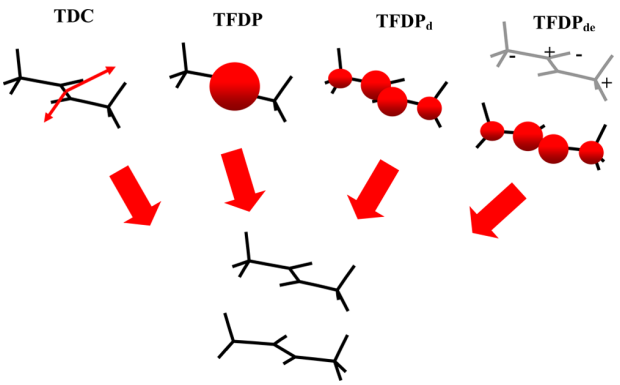


Figure 1. Overview of the transfer scheme used to generate spectra of an NMA dimer. Within the transition dipole coupling model (TDC), transition electric dipole moments (red arrows) are calculated for a monomer and transferred to the dimer components. For the transfer of frequency-dependent polarizabilities (TFDP), the polarizability tensors (symbolized by red circles) are transferred in a similar way. In the TFDP$_d$ variant, the polarizability is distributed over chromophore parts that interact separately after the transfer. By the "e" subscript, we denote a case in which the monomer environment is approximated by atomic partial charges prior to the polarizability computation and transfer.

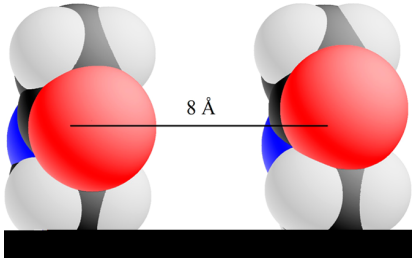


Figure 2. A dimer of two *N*-methylacetamide (NMA) molecules separated by 8 Å; van der Waals atomic radii used for the plot indicate the approximate extension of the electronic cloud. In this case, electron densities of the monomers do not overlap, and it can be expected that the semiempirical transfer schemes provide reasonable results.

absorption and electronic circular dichroism spectra were calculated using the time-dependent density functional theory (TDDFT).[19,20] The frequency-dependent polarizabilities were calculated from the transition moments and excitation energies (i.e., not the Kohn−Sham orbital energies) obtained from the Gaussian output by the SOS approach as outlined elsewhere.[49] Control computation indicated that for a complete set of monoexcited states the SOS method provides virtually the same results as the coupled-perturbed (response) computation but in a shorter time period. The frequency-dependent polarizability components were saved for frequencies covering the entire excitation range incremented in 1 nm steps while applying a frequency uncertainty parameter (Γ) of 10 nm (i.e., variable in the energy-scale) to match the usual experimental band broadening.

We adapted our "cctn" program originally developed for vibrational spectroscopic parameters to enable transfer of the frequency-dependent tensors as well. The transfer is based on the best (least-squares distance method) overlap between the source and target chemical entities and a unitary (rotation) transformation of all Cartesian tensor indices; the details can be found in previous studies.[1−3]

As a "real world" model, we also investigated cyanobacterial photosystem I (Figure 3)[67] for which X-ray coordinates, including 96 porphyrin chromophores, are available as the 1JB0 entry in the protein data bank database (http://www.rcsb.org). A simplified monomer porphyrin molecule (Figure 3) was generated using the 1JB0 coordinates as an initial guess. Minimal relaxation of the geometry was allowed within the constrained (ω_{max} = 300 cm^{-1}, see ref 68) normal mode optimization (NMO)[69,70] at the B3LYP/6-311++G** level and the transition dipole moments and frequency-dependent polarizabilities estimated as for the NMA system. Additionally, two porphyrin dimers were constructed, for which the TFDP and TDC results could be compared to the TDDFT benchmark. The geometries were based on the 1JB0 positions of porphyrins comprising magnesium atom numbers 23357 and 23402, 23.3 Å apart (dimer 1), and magnesia numbers 17407 and 20451, 6.34 Å apart (dimer 2). However, the simplified monomer units lacking the Mg ion were used in the computations. To estimate in detail the computational time, we created larger and smaller arbitrary porphyrin oligomers by adding or deleting other molecules in the 96-mer.

For all systems, eqs 5a and 5b were used to generate the spectra within TFDP using Γ corresponding to the 10 nm

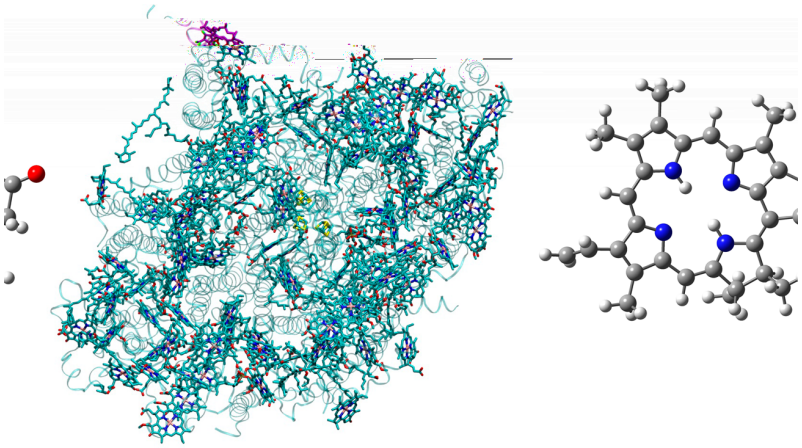


Figure 3. Cyanobacterial photosystem I (from ref 67) with 96 porphyrin residues (left), and a simplified porphyrin monomer (right).

DOI: 10.1021/acs.jctc.5b00136
J. Chem. Theory Comput. 2015, 11, 2210−2220

bandwidth. Likewise, the TDDFT and TDC spectra were generated from the calculated dipole and rotational strengths using Lorentzian bands (full width at half-maximum of 10 nm).

RESULTS AND DISCUSSION

NMA Dipole Polarization Density. In Figure 4, the dipole polarization density (ρ) is plotted as simulated for the NMA molecule at the B3LYP/6-311++G** level for two excitation frequencies corresponding to wavelengths of 138 and 222 nm.

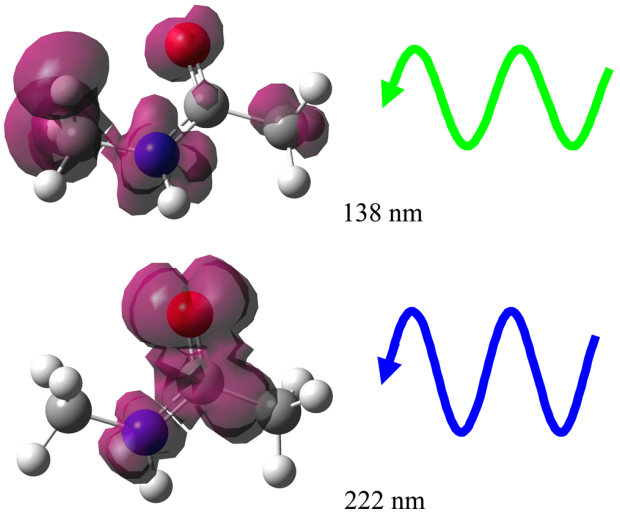


Figure 4. Isodensity surfaces of the dipole polarization density (ρ, see eq 7) for NMA calculated for two excitation wavelengths.

The density corresponds well to the expected properties of the amide chromophore[71] (e.g., predicted theoretically[72] or detected by Raman resonance experiments[73]). For 200 nm, the amide group (HNCO atoms) itself is most spectroscopically active due to the conjugated π-electrons and oxygen electron lone pairs. On the other hand, at 138 nm, it is the aliphatic part of the molecule that is most excited. The dipole polarization density can thus be used as a tool to localize molecular transitions. It justifies the transfer concept based on the locality of some molecular electronic properties and enables us to localize and visualize the relevant electronic transitions.

Distance Dependence of the Dimer Spectra. In Figure 5 (part A), absorption and ECD spectra for three NMA dimers with separation distances of 4.5, 6.5, and 8.5 Å are simulated by TFDP and compared to the benchmark TDDFT computation. As expected, for 4.5 Å in which excited electronic states involving charge-transfer between the NMA molecules significantly contribute to spectral intensities, ECD spectra obtained by the TFDP and TDDFT methods are rather different. Only some spectral features are approximately reproduced, such as the mostly negative signal within 180−200 nm, a positive one within 145−180 nm, and a negative lobe at 135 nm.

Much smaller differences can be found between the TFDP and TDDFT absorption spectra. This reflects the fact that, for separations allowing for a meaningful transfer, the absorption of any chromophore system is always nearly a sum of individual monomers (even for very close separations, the integrated absorption is always proportional to the number of electrons contained in the system;[4] these cases, however, are not investigated here). The main factor determining the splitting of energy levels is the dipole−dipole interaction, where the energy

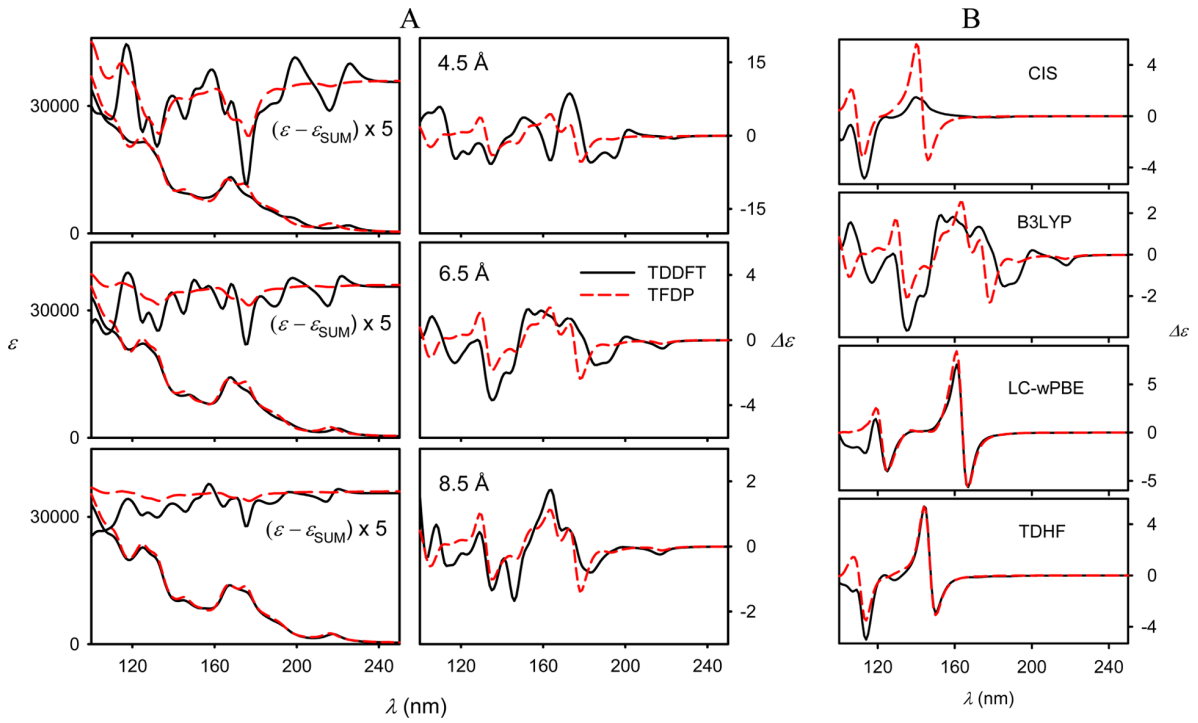


Figure 5. A: Absorption (left) and ECD (right) spectra of an NMA dimer with monomers separated by 4.5, 6.5, and 8.5 Å comparing TDDFT and TFDP. In the absorption panels, the intensity differences with respect to a plain sum of two NMA monomers are plotted as well ($\varepsilon = \varepsilon_{SUM}$, multiplied by 5). B: ECD spectra of the 6.5 Å NMA dimer modeled at the CIS, B3LYP, LC-wPBE, and TDHF levels.

DOI: 10.1021/acs.jctc.5b00136
J. Chem. Theory Comput. 2015, 11, 2210−2220

difference is approximately $\mu_1\mu_2/r^3$, where r is the chromophore distance. For the 215 nm NMA transition, for example, the calculated transition dipole moments were $\mu_1 = \mu_2 = 0.51$ debye, which for the 4.5 Å dimer results in a wavelength splitting of only 0.07 nm. This is further reduced to 0.02 and 0.01 nm for the 6.5 and 8.5 Å dimers, respectively. Note that the mechanism providing the ECD signal is fundamentally different; the monomer ECD is zero, and ECD is thus fundamentally more sensitive to proper modeling of the interchromophore interaction.

For the absorption changes to be emphasized, differences with respect to the sum of the two monomers is inset in the absorption panels. A closer inspection thus reveals small variations, such as very different positions of the longest-wavelength peak (216 nm by TFDP vs 225 nm by TDDFT), and the signal split around 170 nm predicted by TFDP but not by TDDFT. Also, the integral (average) TDDFT absorption intensity in the displayed region is slightly higher than for TFDP by ~4% in the displayed interval of wavelengths.

At 6.5 Å, the absorption spectra yielded by the two methods are nearly identical. The differences between the TDDFT and TFDP spectra are smaller than for 4.5 Å for both absorption and CD. The TFDP curve mimics several TDDFT CD features reasonably well, such as the positive/negative signal around 130 nm, positive signal within 150−170 nm, and a negative peak at 217 nm. Around 180 nm, although both approaches provide a negative lobe, the exact positions of the minimum differ (179 nm for TFDP vs 186 nm for TDDFT). The TDDFT method additionally predicts a shoulder at 191 nm, which is not reproduced by TFDP. For the 6.5 Å distance, the default B3LYP results are compared to the CIS, LC-wPBE, and TDHF calculations in Figure 5B. The CAM-B3LYP functional provided results very similar to LC-wPBE and are not shown. We can see that for cases in which the charge-transfer states are supposed to be limited[61] (LC-wPBE, TDHF), the TFDP transfer results are closer to those obtained by the reference quantum-mechanical model; on the other hand, the CIS rigid molecular orbital treatment[60] results in rather delocalized states, and the transfer performs less successfully.

For the distance of 8.5 Å, a direct interaction of the NMA electronic clouds (approximately delimited by the van der Waals radii, see Figure 2) is almost excluded, and the influence of intermolecular charge-transfer transitions is limited. The TFDP and TDDFT spectra in Figure 5 are thus even more similar, although minor differences still occur. For example, the negative TFDP CD signal at 146 nm is predicted by TDDFT as a shoulder only. Overall, however, the TFDP seems to be suitable for simulating spectra of distant chromophores.

The transferability or charge-transfer phenomena preventing it can be indicated by the shapes of molecular orbitals involved in the electronic transitions. This is shown in Figure 6, where the lowest-unoccupied molecular orbital (LUMO) in the 6.5 Å dimer largely extends over the intermonomer space, thus facilitating the charge-transfer phenomena. On the other hand, for the separation of 8.5 Å, LUMO is predominantly localized on individual NMA molecules, and a direct quantum-chemical interaction is limited.

Comparison of the Transfer Models. For the 8.5 Å NMA dimer, absorption and ECD spectra simulated with the TDC, TFDP, $TFDP_d$, $TFDP_{de}$, and TDDFT approaches are compared in Figure 7. We can see that the TFDP transfer provides spectral shapes that are very close to the TDC model. Unlike TDC, TFDP also comprises contributions of the quadrupole and magnetic-dipole containing terms (see eq 8); however, their contribution is rather limited for distant chromophores.[3,9] As expected, all approaches provide very

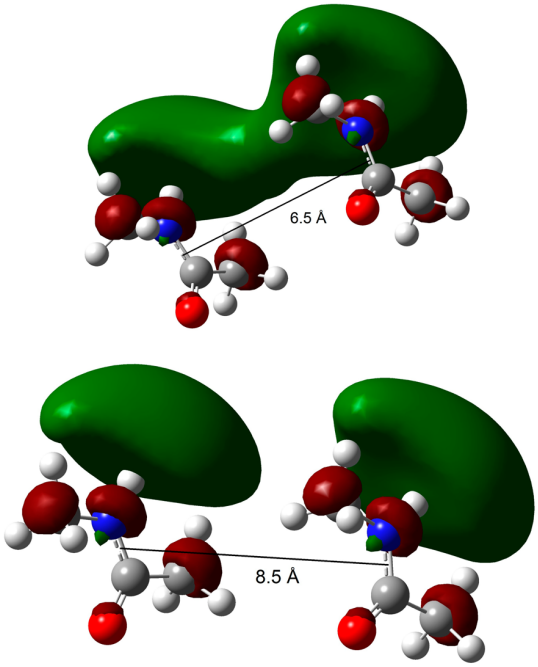


Figure 6. Lowest-unoccupied molecular (LUMO) orbital in the NMA dimer for monomer−monomer separation distances of 6.5 and 8.5 Å.

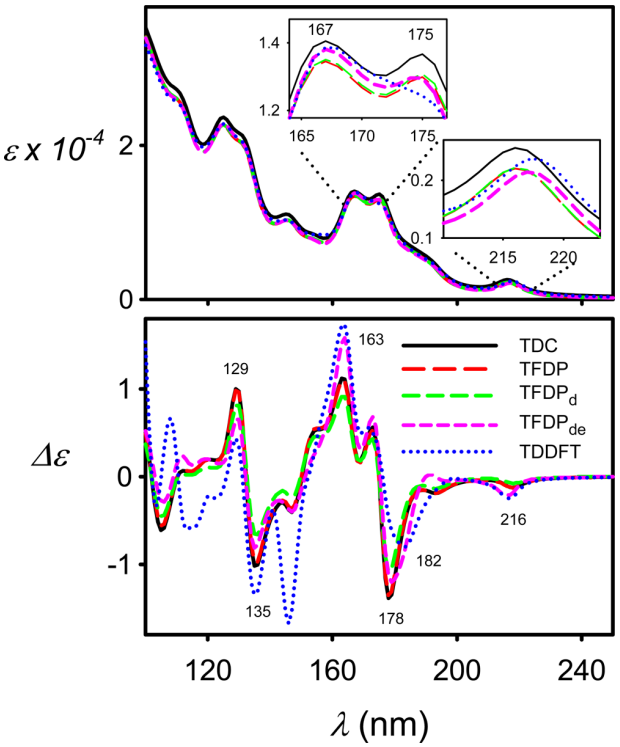


Figure 7. Absorption (top) and ECD (bottom) spectra of the NMA dimer (8.5 Å) as obtained by the TDC, TFDP, $TFDP_d$, $TFDP_e$, and TDDFT approaches; two absorption regions are enhanced in the insets.

DOI: 10.1021/acs.jctc.5b00136
J. Chem. Theory Comput. 2015, 11, 2210−2220

similar absorption intensities; the TDC approach is only slightly higher than for those obtained by other models.

$TFDP_d$ ECD spectra differ only slightly from those of TFDP; intensities of the peaks at 129, 135, 163, and 178 nm differ by ~15%. The $TFDP_{de}$ variant (accounting for the Mulliken charges mimicking the second chromophore) differs more from TFDP and often favorably approaches the TDDFT standard. A subtle change occurs at the longest-wavelength absorption peak that shifts from 216 nm ($TFDP_d$) to 217 nm ($TFDP_{de}$), which is closer to the TDDFT result (218 nm). Around 216 nm, the negative ECD peak given by $TFDP_{de}$ is nearly identical to that from the TDDFT curve. $TFDP_{de}$ also gives the best match for the negative ECD peak at 182 nm. A spectacular improvement owing to the $TFDP_d \rightarrow TFDP_{de}$ step occurs at 163 nm, where the ECD intensity nearly doubles and closely approaches the TDDFT value. For the shorter-wavelength region (<160 nm) comprising high-energetic delocalized excitations, the transfer approach is apparently not that suitable; however, these transitions are irrelevant for most biospectroscopic applications.[71,74] The 145 nm negative TDDFT ECD band, for example, is a charge-transfer transition from the HOMO−2 π orbital localized on one NMA molecule to a LUMO+17 orbital delocalized (almost Rydberg-type) over the whole dimer. Therefore, none of the transfer approaches can reproduce it.

The effect of charge variations in the $TFDP_e$ model is shown in Figure 8. The absorption and ECD spectra of the NMA dimer (8.5 Å) are plotted as obtained with the Mulliken, CHelp, HLY, and MK charges used for the monomer calculation and compared to the TFDP and TDDFT results. All of the charges appear to give very similar spectral changes. Closer inspection reveals that the effects of the CHelp, HLY, and MK charges are very similar, whereas the Mulliken results are the most distinct. In fact, the MU charges in $TFDP_e$ provide the best agreement with TDDFT, as can be documented by the negative ECD band at 182 nm, the positive signal at 200 nm, and the absorption peak at 218 nm. Admittedly, the variations are rather subtle. Still, this is a somewhat surprising result and may be attributed to an accident as the MU population scheme, unlike the others, was not primarily designed to reproduce the electrostatic field of the NMA molecule.

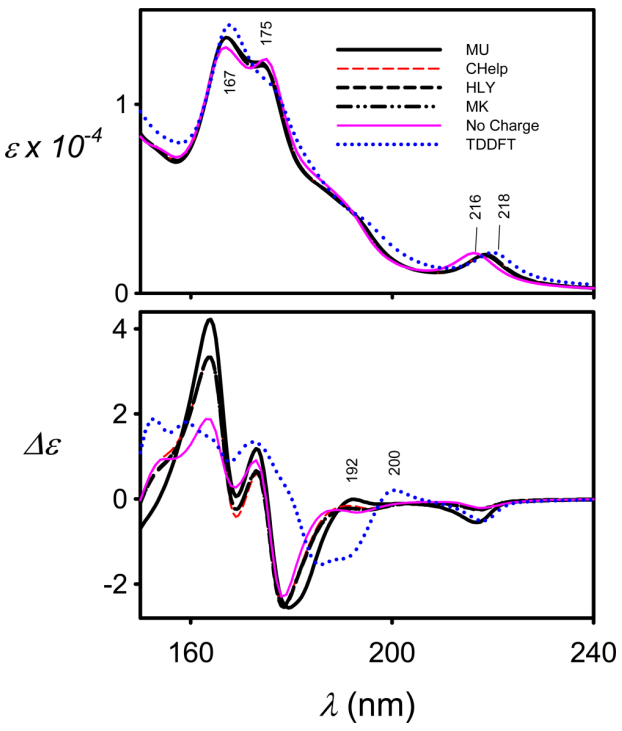


Figure 8. Absorption (top) and ECD (bottom) spectra of the NMA dimer (monomer separation distance of 8.5 Å) as obtained by the $TFDP_e$ model using the MU, CHelp, HLY, and MK charges (as well as no charge) in the transfer as well as the TDDFT result.

Computational times needed to generate the NMA dimer (8.5 Å) electronic spectra are listed in Table 1. The TDC, TFDP, and $TFDP_d$ methods involve computation of electronic excited states for only one molecule. In NMA, molecular symmetry (C_S) can additionally be used for a minor acceleration, reducing the CPU time to 51 min. In $TFDP_e$, the lack of symmetry and the presence of charges almost doubled the computational time to 98 min because each monomer is unique. Still, significant time savings is achieved compared to the full TDDFT calculation lasting 30 h.

Table 1. Computational Times[a] Needed for Various Transfer Models (2000 Dimer Transitions, B3LYP/6-311++G)**

model	time
TDC	51 min
TFDP	51 min
$TFDP_d$	51 min
$TFDP_e$	98 min (2×)
TDDFT	30 h

[a]Intel Xeon CPU E5-2670 0 2.60 GHz.

Note that we chose to test the TFDP method in a wide wavelength interval, and a relatively high number of electronic excited states needed to be calculated to model the whole spectrum. For the NMA dimer, the B3LYP/6-311++G** method provides 396 states within the presented range above 100 nm. Because of the slow decay of the Lorentzian function and the contribution of both real and imaginary polarizability components (ss eqs 3, 8a, and 8b), states below 100 nm slightly contribute to the resultant intensities as well. In practice, however, states below 180 nm are rarely of interest because of the limits of typical spectrometers.

The porphyrin dimer (only CD spectra are plotted in Figure 9) behaves similarly to the NMA dimer. Unlike NMA, however, the inherent chirality of porphyrin (Figure 3) provides a residual CD signal of the monomer ("plain sum" spectra in Figure 9). For a small chromophore separation (6.4 Å, left-hand side), the TDC and TFDP methods are not appropriate. Although they reproduce the increase in intensity, they yield rather unrealistic CD spectral shapes, especially for long-wavelength (~560 nm) transitions. For greater separations (right-hand side), however, both of the approximation methods provide reasonable results. The TFDP CD spectrum is more similar to that of TDDFT, reproducing the fine features around 400 nm and the negative band at 570 nm better.

As indicated in the Introduction, a question arises when TFDP and similar semiclassical methods converge to the exact quantum-mechanical results. Although this topic is very complex and goes beyond the scope of the present study, the results on the NMA and porphyrin dimers suggest two principle limitations of the transfer. First, for close chromophores, the strong nonelectrostatic interaction between

DOI: 10.1021/acs.jctc.5b00136
J. Chem. Theory Comput. 2015, 11, 2210−2220

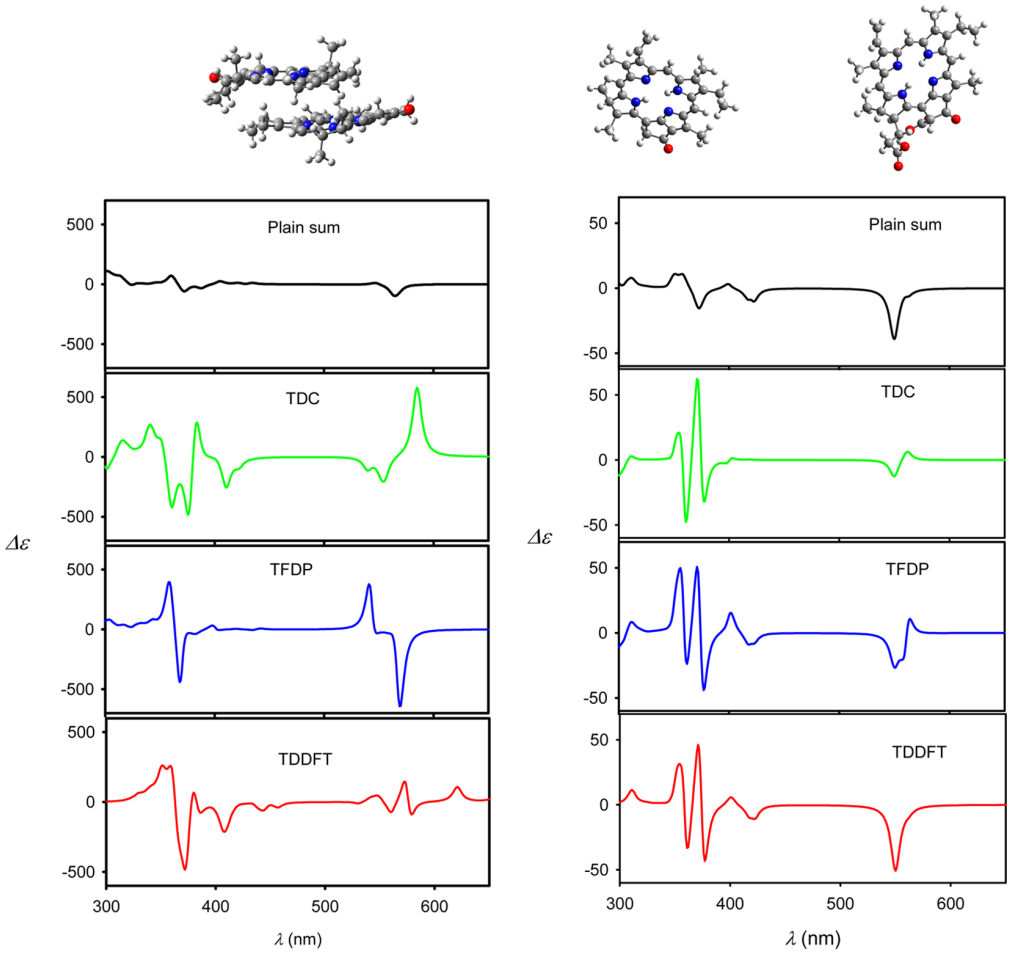


Figure 9. ECD spectra generated for two porphyrin dimers by the four approaches; the distances between the mass centers of the porphyrin monomers were 6.4 Å (left) and 23.3 Å (right).

them cannot be easily described (e.g., using dipole−dipole approximation) and requires the full quantum-mechanical treatment. It is, nevertheless, interesting that even in this case some features in the spectrum can be explained with the empirical models (e.g., the "±" couplet around 360 nm in Figure 9, reproduced by both TFDP and TDDFT). For distant chromophores, the electrostatic interaction between them seems to prevail, which justifies the transfer approach. However, the classical (transfer) and quantum models do not need to converge to the same limit either because of the possibility of charge transfer and delocalized Rydberg-like states. We found it difficult to model the "infinite distance" case because of the limited precision of the quantum-chemical methods optimized for isolated molecules. Fortunately, the classical treatment seems to describe the most important spectral features reasonably well; thus, we find it useful to explore the potential of the transfer methods, such as TFDP. They are computationally efficient, make it easier to understand the light-scattering and absorption phenomena, and are amenable to accuracy improvement in the future (e.g., using more advanced chromophore-embedding schemes).

Finally, we compare the TDC and TFDP CD and absorption spectra generated for the photosystem I-derived 96-mer in Figure 10. We provide this as an example of a potential application only; however, it is noteworthy that the theoretical

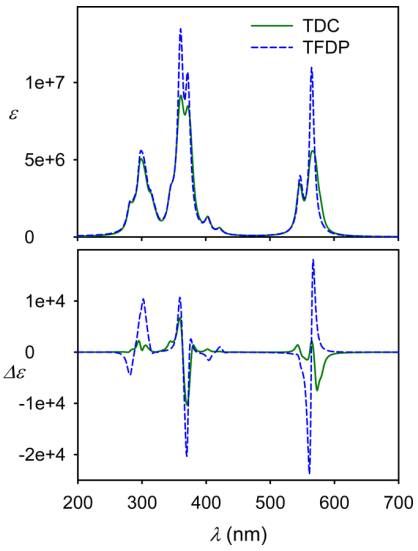


Figure 10. ECD and absorption spectra generated for the porphyrin 96-mer of the photosystem I unit shown in Figure 3

spectra are compatible with an earlier experimental observation in similar systems.[75−77] For illustration, we replot the

DOI: 10.1021/acs.jctc.5b00136
J. Chem. Theory Comput. 2015, 11, 2210−2220

experimental absorption and ECD spinach photosystem spectra from ref 77 in Figure 11. The simplified model used in the

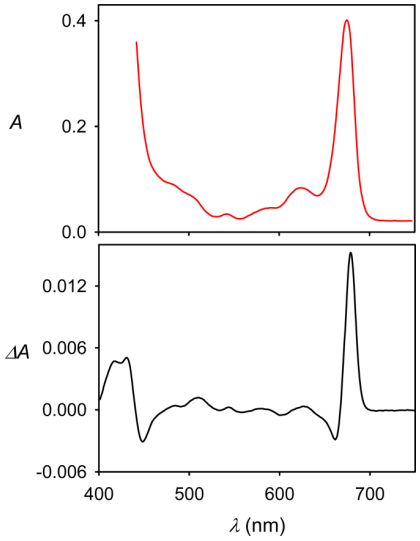


Figure 11. Absorption (A) and ECD (ΔA) experimental spectra of a spinach photosystem according to ref 77. The ECD signal in the Soret region (<400 nm) may be obscured by the large absorption signal.

present study qualitatively corresponds to the principle experimental features (i.e., the dominance of the Soret; calculated at 380 nm, experimentally around 420 nm), Q (580/680 nm) porphyrin absorption bands,[78] and a relatively large dissymmetry factor (ratio of CD to absorption) within 10^{-2} to 10^{-3}, although a more detailed comparison is not relevant at this stage. The TDC and TFDP models give a similar CD signal in the Soret region (Figure 10), whereas they differ in the Q-bands. The difference can be attributed to many close-distance porphyrin pairs in the photosystem in which case the results from the transfer methods become less predictable, as shown in Figures 5 and 9.

An interesting insight into the different nature of the TFDP and TDC methods is provided by the dependence of computational time on the system size plotted in Figure 12. Times of TDDFT calculation needed to estimate monomer properties were excluded as they are approximately the same for both methods. From Figure 12 it is obvious that TFDP is not only more time-efficient compared to TDC, but it also has quite different "scaling" with respect to the number of

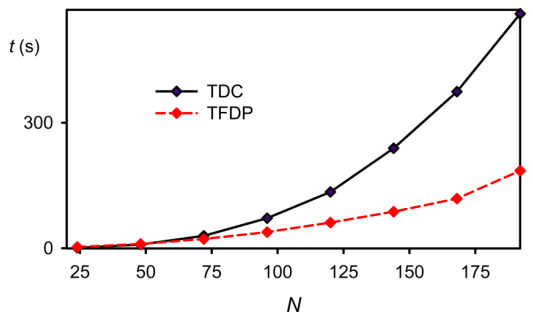


Figure 12. Time required for the TDC and TFDP computation of ECD spectra dependent on the number of porphyrin molecules in an arbitrary oligomer system based on photosystem I.

monomer units (N). Indeed, the diagonalization of the interaction Hamiltonian needed for TDC inherently scales as N^3.[79] The computational time also increases sharply due to the many electronic transitions present in the relatively large monomer units. On the other hand, the pairwise interaction described by eqs 8a and 8b for TFDP provides the N^2 scaling, which can further be reduced to almost linear dependence by a preselection of interacting pairs on the basis of interchromophore distances. The state splitting obtained in TDC by the diagonalization is in the TFDP encoded in the shape of the frequency functions.

CONCLUSIONS

In search of transferable quantities that would be widely applicable to simulate optical properties of large molecular systems, we investigated the transfer of frequency-dependent polarizabilities (TFDP). This was significantly facilitated by an efficient TDDFT implementation providing dynamic (frequency-dependent) electric dipole and higher order polarizabilities in a wide range of transition frequencies needed for practical spectroscopy. The TFDP method provided results at least comparable with a similar semiempirical model (TDC). Fine transfer parameters could additionally be varied to account for the chromophore environment and fine polarizability localization, as documented for the $TFDP_d$ and $TFDP_e$ variants. The scaling of TFDP computational time with respect to the system size is also much more favorable than for TDC. Therefore, the polarizability transfer appears to be the method of choice for approximate simulations of electronic spectra for a wide range of large molecular systems.

AUTHOR INFORMATION

Corresponding Author

*E-mail: bour@uochb.cas.cz.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The present study was supported by the Academy of Sciences (M200551205) and Grant Agency (P208/11/0105, 13-03978S, and 15-09072S) of the Czech Republic. We thank Dr. R. Pelc for useful comments on the manuscript. We used computational sources covered by MetaCentrum projects number CZ.1.05/3.2.00/08.0144 and LM2010005.

REFERENCES

(1) Bouř, P.; Sopková, J.; Bednárová, L.; Maloň, P.; Keiderling, T. A. *J. Comput. Chem.* **1997**, *18*, 646−659.
(2) Bieler, N. S.; Haag, M. P.; Jacob, C. R.; Reiher, M. *J. Chem. Theory Comput.* **2011**, *7*, 1867−1881.
(3) Yamamoto, S.; Li, X.; Ruud, K.; Bouř, P. *J. Chem. Theory Comput.* **2012**, *8*, 977−985.
(4) Barron, L. D. *Molecular Light Scattering and Optical Activity*; Cambridge University Press: Cambridge, 2004.
(5) Choi, J. H.; Kim, J. S.; Cho, M. *J. Chem. Phys.* **2005**, *122*, 174903.
(6) Stenner, R. S. *J. Phys. Chem. B* **2012**, *116*, 4141−4153.
(7) Yamamoto, S.; Kaminský, J.; Bouř, P. *Anal. Chem.* **2012**, *84*, 2440−2451.
(8) Andrushchenko, V.; Wieser, H.; Bouř, P. *J. Phys. Chem. B* **2004**, *108*, 3899−3911.
(9) Yamamoto, S.; Bouř, P. *Collect. Czech. Chem. Commun.* **2011**, *76*, 567−583.
(10) Holzwarth, G.; Chabay, I. *J. Chem. Phys.* **1972**, *57*, 1632−1635.
(11) Rabenold, D. A. *J. Chem. Phys.* **1980**, *12*, 5942−5946.

DOI: 10.1021/acs.jctc.5b00136
J. Chem. Theory Comput. 2015, 11, 2210−2220

(12) Hirst, J. D. *J. Chem. Phys.* **1998**, *109*, 782−788.
(13) Gilbert, A. T. B.; Hirst, J. D. *J. Mol. Struct.: THEOCHEM* **2004**, *675*, 53−60.
(14) Besley, N. A.; Oakley, M. T.; Cowan, A. J.; Hirst, J. D. *J. Am. Chem. Soc.* **2004**, *126*, 13502−13511.
(15) Jiang, J.; Abramavisius, D.; Bulheller, B. M.; Hirst, D. M.; Mukamel, S. *J. Phys. Chem. B* **2010**, *114*, 8270−8277.
(16) Woody, R. W. *Monatsh. Chem.* **2005**, *136*, 347−366.
(17) Koslowski, A.; Sreerama, N.; Woody, R. W. Theoretical Approach To Electronic Optical Activity. In *Circular Dichroism Principles and Applications*; Nakanishi, K., Berova, N., Woody, R. W., Eds. Wiley-VCH: New York, 2000; pp 97−132.
(18) Bulheller, B. M.; Miles, A. J.; Wallace, B. A.; Hirst, J. D. *J. Phys. Chem. B* **2008**, *112*, 1866−1874.
(19) Jamorski, C.; Casida, M. E.; Salahub, D. R. *J. Chem. Phys.* **1996**, *104*, 5134−5147.
(20) Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2002**, *116*, 7433−7447.
(21) Furche, F.; Ahlrichs, R.; Wachsmann, C.; Weber, E.; Sobanski, A.; Vögtle, F.; Grimme, S. *J. Am. Chem. Soc.* **2000**, *122*, 1717−1724.
(22) Bak, K. L.; Hansen, A. E.; Ruud, K.; Helgaker, T.; Olsen, J.; Jørgensen, P. *Theor. Chim. Acta* **1995**, *90*, 441−458.
(23) Autschbach, J.; Ziegler, T.; van Gisbergen, S. J. A.; Baerends, E. J. *J. Chem. Phys.* **2002**, *116*, 6930−6940.
(24) Rogers, D. M.; Hirst, J. D. *Chirality* **2004**, *16*, 234−243.
(25) Hirst, J. D.; Colella, K.; Gilbert, A. T. B. *J. Phys. Chem. B* **2003**, *107*, 11813−11819.
(26) Rogers, D. M.; Besley, N. A.; O'Shea, P.; Hirst, J. D. *J. Phys. Chem. B* **2005**, *109*, 23061−23069.
(27) Kaminský, J.; Kubelka, J.; Bouř, P. *J. Phys. Chem. A* **2011**, *115*, 1724−1742.
(28) Cammi, R.; Corni, S.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **2005**, *122*, 104513.
(29) Mennucci, B.; Cappelli, C.; Cammi, R.; Tomasi, J. *Chirality* **2011**, *23*, 717−729.
(30) Bouř, P. *J. Phys. Chem. A* **1999**, *103*, 5099−5104.
(31) Long, D. A. *Proc. R. Soc. London, Ser. A* **1953**, *217*, 203.
(32) Barron, L. D.; Clark, B. P. *J. Raman Spectrosc.* **1982**, *13*, 155−159.
(33) Barron, L. D.; Escribano, J. R.; Torrance, J. F. *Mol. Phys.* **1986**, *57*, 653−660.
(34) Prasad, P. L.; Nafie, L. A. *J. Chem. Phys.* **1979**, *70*, 5582−5588.
(35) Olsen, J. M.; Aidas, K.; Kongsted, J. *J. Chem. Theory Comput.* **2010**, *6*, 3721−3734.
(36) List, N. H.; Beerepoot, M. T. P.; Olsen, J. M. H.; Gao, B.; Ruud, K.; Jensen, H. J. A.; Kongsted, J. *J. Chem. Phys.* **2015**, *142*, 034119.
(37) Thellamurege, N. M.; Hirao, H. *J. Phys. Chem. B* **2014**, *118*, 2084−2092.
(38) Droz, M. H.; Zhou, X.; Shedge, S. V.; Wesolowski, T. A. *Theor. Chem. Acc.* **2014**, *133*, 1405.
(39) Nasluzov, V. A.; Ivanova, E. A.; Shor, A. M.; Vayssilov, G. N.; Birkenheuer, U.; Rösch, N. *J. Phys. Chem. B* **2003**, *107*, 2228−2241.
(40) Sreerama, N.; Woody, R. W. Circular dichroism of peptides and proteins. In *Circular Dichroism Principles and Applications*; Nakanishi, K., Berova, N., Woody, R. W., Eds. Wiley-VCH: New York, 2000; pp 601−620.
(41) Keiderling, T. A. Circular Dichroism. In *Circular Dichroism: Principles and Applications*, 2nd ed.; Berova, N., Nakanishi, K., Woody, R. W., Eds. Wiley: New York, 2000; pp 621−666.
(42) Johnson, W. C. CD of Nucleic Acids. In *Circular Dichroism: Principles and Applications*; Berova, N., Nakanishi, K., Woody, R. W., Eds. Wiley-VCH: New York, 2000; pp 703−718.
(43) King, S. M.; Johnson, W. C. *Proteins* **1999**, *35*, 313−320.
(44) Johnson, W. C. *Proteins* **1999**, *35*, 307−312.
(45) McCann, D. M.; Stephens, P. J. *J. Org. Chem.* **2006**, *71*, 6074−6098.
(46) Nafie, L. A.; Freedman, T. B. Vibrational optical activity theory. In *Circular Dichroism. Principles and Applications*; 2nd ed.; Berova, N., Nakanishi, K., Woody, R. W., Eds. Wiley-VCH: New York, 2000; pp 97−131.
(47) Bouř, P. *Chem. Phys. Lett.* **1998**, *288*, 363−370.
(48) Charney, E. *The Molecular Basis of Optical Activity*. Wiley-Interscience: New York, 1979.
(49) Štěpánek, P.; Bouř, P. *J. Comput. Chem.* **2013**, *34*, 1531−1539.
(50) Štěpánek, P.; Bouř, P. *J. Comput. Chem.* **2015**, *36*, 723−730.
(51) Bouř, P.; Keiderling, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 9100−9105.
(52) Tinoco, I. *Radiat. Res.* **1963**, *20*, 133.
(53) Zhong, W.; Gulotta, M.; Goss, D. J.; Diem, M. *Biochemistry* **1990**, *29*, 7485−7491.
(54) Andrushchenko, V.; Bouř, P. *J. Comput. Chem.* **2008**, *29*, 2693−2703.
(55) Janesko, B. G.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 124704.
(56) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision D01; Gaussian, Inc.: Wallingford, CT, 2009.
(57) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(58) Šebek, J.; Gyurscik, B.; Šebestík, J.; Kejík, Z.; Bernárová, L.; Bouř, P. *J. Phys. Chem. A* **2007**, *111*, 2750−2760.
(59) Šebek, J.; Kejík, Z.; Bouř, P. *J. Phys. Chem. A* **2006**, *110*, 4702−4711.
(60) Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *J. Phys. Chem.* **1992**, *96*, 135−149.
(61) Yanai, T.; Tew, D.; Handy, N. *Chem. Phys. Lett.* **2004**, *393*, 51−57.
(62) Vydrov, O. A.; Scuseria, G. E.; Perdew, J. P. *J. Chem. Phys.* **2007**, *126*, 154109.
(63) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 2343−2346.
(64) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129−145.
(65) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894−905.
(66) Hu, H.; Lu, Z.; Yang, W. *J. Chem. Theory Comput.* **2007**, 1004−1013.
(67) Jordan, P.; Fromme, P.; Witt, H. T.; Klukas, O.; Saenger, W.; Krauß, N. *Nature* **2001**, *411*, 909−917.
(68) Hudecová, J.; Hopmann, K. H.; Bouř, P. *J. Phys. Chem. B* **2012**, *116*, 336−342.
(69) Bouř, P.; Keiderling, T. A. *J. Chem. Phys.* **2002**, *117*, 4126−4132.
(70) Bouř, P. *Collect. Czech. Chem. Commun.* **2005**, *70*, 1315−1340.
(71) Berova, N.; Nakanishi, K.; Woody, R. W. *Circular Dichroism Principles and Applications*; Wiley-VCH: New York, 2000.
(72) Goldmann, E.; Asher, S. A.; Mukamel, S. *Phys. Chem. Chem. Phys.* **2001**, *3*, 2893−2903.
(73) Xiong, K.; Punihaole, D.; Asher, S. A. *Biochemistry* **2012**, *51*, 5822−5830.
(74) Johnson, W. C., Jr. Circular dichroism instrumentation. In *Circular Dichroism and the Conformational Analysis of Biomolecules*, Fasman, G. D., Ed. Plenum Press: New York, 1996; pp 635−652.
(75) Harada, J.; Mizoguchi, T.; Tsukatani, Y.; Noguchi, M.; Tamiaki, H. *Sci. Rep.* **2012**, 671.
(76) Shubin, V. V.; Roegner, M.; El-Mohsnawy, E.; Terekhova, I. V.; Schlodder, E.; Karapetyan, N. V. *Appl. Biochem. Microbiol.* **2010**, *46*, 299−307.
(77) Kropacheva, T. N.; Germano, M.; Zucchelli, G.; Jennings, R. C.; Gorkom, H. J. *Biochim. Biophys. Acta* **2005**, *1709*, 119−126.

DOI: 10.1021/acs.jctc.5b00136
J. Chem. Theory Comput. 2015, 11, 2210−2220

(78) Gouterman, M. *J. Mol. Spectrosc.* **1961**, *6*, 138−163.
(79) Press, W. H.; Teukolsky, S. A.; Vetteling, W. T.; Flannery, B. P. *Numerical Recipes in Fortran*, 2nd ed.; Cambridge University Press: New York, 1992.

DOI: 10.1021/acs.jctc.5b00136
J. Chem. Theory Comput. 2015, 11, 2210−2220

10 Other Publications

10.1 Appendix H – *PCCP* 2016, 18, 2130.

PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2016, **18**, 2130

Simulation of Raman optical activity of multi-component monosaccharide samples†

Adéla Melcrová,[ab] Jiří Kessler,[a] Petr Bouř*[a] and Jakub Kaminský*[a]

Received 14th July 2015,
Accepted 8th December 2015

DOI: 10.1039/c5cp04111b

www.rsc.org/pccp

Determination of the saccharide structure in solution is a laborious process that can be significantly enhanced by optical spectroscopies. Raman optical activity (ROA) spectra are particularly sensitive to the chirality and conformation. However, the interpretation of them is largely dependent on computational tools providing a limited precision only. To understand the limitations and the link between spectral shapes and the structure, in the present study we measured and interpreted using a combination of molecular dynamics (MD) and density functional theory (DFT) Raman and ROA spectra of glucose and mannose solutions. Factors important for analyses of mixtures of conformers, anomers, and different monosaccharides are discussed as well. The accuracy of the simulations was found to be strongly dependent on the quality of the hydration model; the dielectric continuum solvent model provided lower accuracy than averaging of many solvent–solute clusters. This was due to different conformer weighting rather than direct involvement of water molecules in scattering recorded as ROA. However, the cluster-based simulations also failed to correctly reproduce the ratios of principal monosaccharide forms. The best results were obtained by a combined MD/DFT simulation, with the ratio of α- and β-anomers and the $-CH_2OH$ group rotamers determined experimentally by NMR. Then a decomposition of experimental spectra into calculated subspectra provided realistic results even for the glucose and mannose mixtures. Raman spectra decomposition provided a better overall accuracy (~5%) than ROA (~10%). The combination of vibrational spectroscopy with theoretical simulations represents a powerful tool for analysing the saccharide structure. Conversely, the ROA and Raman data can be used to verify the quality of MD force fields and other parameters of computational modeling.

Introduction

Raman optical activity (ROA), a small difference in scattering of right- and left-circularly polarized light, is a useful probe of structure, applicable to a wide range of chiral molecules. Since its discovery[1] it has been used for small molecules, proteins, nucleic acids and whole viruses.[2–4] The spectra are often complex and their interpretation not trivial at all. Fortunately, a steady development of theoretical chemistry in the last few decades provided useful interpretation tools and significantly stimulated the experimental field. The possibility to simulate Raman and ROA intensities by available quantum chemical codes[5,6] helps to shed light on the behavior of a wide range of molecular systems.[7–16]

Carbohydrates are known to support cellular structures (*e.g.* bacterial or plant cell walls) and participate in numerous events in living organisms. Long-chain glycosaminoglycans fill the extracellular space, influence angiogenesis, immunosuppression, cell differentiation, ovulation, embryogenesis, and regeneration. Small sugars act as anti-inflammatory and angiogenic agents, and support immune and signaling systems.[17–19] However, current knowledge about their three-dimensional structure and its relation to the function is far from complete and standard methods for structural characterization of biomolecules often fail for carbohydrates.[20] Some saccharides are hard to crystallize or provide poor NMR spectra because of peak overlaps and broadening due to fast relaxation. Sugars also typically lack active chromophores needed for absorption or electronic circular dichroism (ECD) measurements in the usual ultraviolet and visible regions (~190–800 nm). Such chromophores may be introduced synthetically, but this may perturb the saccharide geometry. Saccharides are typically studied in an aqueous environment causing many troubles in infrared techniques. In this context, ROA spectroscopy appears to be a very

[a] *Institute of Organic Chemistry and Biochemistry, AS CR, Flemingovo nám. 2, 166 10 Prague, Czech Republic. E-mail: kaminsky@uochb.cas.cz, bour@uochb.cas.cz; Fax: +420-220183-579*
[b] *J. Heyrovský Institute of Physical Chemistry, AS CR, Dolejškova 2155, 182 23 Prague, Czech Republic*

† Electronic supplementary information (ESI) available: Relative potential energy distributions calculated at the B3LYP/6-311++G**/CPCM level for **DG**; convergence of the cluster averaging for α-DG; comparison of MD and DFT geometry distributions for α-**DG**, β-**DG**, α-**DM** and β-**DM**; and the experimental spectra of **DG** and **DM** mixtures. See DOI: 10.1039/c5cp04111b

convenient alternative, as it yields relevant data on the carbohydrate structure and interactions.[21–33] For example, early studies on small saccharides[21–25,34] provided an insight into the anomeric configuration, relative configurations of the hydroxyl, ring puckering and the glycosidic linkage. In polysaccharides and glycoproteins, ROA was used to track the secondary and tertiary structure.[20,27,32,35–38]

Simulations of saccharide ROA spectra are, however, rather complicated due to molecular flexibility and strong polar interactions with the solvent.[29,39,40] Realistic description of carbohydrate-water interactions determining the conformation of the ring and hydroxyl groups is needed for the results to be useful.[29–31,39,41] Chiroptical methods are generally very sensitive to the solvent–solute interactions, and comparable solvent-induced features have also been observed in ECD, vibrational circular dichroism or optical rotation of several systems.[42–46] In our opinion, further quantitative estimates of the solvent and conformational factors are thus needed to make the spectroscopic studies of saccharides more reliable and to extract maximum information from the spectra. We show, for example, that a relatively small number of conformers is sufficient to reproduce the spectra, and that water–sugar interactions are important for conformer weighting (less so for water spectra contributions).

To understand spectra of complex polysaccharides and their conjugates with proteins, individual parts have to be carefully controlled. We thus consider increasing the number of components and investigated samples (*e.g.*, α and β anomers, as well as glucose/mannose mixture) as the first step in this direction. Anomers are stereoisomers often occurring in cyclic saccharides, differing only in the configuration of the hemiacetal (anomeric) carbon. Individual anomers of reducing sugars transform to their inverse form *via* the process of anomerization (mutarotation), and in solution an equilibrium of the two anomeric forms is usually established quite quickly. However, the anomer ratio ("α"/"β") is specific for each sugar. As the two anomers differ in reducing configuration at one chiral center only, their ROA spectra are not mere "mirror images"; the anomerization (α ↔ β) significantly changes the spectral pattern.

Another objective is to estimate as to how the ROA spectra can provide feedback on the quality of molecular dynamics (MD) force fields, accuracy of density functional theory (DFT), and solvent models. For example, available computational tools do not appear to provide completely satisfactory estimates of the anomeric and $-CH_2OH$ conformational species; experimental ratios taken from previous NMR studies led to better results. As proven by earlier studies[30,41,47] the combination of quantum and molecular dynamics is used as the best tool to tackle the flexible and hydrated sugar molecules. The sensitivity of ROA to monosaccharide structure was tested for arbitrary monosaccharide mixtures decomposed into both calculated and experimental subspectra.

Methods

Raman and ROA spectra measurements

D and L glucose and mannose (Fig. 1) were purchased from Carbosynth and dissolved in milli-Q water to concentrations of 60 mg/100 μL. The samples were left undisturbed for several hours to equilibrate various sugar forms; required equilibration times were estimated by NMR. Backscattered Raman and scattered circular polarization (SCP) ROA spectra were recorded on a ChiralRAMAN-2X (Biotools Inc.) spectrometer equipped with Opus diode-pumped solid-state laser emitting 532 nm green light. Laser power at the head was about 350 mW (approximately 2/3 reached the sample), total acquisition times were ~20 h for each sample contained in a fused silica cell (3 mm optical path, 60 μL sample volume). Fluorescence was reduced by removing sample impurities by active carbon or quenched by leaving the sample in the laser beam for an hour before measurement. Raman spectra are presented without subtracted solvent signal as well as polynomial baseline. Luminescence standard material (SRM 2243) was used for ROA and Raman intensity calibration.

Simulations of the spectra

Programs Gaussian 09,[48] MacroModel,[49] and Amber12[50] were used for the quantum and molecular dynamics simulations. Only the pyranose forms of glucose and mannose were considered as they are prevalent in solutions (>99%, see ref. 51 and 52). We used two models of sugar geometries, a systematic conformer search coupled with a DFT calculation, and averaging of MD solvent–solute clusters.

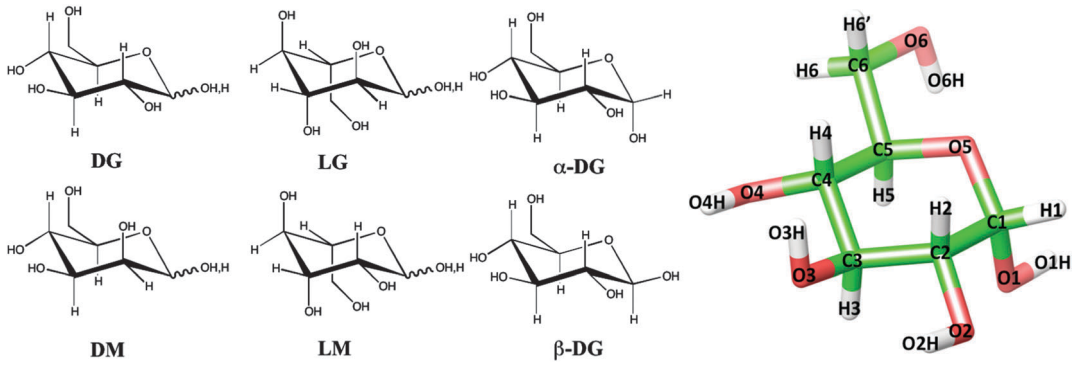


Fig. 1 Studied compounds, D- and L-glucose (**DG** and **LG**), D- and L-mannose (**DM** and **LM**), example of the α- and β-anomers for **DG**, and used atom numbering shown for α-**DG**.

The systematic conformer search was done within MacroModel using a hybrid search method, employing several simulated annealings followed by 10 000 cycles of large-scale low mode search steps. The MMFF94s force field, temperature of 1000 K, 0.5 ps annealing time, 3 fs time step, the GB/SA water solvation model, 100 kcal mol^{-1} energy window and 0.5 Å cutoff for eliminating redundant conformers were used. Further computations were done in Gaussian. Produced geometries (348 for α-**DG**, 350 for β-**DG**, 423 for α-**DM**, and 396 for β-**DM**) were optimized at the HF/6-31G**/CPCM[53] level, and the lowest-energy conformers (relative energies below 5 kcal mol^{-1}) were re-optimized at the B3LYP/6-311++G**/CPCM approximation level with the D2 dispersion correction.[54] Raman and ROA spectra were generated at the same B3LYP/6-311++G**/CPCM level for conformers with populations higher than 1%. As recommended previously,[55,56] Boltzmann populations were also calculated from the CCSD(T)/6-311++G**/CPCM single-point energies; however, the CCSD(T) populations were nearly the same, and thus only B3LYP populations are presented and were used for spectra generation.

The MD cluster averaging was based on the separate MD simulations of α and β anomers of **DG** and **DM**. One molecule of sugar was placed into a cubic box (23 Å a side) and surrounded by 400 water molecules. For the initial geometry only the $^4C_1$ chair conformation was considered as other pyranose ring conformations are too high in energy.[57–59] The GLYCAM06[60] (sugars) and TIP3P[61] (water) force fields were used. A short heating of the system from 0 to 300 K for 20 ps under *NVT* conditions and 100 ps *NpT* equilibration preceded the production run, performed for 10 ns under the *NpT* conditions at the temperature of 300 K, pressure of 1 atm, and 1 fs integration step. Snapshot geometries were saved during the production run each 1 ps and the parallel variable method[62] was used to reduce the number of snapshots to 70, which were then used to generate the spectra.

In the snapshots, only solvent molecules in the first solvation sphere were retained. This was achieved with a 2 Å cutoff distance between any solute and solvent atoms. The solute–solvent clusters were generated using our (''XSHELL'') software.[63] Geometries of the resultant clusters comprised 5–11 water molecules and were partially optimized using the normal mode vibrational coordinates with the normal mode frequency limit[64] ω_{max} of 300 cm^{-1}. The normal mode optimizations was found previously to be an excellent tool for optimization of MD snapshots.[65–67] It has been particularly developed[64,68,69] for computation of vibrational spectra, as it allows to relax the higher-frequency modes of interest, whereas the MD geometry dispersion reflecting non-zero temperature is largely conserved.

The Raman and ROA intensities were then calculated at the B3LYP/6-311++G**/CPCM level (same as for the systematic conformer search). Polarizability derivatives of water atoms were deleted before the spectral simulations, which significantly minimizes computational ROA artifacts resulting from incomplete averaging of solvent orientation.[65] (Note also that Raman experimental spectra are presented with the water background subtracted.) Backscattered Raman and SCP ROA intensities were convoluted with Lorentzian peaks of full width of half-maximum of 20 cm^{-1}, and taking into account the Boltzmann temperature factor at 300 K. As a rough indicator of the cluster averaging, a relative error $\delta(N)$ was estimated as:

$$\delta = \int_{\omega_{min}}^{\omega_{max}} |S_N(\omega) - S_M(\omega)| d\omega \Big/ \int_{\omega_{min}}^{\omega_{max}} |S_M(\omega) d\omega|, \quad (1)$$

where S_N is a spectrum from N averaged clusters ($1 \leq N \leq 70$), $M = 70$, $\omega_{min} = 200$ cm^{-1}, and $\omega_{max} = 2000$ cm^{-1}. The resultant spectra were obtained by equal weighting of all clusters, as well as selective weighting of $-CH_2OH$ rotamers according to NMR data. For the latter method, the clusters were divided into three groups according to their $-CH_2OH$ orientation, the subspectra generated with equal cluster weighting for each group separately, and the resultant spectra of rotamers were mixed according to experimental NMR ratios taken from ref. 70 and 71.

For further tests selected experimental spectra $S(\omega)$ were decomposed into normalized calculated or experimental subspectra $s_i(\omega)$ as:

$$S(\omega) = \sum_{i=1}^{n} c_i s_i(\omega), \quad (2)$$

where the coefficients were determined by minimization of $\int_{\omega 1}^{\omega 2} (S(\omega) - \sum_{i=1}^{n} c_i s_i(\omega))^2 d\omega + \alpha \sum_{i=1}^{n} (c_i - 1/n)^2 \rightarrow \min$, with ω_1 = 100 cm^{-1} and ω_2 = 1600 cm^{-1}. The parameter α = 0.00001 was introduced to prevent negative coefficient; the coefficients were normalized after the decomposition to one $\left(\sum_{i=1}^{n} c_i = 1\right)$. Then $n = 2$, because the experimental spectra were decomposed[16] into calculated spectra of free **DG** and **DM** simulated as described above. Similarly, the experimental spectra of mixtures were decomposed to experimental spectra of individual components in order to establish the accuracy of the decomposition method.

In the reviewing process, it was suggested to compare our simulations to the ''QM/MM'' approach pursued in ref. 30, 41 and 47. For this purpose, the α-**DG** spectra were simulated with various models, as for this anomer experimental data are available in ref. 72. As a simplest solvent model, the explicit water molecules in our clusters were deleted and the solute geometries fully optimized at the B3LYP/6-311++G**/CPCM level. To follow the ''QM/MM'' approach form ref. 30, 41 and 47, the Amber TIP3P[61] water force field was used for the solvent molecules in all 70 clusters, the clusters were fully optimized using the ONIOM method[73–76] and the B3LYP/6-311++G** level for the solute, and Raman and ROA obtained at the same QM/MM level. All spectra were compared with those obtained by our default QM/QM (labeled as MD approach in this work) procedure.

Results

Experimental spectra

Experimental Raman and ROA spectra are presented in Fig. 2. They are consistent with previous data;[21,24] our results additionally

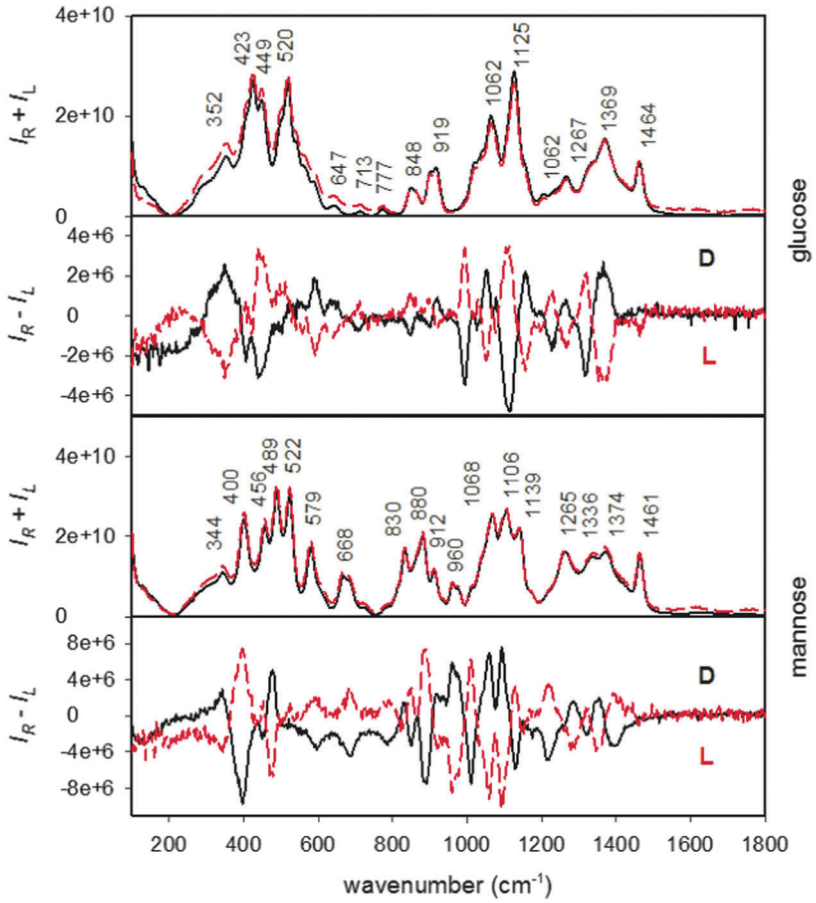


Fig. 2 Experimental Raman ($I_R + I_L$) and ROA (($I_R - I_L$)) spectra of D- and L-glucose and mannose enantiomers.

include the region below 600 cm^{-1}. The main features of mannose also resemble those in α-D-mannoside.[24] The ROA baseline below 200 cm^{-1} is affected by instrumental artifacts, the region of 200–600 cm^{-1}, however, contains distinct ROA features in intensity comparable with the higher-frequency bands.

As discussed before,[24] monosaccharide vibrational spectrum can be approximately divided into a low wavenumber (<600 cm^{-1}), anomeric (600–950 cm^{-1}), fingerprint (950–1200 cm^{-1}) and CH_2 and –COH deformation (>1200 cm^{-1}) regions. This can be complemented by calculated potential energy distributions (PEDs), some of them presented in Fig. S1 (ESI†). For example, the OH bending motion is relatively restricted to the vicinity of 1200 cm^{-1}, whereas most other coordinates contribute throughout the vibrational spectrum. Below 600 cm^{-1} torsional–deformational modes of the six-member ring prevail.

In the low wavenumber range (<600 cm^{-1}), Raman and ROA bands sensitive to the anomeric forms often comprise exo- and endocyclic torsion and bending deformations coupled with exo- and endocyclic deformations. The "true" anomeric region (600–950 cm^{-1}) typically comprises C–C and C–O stretching, C–O–C bending, and CH_2 rocking. The experimental bands at around 850 cm^{-1} were found to be characteristic for the α anomeric hydroxyl.[22–25,34] The 848 cm^{-1} **DG** Raman band appears at 880 cm^{-1} for **DM** (Fig. 2). ROA spectra exhibit even greater sensitivity to the anomeric configuration in this region.[23–25,34] β-Anomer sugars have a very small ROA signal between ~700–800 cm^{-1}, while there are several strong signals for the α-anomers.[23–25,34] Also the ROA signal of mannose (predominantly α) is stronger than for glucose in this region.

ROA intensities within 800–950 cm^{-1} are weak (Fig. 2). For **DG**, the 848 and 890 cm^{-1} negative bands correspond to axial and equatorial positions of the anomeric hydroxyl, respectively. **DM** has a very different ROA sign pattern here. In the fingerprint region there are two intense Raman bands (1062 and 1125 cm^{-1}) of glucose and three (1068, 1106 and 1139 cm^{-1}) of mannose, and rich ROA features. As discussed before[39] the conformation of the hydroxymethyl (–CH_2OH) group significantly influences Raman and ROA band intensities in this region, but ribo- and xylo-saccharides provide similar patterns. Similarly, the signal in the highest wavenumber region is quite complex, and detailed vibrational analysis can be found elsewhere.[23,24] For example, COH bending and CH_2OH twisting motions participate on the ROA intensity pattern, and the ROA signal at 1260 cm^{-1} is characteristic for β-D-glucose.[25]

Experimental *vs.* computed anomeric ratios

Although the unbiased conformational scans including systematic conformer search and molecular dynamics provide a good basis for estimation of the prevalent saccharide conformers, accurate modeling and interpretation of the spectra require a more careful

Table 1 Calculated and experimental populations (%) of α-anomeric forms of glucose and mannose

	Calculation		Experiment[d]
	α-Sum		α-Sum
DG	46[a]	44[b] (39[c])	37
DM	56[a]		67

[a] Based on B3LYP/6-311++G**/CPCM Gibbs energies. [b] Ref. 81. [c] Ref. 80. [d] Ref. 51 and 52.

consideration of the anomeric equilibrium. As previously determined by NMR (*cf.* Table 1, ref. 51 and 52) the α/β ratios in glucose and mannose significantly differ. This was explained by mutually competing anomeric and 1,3-diaxial effects.[77–79] Experimentally, the populations of the α-anomer in glucose and mannose water solutions are 37% and 67%, respectively. Our computed Boltzmann populations (46% and 56%, Table 1) based on anomer relative energies follow the experimental trend with an error similar to that observed in other studies.[80,81] However, such small inconsistencies in the anomeric ratios have a bigger effect on the spectra as documented for **DG** in Fig. 3, where Raman and ROA spectra of α- and β-forms of **DG** are presented as calculated for the prevalent $^4C_1$ main chain conformer. For example, the characteristic glucose and mannose Raman and ROA patterns at around 300 cm^{-1} largely stemming from the different α- and β-content are to some extent caused by deformation vibrations in the vicinity of the anomeric carbon. Therefore, as another option, we investigate weighting of the computed spectra using the experimental α/β form ratios.

The effect of –CH_2OH rotamers

A similar situation occurs with the hydroxymethyl group conformations (*gg*/*gt*/*tg*, see Fig. 4, top, for definition). Their calculated ratios (Table 2) differ by up to 30% from the experimental ones. The sole effect of the *gg*/*gt*/*tg* conformer ratio on the Raman and ROA spectra can be documented for glucose, for which experimental spectra of pure β-anomer were previously obtained by W. Hug *et al.*[72] In Fig. 4, lower part, we compare spectra obtained by direct averaging of 70 MD clusters (top) with those corrected for the experimental *gg*/*gt*/*tg* ratios from Table 2 (middle) and experiment (bottom). Although the –CH_2OH group rotation do not change the resultant Raman and ROA spectral patterns so much as the anomerization (*cf.* Fig. 3), several spectral features are sensitive to it. Incorporation of the experimental weights leads to a better reproduction of the relative Raman intensities of the 500/520 cm^{-1} bands, and more realistic ROA signal at around 420 cm^{-1} and within 800–1400 cm^{-1}. As expected, the ROA spectra are more sensitive to the conformational changes than the Raman scattering.

The ROA spectroscopy thus offers itself for future development and quality assessments of the MD saccharide force fields. At the same time, force field amelioration enabling to include the fine rotational effects associated with small energy differences appear as a complex problem, because of the solvent-solute interactions, *etc.* On the other hand, the six membered ring conformation seems to be rigid and relatively reliably determined by MD or the DFT/CPCM-based conformational scan; *e.g.* $^1C_4$ chairs of **DG** were higher in energy by about 12 kJ mol^{-1} (for the α anomer) and 19 kJ mol^{-1} (for the β anomer) if compared to the $^4C_1$ conformers.

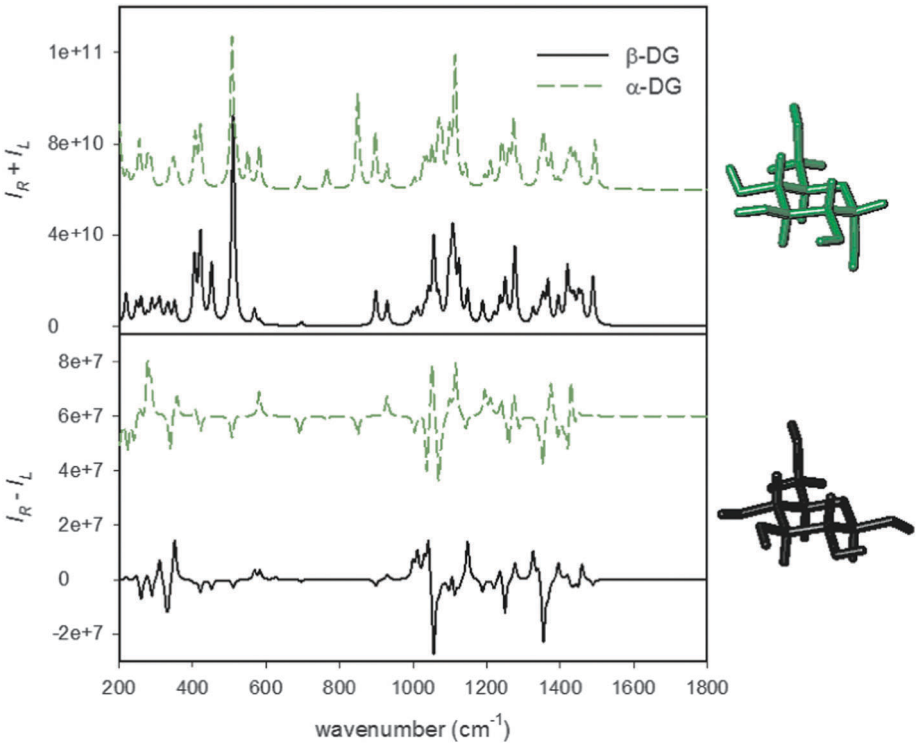


Fig. 3 Calculated (B3LYP/6-311++G**/CPCM) Raman and ROA spectra of α- and β-**DG**, for a $^4C_1$ –CH_2OH conformer.

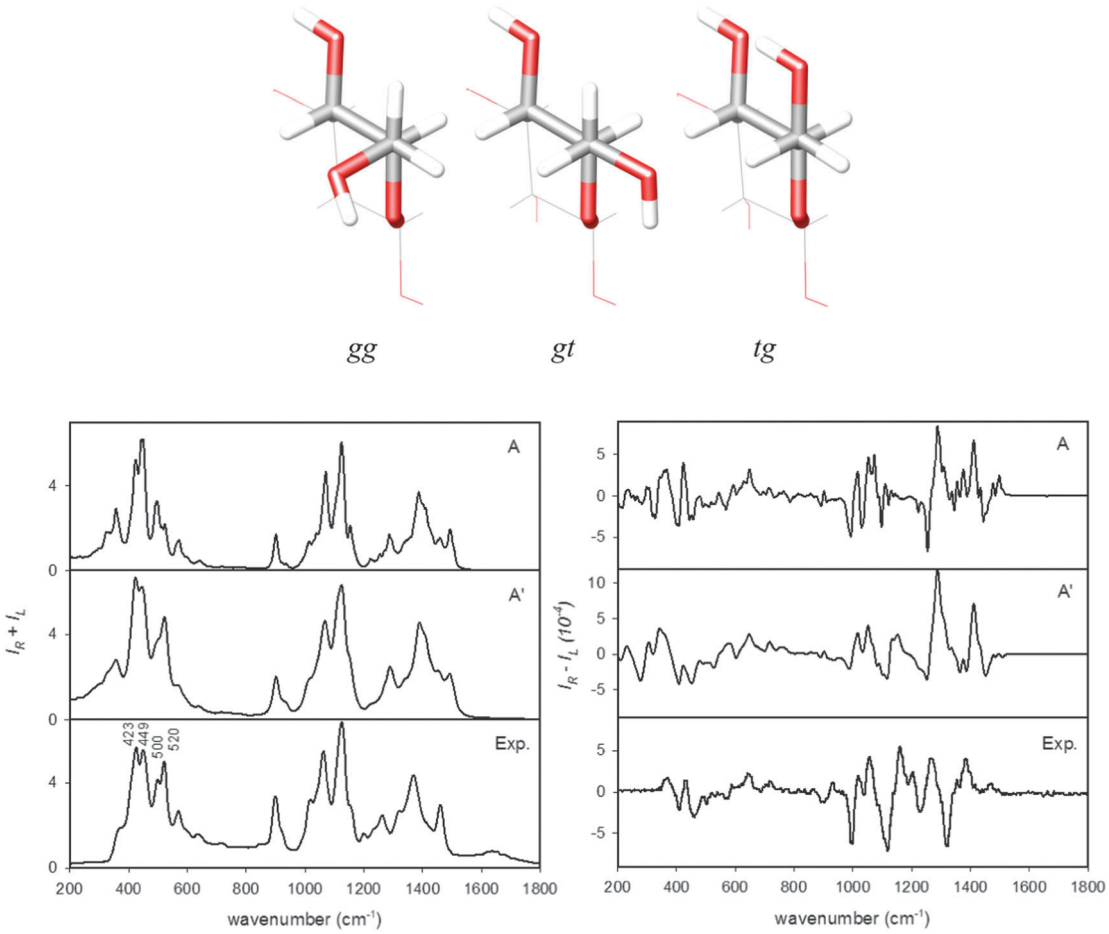


Fig. 4 Raman (left) and ROA (right) spectra of the β anomer of **DG** calculated by averaging using the equal weighting of 70 MD clusters without (A) and with (A′) the *gg*/*gt*/*tg* ratio adjusted according to NMR data (Table 2), and experiment from ref. 72.

Table 2 Experimental and calculated –CH_2OH rotamer populations (*gg*/*gt*/*tg*, in %)

	DFT[a]	MD-GLYCAM06	DFT (ref. 88 and 89)	Exp. (ref. 70 and 71)
DG	39/47/14	24/76/0	38/41/21	54/44/2
DM	41/24/35	26/74/0	59/3/38	52/43/5

[a] B3LYP/6-311++G**/CPCM Gibbs energy.

This is in agreement with the DFT results of Ionescu for penta-*O*-methyl-D-glucopyranose.[59] Similar energetic differences (18 and 20 kJ mol^{-1} for α-**DM** and β-**DM**, resp.) were observed for mannose.

Simulated spectra

Raman and ROA glucose and mannose spectra simulated with the DFT/PCM and cluster/MD models, without and with the NMR conformer correction, are plotted in Fig. 5 and 6. As expected, the Raman spectra (Fig. 5) are less influenced by the computational methodology, and all simulations reproduced the Raman main features observed in experiment. In comparison with the DFT/CPCM conformer averaging, the MD model better reproduces several positions and relative Raman band intensities within the entire spectral range. The anomeric and –CH_2OH rotational NMR correction has almost no effect on mannose Raman intensities, but causes minor changes in the glucose spectra, such as improvement of relative intensities of the 1062/1125 cm^{-1} bands, and the shape of the 520 cm^{-1} peak (numbers relate to experiment).

For the ROA spectra (Fig. 6) the DFT/CPCM model gives a rather poor agreement with the experiment, in particular in the low-frequency (<600 cm^{-1}) wavenumber region. This can be explained by direct participation of water molecules in the low-wavenumber vibrations,[30,41,82] which cannot be realistically simulated with the dielectric CPCM solvent model. At the same time, OH torsional modes (*cf.* Fig. S1, ESI†) contribute heavily here, conformation of which is very dependent on the interactions with the water environment as well. Within 800–1800 cm^{-1}, the simulated DFT/CPCM spectra follow the experiment more; however, although a band-to-band comparison is rarely possible. The NMR corrections qualitatively do not improve the overall quality of the simulation.

On the other hand, the MD calculations provide more realistic ROA patterns, almost to the same extent as for the Raman intensities. Also, the NMR correction appears more useful here than for the CPCM results. For example, the 344 cm^{-1} positive and 1112 cm^{-1} negative ROA **DG** bands are obtained with it, easily

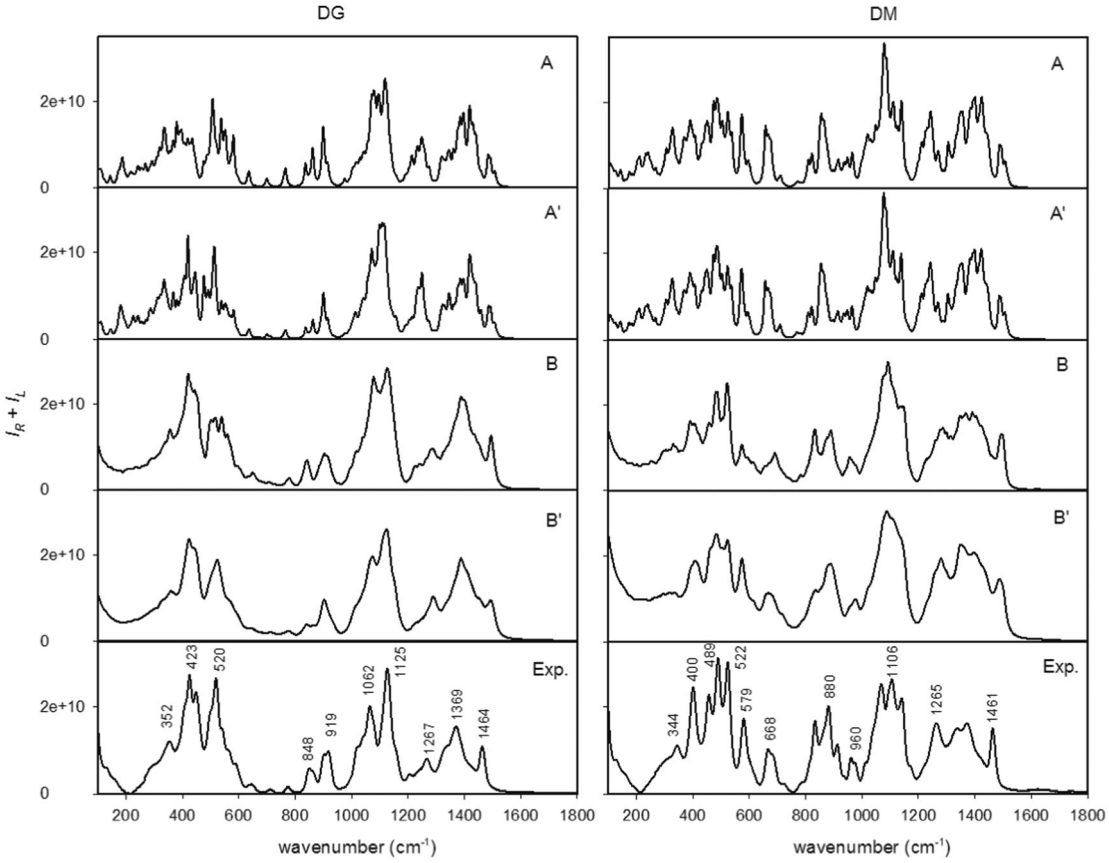


Fig. 5 Raman **DG** and **DM** spectra calculated by the plain (A) DFT/PCM and MD (B) computations, and with the α/β and *gg*/*gt* form populations adjusted according to the experiment[51,52] (A′ and B′), and the experiment.

assignable to their experimental counterparts. Similarly, for **DM**, the positive 1060/1090 cm^{-1} double band or the 1320 cm^{-1} negative band could not be obtained with the plain MD averaging. Experimentally as well as in theory, **DG** and **DM** ROA patterns are rather similar within 200–600 cm^{-1} (a broad "+–+" w-shape signal), whereas there are more characteristic features above 1000 cm^{-1}.

The comparison (Fig. 7) reveals that the involvement of water molecules explicitly in the Raman/ROA computation brings a minor effect only; the CPCM and QM/QM (*i.e.* MD approach in this work) spectral shape are quite similar, although minor spectral features, *e.g.* at around 1400 cm^{-1}, may be better reproduced by the more advanced QM/QM procedure. This is consistent with data obtained on similar systems previously.[65] However, the "explicit" water environment is necessary to properly weigh the sugar conformers; indeed, relying on the continuum solvent model only (*cf.* spectra DFT and DFT′, Fig. 6) provides much inferior results.

The QM/MM methodology for α-**DG** provided very good Raman shapes, better than QM/QM; however, the latter method led to more realistic ROA profile (especially in the region below 400 cm^{-1}, as well as at around 1400 cm^{-1}). This may be caused by a limited cluster size as discussed in the previous studies.[30,41,47] Indeed, the QM/MM methodology enables to work with much larger clusters than QM/QM, containing up to several hundreds of water molecules.[47] We can thus see that the combination of quantum and molecular dynamics provides many ways of spectral simulations, and their performance should be carefully tested, because of their semi-empirical character.

MD *versus* CPCM geometries

The differences in the description of the sugar structures obtained by the DFT/CPCM (with DFT-based conformer distribution) and MD modeling can be seen when histograms of individual torsion angles are compared, as exemplified for **DG** in Fig. 8 (*cf.* Fig. S3, ESI,† for **DM**). This reveals, for example, the fundamental difference in the dispersion of the torsion angles, much more limited for DFT/CPCM. This is explicable by the inadequate description of the hydrogen bonds by the CPCM model, not able to describe their local and partially covalent character.[83] In addition, MD provides conformers (angle values) that are not supported by DFT/CPCM within low relative energies. There are also minor differences in equilibrium torsion angles of individual conformers, usually not exceeding 10–20°.

Glucose–mannose mixtures

Analysis of mixtures of different sugars *via* Raman spectroscopy is a convenient analytical method that may be used, for example, in the food industry.[84–87] A decomposition of an experimental spectrum into theoretical spectra also provides

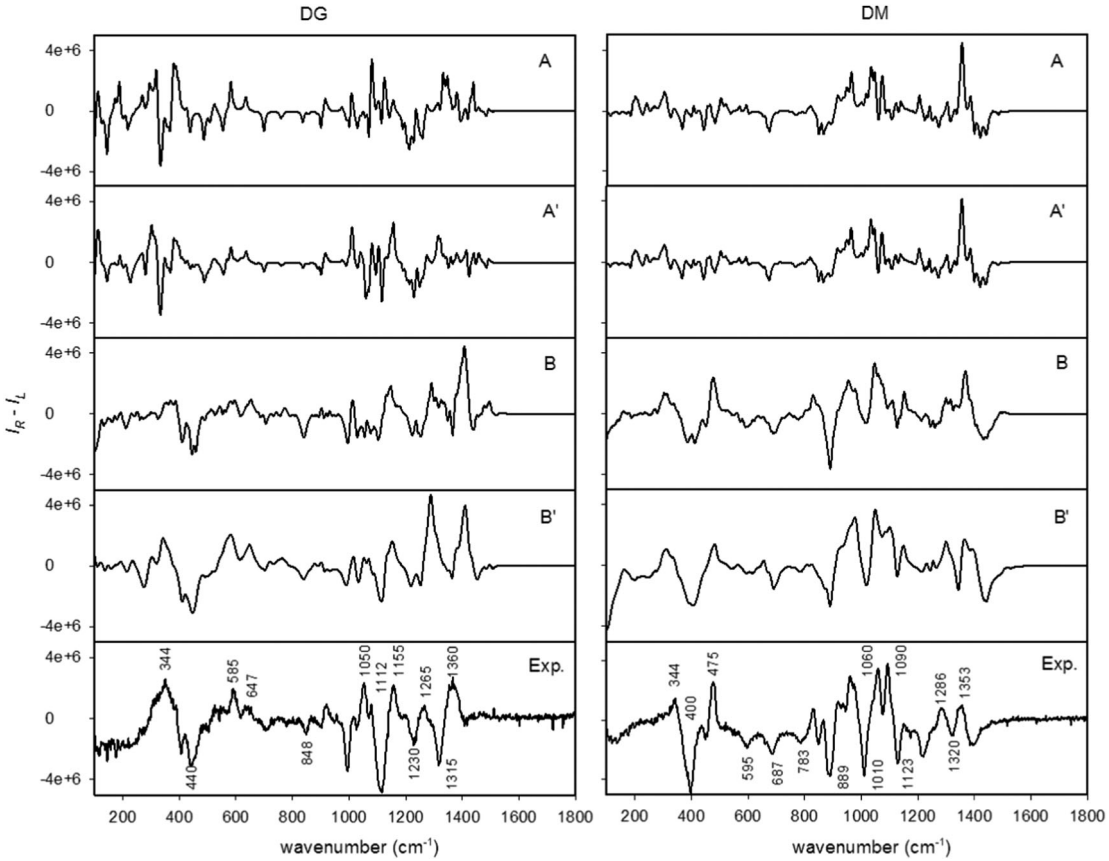


Fig. 6 ROA **DG** and **DM** spectra calculated by the plain (A) DFT/PCM and MD (B) computations, and with the α/β and *gg*/*gt* form populations adjusted according to the experiment[51,52] (A′ and B′), and the experiment.

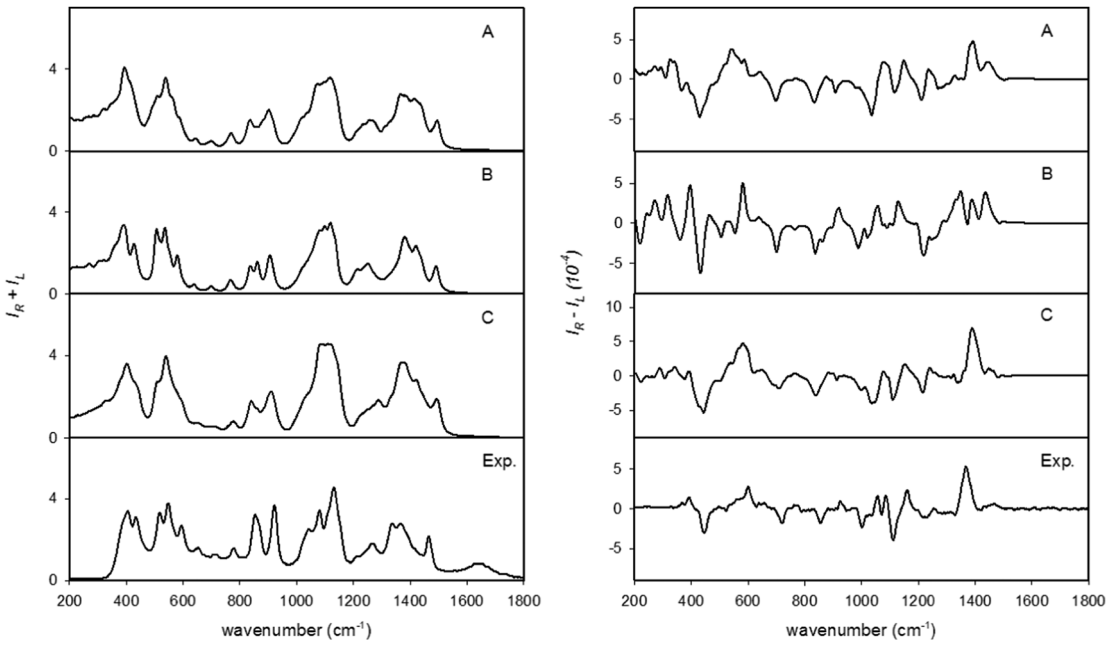


Fig. 7 Different solvent models: Raman and ROA spectra of the α-**DG** calculated by averaging of 70 MD cluster geometries using the equal weighting, with all water molecules deleted (A), using the Amber MM level for the waters (B), using the DFT level also for the waters (C), and the experiment. The *gg*/*gt*/*tg* ratio was adjusted according to NMR data (Table 2), experimental spectra of the α-anomer are redrawn from ref. 72.

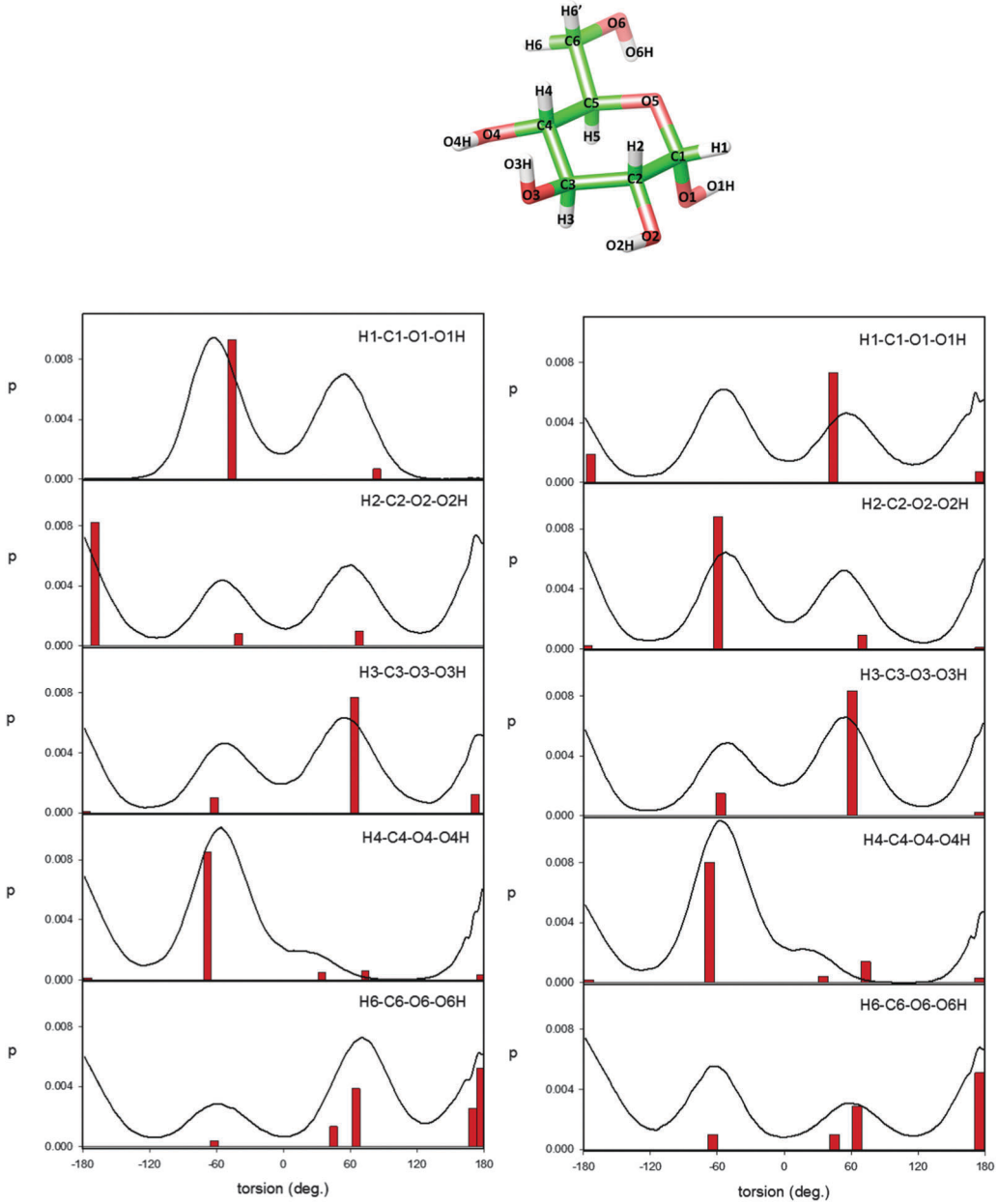


Fig. 8 Selected torsional angles as obtained by MD and DFT computations, for α-**DG** (left) and β-**DG** (right). MD probability histograms (black line) were obtained during 10 ns simulations with the GLYCAM06 force field. The DFT distribution bars (red, in arbitrary y-scale) were obtained from the B3LYP/6-311++G**/CPCM conformational scan and Boltzmann weighting.

an important feedback on the accuracy of the modeling, and widens the application range of ROA spectroscopy. The present accuracy can be estimated in Fig. 9, where the **DG** content in **DG**/**DM** mixtures as obtained from the Raman and ROA spectra is plotted against the exact ratios. Apparently, the decomposition provides the ratios with an error of 5–20%, similar when done with the experimental and calculated sub-spectra. This result suggests that the simulation has approached a reasonable precision, and it also justifies the use of the Raman and ROA spectroscopy for determination of different sugar forms. In particular, the error in calculated spectral intensities does not seem to have a dramatic effect on the resultant precision of the whole decomposition process; the precision appears limited more by the experimental error of the concentrations, Raman and ROA spectral noise, and baseline drift. At present, Raman spectra provide better precision of the decomposition (~5–10% error in conformer ratios) than ROA (~10–20%).

How the **DG**:**DM** ratio in a mixture is reflected by spectral changes in different wavenumber regions can be seen in Fig. S4 (ESI†) with experimental Raman and ROA spectra of 1:0, 3:1, 1:1, 2:3, 1:3 and 0:1 mixtures. For the 1:3 and 3:1 **DG**:**DM** mixtures the experimental and calculated Raman and ROA spectra are plotted in Fig. 10. Clearly, the spectroscopy is able to discriminate between various concentration ratios, and the

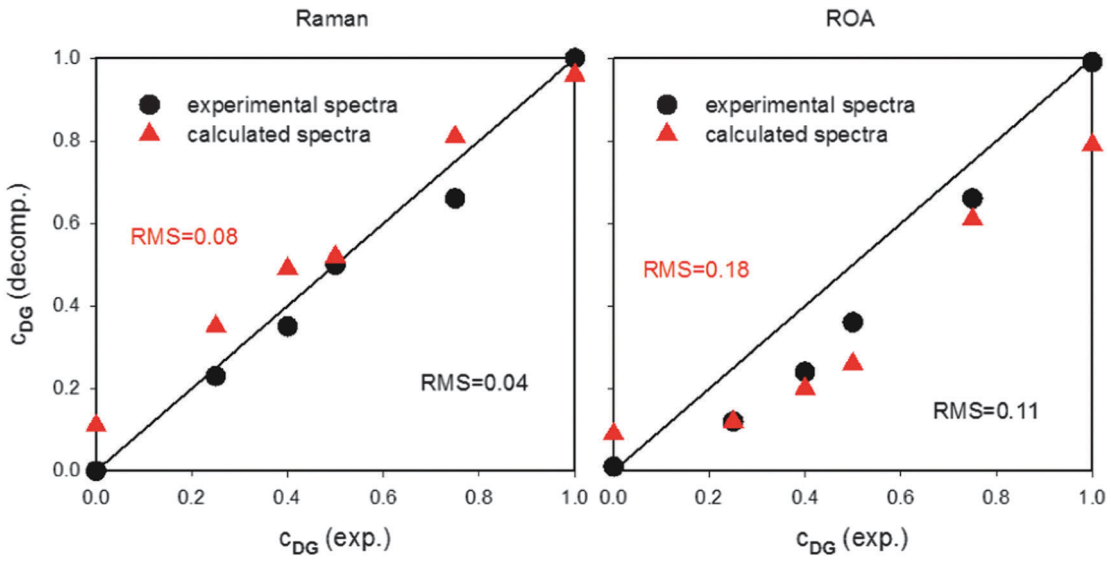


Fig. 9 **DG** content in **DG**/**DM** mixtures as obtained by decomposition of their Raman (left) and ROA (right) spectra into experimental (black circles) and calculated (red triangles) sub-spectra of individual components, see the Methods (eqn (2)) for the decomposition algorithm.

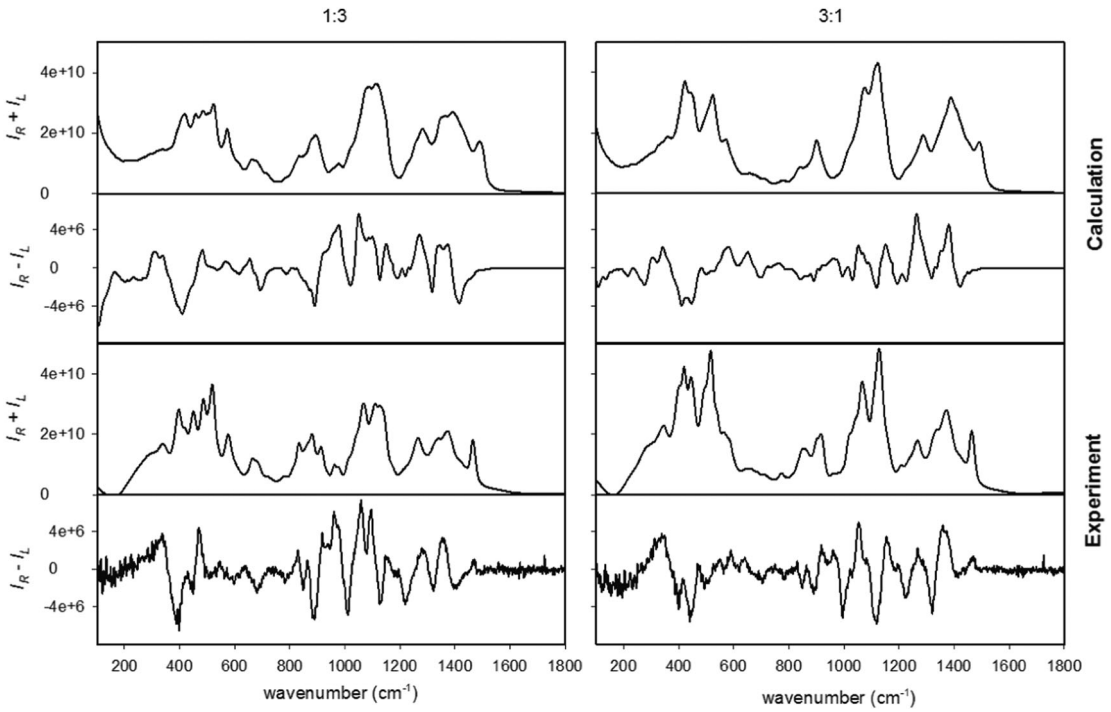


Fig. 10 Simulated and experimental spectra of 1 : 3 (left) and 3 : 1 (right) **DG**/**DM** mixtures. The adjusted MD spectra of **DG** and **DM** were used for the simulation.

theory is able to capture this. For example, the intensity of the 340–520 cm^{-1} **DG** Raman bands drops almost by 25% for higher **DM** content, for which a new band appears at 668 cm^{-1}. Also the Raman bands at 800–900 cm^{-1} and 1100–1200 cm^{-1} are very useful indicators, varying in intensity by 20–50%. As expected, the ROA spectra are sensitive even more, with many bands even changing sign according to actual sugar composition.

Conclusions

We explored several computational technologies with respect to their performance to simulate the Raman and ROA spectra of monosaccharides and their mixtures (α/β anomers, *gg*/*gt*/*tg* conformers and the glucose/mannose species). We found this important for better understanding of molecular behavior as well as for future applications of the spectroscopies.

Simulations based on the systematic conformational scan with the polarizable solvent model (DFT/CPCM) provided notably worse results than the combined MD/DFT cluster methodology, which better reflects the strong interaction between the sugars and water molecules. The MD/DFT approach was found to be limited by the inaccuracy of MD force field, which predicted the anomeric and $-CH_2OH$ form ratios associated with tiny energy differences with a significant error. The simulations then profited from experimental NMR data, which to some extent spoiled the "*a priori*" parameter-less approach, but enabled to better understand the link between the spectra and the structure. The results also indicate that Raman and ROA spectroscopy can be used as excellent tools for future improvements of saccharide force fields (although this goes beyond the scope of present work), providing feedback on tiny structural changes. The decomposition of glucose/mannose mixtures provided conformer ratios with an error of 5–20%, and the error was similar for experimental and calculated sub-spectra.

Acknowledgements

This work was supported by the Czech Science Foundation (16-00270S and 13-03978S). This work was supported by the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), funded by the European Regional Development Fund and the national budget of the Czech Republic *via* the Research and Development for Innovations Operational Program, as well as Czech Ministry of Education, Youth and Sports *via* the project Large Research, Development and Innovations Infrastructures (LM2011033). Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure Meta Centrum, provided under the program "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005), is greatly appreciated, as well as access to the CERIT-SC computing and storage facilities provided under the program Center CERIT Scientific Cloud, part of the Operational Program Research and Development for Innovations, reg. no. (CZ. 1.05/3.2.00/08.0144). M. J. Kaminský was also supported by the grants (Czech Science Foundation; 14-03564S, 15-09072S).

Notes and references

1 L. D. Barron, M. P. Bogaard and A. D. Buckingham, *J. Am. Chem. Soc.*, 1973, **95**, 603–605.
2 L. D. Barron, L. Hecht, I. H. McColl and E. W. Blanch, *Mol. Phys.*, 2004, **102**, 731–744.
3 L. D. Barron, E. W. Blanch, I. H. McColl, C. D. Syme, L. Hecht and K. Nielsen, *Spectrosc. Int. J.*, 2003, **17**, 101–126.
4 E. W. Blanch, L. Hecht and L. D. Barron, *Methods*, 2003, **29**, 196–209.
5 R. D. Amos, *Chem. Phys. Lett.*, 1982, **87**, 23.
6 K. Ruud, T. Helgaker and P. Bouř, *J. Phys. Chem. A*, 2002, **106**, 7448–7455.
7 J. Šebestík and P. Bouř, *J. Phys. Chem. Lett.*, 2011, **2**, 498–502.
8 J. Šebestík and P. Bouř, *Angew. Chem., Int. Ed.*, 2014, **53**, 9236–9239.
9 S. Yamamoto and P. Bouř, *Angew. Chem., Int. Ed.*, 2013, **51**, 11058–11061.
10 S. Yamamoto, H. Watarai and P. Bouř, *ChemPhysChem*, 2011, **12**, 1509–1518.
11 P. Bouř, *J. Comput. Chem.*, 2001, **22**, 426–435.
12 S. Yamamoto and P. Bouř, *J. Comput. Chem.*, 2013, **34**, 2152–2158.
13 S. Yamamoto, J. Kaminský and P. Bouř, *Anal. Chem.*, 2012, **84**, 2440–2451.
14 V. Parchaňský, J. Kapitán and P. Bouř, *RSC Adv.*, 2014, **4**, 57125–57136.
15 V. Parchaňský, J. Kapitán, J. Kaminský, J. Šebestík and P. Bouř, *J. Phys. Chem. Lett.*, 2013, **4**, 2763–2768.
16 L. Benda, P. Štěpánek, J. Kaminský and P. Bouř, in *Compr. Chirality*, ed. E. M. Carreira and H. Yamamoto, Elsevier, Spectroscopic Analysis: Ab initio Calculation of Chiroptical Spectra, 2012, vol. 8, pp. 520–544.
17 R. Stern, A. A. Asari and K. N. Sugahara, *Eur. J. Cell Biol.*, 2006, **85**, 699–715.
18 T. Feizi and W. Chai, *Nat. Rev. Mol. Cell Biol.*, 2004, **5**, 582–588.
19 A. Varki, and J. B. Lowe, in *Essentials of Glycobiology*, ed. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R Bertozzi, G. W. Hart and M. E. Etzler, Biological Roles of Glycans, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2nd edn, 2009.
20 N. R. Yaffe, A. Almond and E. W. Blanch, *J. Am. Chem. Soc.*, 2010, **132**, 10654–10655.
21 L. D. Barron, A. R. Gargaro and Z. Q. Wen, *Carbohydr. Res.*, 1991, **210**, 39–49.
22 A. F. Bell, L. Hecht and L. D. Barron, *J. Raman Spectrosc.*, 1993, **24**, 633–635.
23 A. F. Bell, L. Hecht and L. D. Barron, *Spectrochim. Acta, Part A*, 1995, **51**, 1367–1378.
24 Z. Q. Wen, L. D. Barron and L. Hecht, *J. Am. Chem. Soc.*, 1993, **115**, 285–292.
25 A. F. Bell, L. D. Barron and L. Hecht, *Carbohydr. Res.*, 1994, **257**, 11–24.
26 A. F. Bell, L. Hecht and L. D. Barron, *Chem. – Eur. J.*, 1997, **3**, 1292–1298.
27 T. R. Rudd, R. Hussain, G. Siligardi and E. A. Yates, *Chem. Commun.*, 2010, **46**, 4124–4126.
28 F. Zhu, N. W. Isaacs, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 2005, **127**, 6142–6143.
29 J. Kaminský, J. Kapitán, V. Baumruk, L. Bednárová and P. Bouř, *J. Phys. Chem. A*, 2009, **113**, 3594–3601.
30 J. R. Cheeseman, M. S. Shaik, P. L. A. Popelier and E. W. Blanch, *J. Am. Chem. Soc.*, 2011, **133**, 4991–4997.
31 S. Luber and M. Reiher, *J. Phys. Chem. A*, 2009, **113**, 8268–8277.
32 C. Johannessen, R. Pendrill, G. Widmalm, L. Hecht and L. D. Barron, *Angew. Chem., Int. Ed.*, 2011, **50**, 5349–5351.
33 L. Ashton, P. D. Pudney, E. W. Blanch and G. E. Yakubov, *Adv. Colloid Interface Sci.*, 2013, **66**, 166–200.

34 A. F. Bell, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 1994, **116**, 5155–5161.
35 C. Mensch, R. Pendrill, G. Widmalm and C. Johannessen, *ChemPhysChem*, 2014, **15**, 2252–2254.
36 L. D. Barron, A. R. Gargaro, Z. Q. Wen, D. D. MacNicol and C. Butters, *Tetrahedron: Asymmetry*, 1990, **1**, 513–516.
37 A. F. Bell, L. Hecht and L. D. Barron, *Chem. – Eur. J.*, 1997, **3**, 1292–1298.
38 A. F. Bell, L. Hecht and L. D. Barron, *J. Raman Spectrosc.*, 1995, **26**, 1071–1074.
39 N. A. Macleod, C. Johannessen, L. Hecht, L. D. Barron and J. P. Simons, *Int. J. Mass Spectrom.*, 2006, **253**, 193–200.
40 M. Pecul, A. Rizzo and J. Leszczynski, *J. Phys. Chem. A*, 2002, **106**, 11008–11016.
41 S. T. Mutter, F. Zelinski, J. R. Cheeseman, C. Johanessen, P. L. A. Popelier and E. W. Blanch, *Phys. Chem. Chem. Phys.*, 2015, **17**, 6016–6027.
42 J. Hilario, J. Kubelka and T. A. Keiderling, *J. Am. Chem. Soc.*, 2003, **125**, 7562–7574.
43 V. W. Jürgensen and K. Jalkanen, *Phys. Biol.*, 2006, **3**, S63–S79.
44 M. Losada, H. Tran and Y. Xu, *J. Chem. Phys.*, 2008, **128**, 014508.
45 M. Losada, P. Nguyen and Y. Xu, *J. Phys. Chem. A*, 2008, **112**, 5621–5627.
46 M. R. Poopari, Z. Dezhahang, G. Yang and Y. Xu, *ChemPhysChem*, 2012, **13**, 2310–2321.
47 F. Zielinski, S. T. Mutter, C. Johanessen, E. W. Blanch and P. L. A. Popelier, *Phys. Chem. Chem. Phys.*, 2015, **17**, 21799–21809.
48 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 Jr., 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, Revision D.01*, Gaussian, Inc., Wallingford CT, 2009.
49 (*a*) F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson and W. C. Still, *J. Comput. Chem.*, 1990, **11**, 440; (*b*) *MacroModel, version 9.1*, Schrödinger, LLC, New York, NY, 2005.
50 D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossvai, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, *AMBER 12*, University of California, San Francisco, 2012.
51 S. J. Angyal, *Angew. Chem., Int. Ed.*, 1969, **8**, 157–166.
52 S. J. Angyal, *Carbohydr. Res.*, 1994, **263**, 1–11.
53 A. Klamt, in *The Encyclopedia of Computational Chemistry*, ed. P. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III and P. R. Schreiner, John Wiley & Sons, Chichester, 1998, vol. 1, pp. 604–615.
54 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799.
55 G. I. Csonka and J. Kaminský, *J. Chem. Theory Comput.*, 2011, **7**, 988–997.
56 W. M. C. Sameera and D. A. Pantazis, *J. Chem. Theory Comput.*, 2012, **8**, 2630–2645.
57 B. J. Smith, *J. Am. Chem. Soc.*, 1997, **119**, 2699–2706.
58 X. Biarnés, A. Ardevol, A. Planas, C. Rovira, A. Laio and M. Parrinello, *J. Am. Chem. Soc.*, 2007, **129**, 10686–10693.
59 A. R. Ionescu, A. Berces, M. Z. Zgierski, D. M. Whitfield and T. Nukada, *J. Phys. Chem. A*, 2005, **109**, 8096–8105.
60 K. N. Kirschner, A. B. Yongye, S. M. Tschampel, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem.*, 2008, **29**, 622–655.
61 W. L. Jorgensen, J. Chandrasekhar and J. D. Madura, *J. Chem. Phys.*, 1983, **79**, 926–935.
62 J. Kessler, M. Dračínský and P. Bouř, *J. Comput. Chem.*, 2012, **34**, 366–371.
63 XSHELL, P. Bouř, http://hanicka.uochb.cas.cz/~bour/programs/F/xshell.f, Praha, 2015.
64 J. Hudecová, K. Hopmann and P. Bouř, *J. Phys. Chem. B*, 2012, **116**, 336–342.
65 K. Hopmann, K. Ruud, M. Pecul, A. Kudelski, M. Dračínský and P. Bouř, *J. Phys. Chem. B*, 2011, **115**, 4128–4137.
66 J. Kapitán, M. Dračínský, J. Kaminský, L. Benda and P. Bouř, *J. Phys. Chem. B*, 2010, **114**, 3574–3582.
67 M. Dračínský and P. Bouř, *J. Chem. Theory Comput.*, 2010, **6**, 288–299.
68 P. Bouř and T. A. Keiderling, *J. Chem. Phys.*, 2002, **117**, 4126–4132.
69 P. Bouř, *Collect. Czech. Chem. Commun.*, 2005, **70**, 1315–1340.
70 Y. Nishida, H. Ohrui and H. Meguro, *Tetrahedron Lett.*, 1984, **25**, 1575–1578.
71 K. Bock and J. O. Duus, *J. Carbohydr. Chem.*, 1994, **13**, 513–543.
72 W. Hug, in *Encyclopedia of Spectroscopy and Spectrometry*, ed. J. Linfon, Academic Press, 2nd edn, 2010, pp. 2387–2405.
73 S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma and M. J. Frisch, *J. Mol. Struct.*, 1999, **462**, 1–21.
74 T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery Jr., K. Morokuma and M. J. Frisch, *J. Chem. Theory Comput.*, 2006, **2**, 815–826.
75 F. Clemente, T. Vreven and M. J. Frisch, *Quantum Biochemistry*, Wiley VCH, Weinheim, 2010.

76 T. Vreven and K. Morokuma, *Continuum Solvation Models in Chemical Physics: From Theory to Applications*, Wiley, 2008.
77 J. T. Edward, *Chem. Ind.*, 1955, 1102–1104.
78 R. U. Lemieux and G. Huber, *J. Am. Chem. Soc.*, 1956, **78**, 4117–4119.
79 R. U. Lemieux and P. Chu, presented in part at the Abstracts of Papers; 133rd National Meeting of the American Chemical Society, Washington D.C., 1958.
80 C. O. da Silva, B. Mennucci and T. Vreven, *J. Org. Chem.*, 2004, **69**, 8161–8164.
81 S. E. Barrows, J. W. Storer, C. J. Cramer, A. D. French and D. G. Truhlar, *J. Comput. Chem.*, 1998, **19**, 1111–1204.
82 J. Kapitán, V. Baumruk, V. Kopecký Jr., R. Pohl and P. Bouř, *J. Am. Chem. Soc.*, 2006, **128**, 13451–13462.
83 P. Bouř, D. Michalík and J. Kapitán, *J. Chem. Phys.*, 2005, **122**, 144501.
84 B. Özbalci, I. H. Boyaci, A. Topcu, C. Kadılar and U. Tamer, *Food Chem.*, 2013, **136**, 1444–1452.
85 M. M. Paradkar and J. Irudayaraj, *Food Chem.*, 2001, **76**, 231–239.
86 M. M. Paradkar, S. Sakhamuri and J. Irudayaraj, *J. Food Sci.*, 2002, **67**, 2009–2015.
87 K. Ilaslan, I. H. Boyaci and A. Topcu, *Food Control*, 2014, **48**, 56–61.
88 M. Appell, G. Strati, J. L. Willet and F. A. Momany, *Carbohydr. Res.*, 2004, **339**, 537–551.
89 M. Appell, J. L. Willet and F. A. Momany, *Carbohydr. Res.*, 2005, **340**, 459–468.

10.2 Appendix I – *Anal. Chem.* 2016, 88, 8878.

analytical chemistry

Article

pubs.acs.org/ac

Detection of Sugars via Chirality Induced in Europium(III) Compounds

Tao Wu,[†] Jiří Průša,[§,†] Jiří Kessler,[⊥,†] Martin Dračínský,[†] Jan Valenta,[*,‡] and Petr Bouř[*,†]

[†]Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Flemingovo náměstí 2, Prague, 16610, Czech Republic

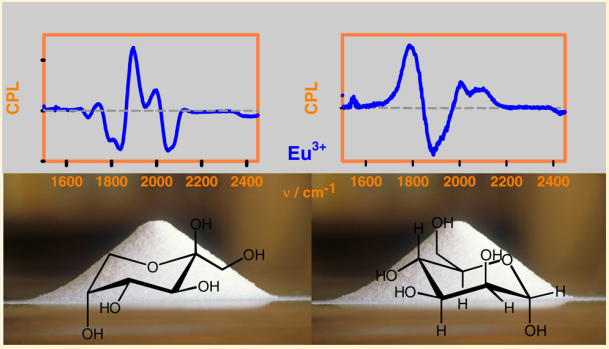
[§]Department of Analytical Chemistry, University of Chemistry and Technology, Technická 5, Prague, 16628, Czech Republic

[⊥]Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University, Hlavova 8, Prague, 12840, Czech Republic

[‡]Faculty of Mathematics and Physics, Charles University, Ke Karlovu 3, Prague, 12116, Czech Republic

Supporting Information

ABSTRACT: Detection and resolution of simple monosaccharides are difficult tasks because their structure is quite similar. The present study shows that circularly polarized luminescence (CPL) induced in europium complexes provides very specific spectral patterns for fructose, mannose, glucose, and galactose. Differences were also observed between bare Eu^{3+} ion and its complexes, when interacting with these sugars. The CPL spectra were measured on a Raman optical activity (ROA) spectrometer, which ensured high fluorescence intensity owing to the strong 532 nm laser excitation. The induced fluorescence was recorded in the same spectrum as the vibrational Raman bands. On the basis of the ligand field theory, most fluorescence spectral peaks could be assigned to *f*-shell europium transitions. Additional information on the interaction of the lanthanide with the sugar component was provided by measurement of time-dependent fluorescence, as formation of different complexes led to variations in fluorescence decay times. In nuclear magnetic resonance (NMR), the paramagnetic metal ion interacting with the sugars caused specific changes in ^{13}C chemical shifts. The spectroscopic data and molecular dynamics modeling showed that the interaction between the monosaccharides and Eu ion is rather weak due to the competition of the OH sugar groups with water molecules. However, multiple binding modes are possible, which contributes to the complexity and specificity of the spectra. The induced chirality and fluorescence spectra thus appear to be convenient means for monosaccharide detection and identification.



The prospect of detecting saccharides in physiologically relevant conditions attracts attention because it opens the way to monitoring and control of a wide range of biological processes including cellular recognition, immune response, and regulation of enzymatic activity. Metal coordination of natural carbohydrates in aqueous solution is particularly suitable for this purpose as it often provides remarkable selectivity and stability.[1−3] However, the stability of the complexes strongly depends on experimental conditions. A typical problem in detecting carbohydrates in an aqueous environment is a competition of the sugar receptor with the hydroxyl groups of water. Complexation of sugars with lanthanides can be followed by infrared or microwave spectroscopy where, however, rather unspecific changes were often observed.[3−5]

Fluorescence of the lanthanide complexes is much more sensitive even to weak interaction with saccharides and has been suggested to detect and identify neutral sugars including cancer biomarkers.[6] The interaction and specificity can be conveniently tuned by varying the metal and/or ligands.[3,7,8] For many metals and their complexes, the affinity to sugars and consequent stability constants are comparable to the older and still more common carbohydrate sensors based on the boronic acid.[9]

Circularly polarized luminescence (CPL), differential emission of the left- and right-circularly polarized light, is potentially even more attractive than the total luminescence, as CPL bands can be either positive or negative. CPL spectra are thus more specific, making it possible to distinguish more electronic transitions, and the information is easier to read. Lanthanide ions or their complexes are usually not chiral, but the chirality can be induced by the sugar component.[10,11] At the same time, the specific electronic structure of lanthanides[12] allows for a very high dissymmetry factor ($g = 2(I_R - I_L)/I$, i.e., twice the ratio of CPL to total luminescence, where I_R and I_L are the

Received: June 30, 2016
Accepted: August 15, 2016
Published: August 15, 2016

DOI: 10.1021/acs.analchem.6b02505
Anal. Chem. 2016, 88, 8878−8885

intensity of the right- and left-circularly polarized light, respectively).[13,14]

CPL measurements, however, are often difficult to do as sensitivity of CPL spectrometers is limited. In the present study, we use the Raman optical activity (ROA) spectrometer fitted with a strong 532 nm laser excitation source and a sensitive setup for detecting the difference in circular polarization.[15] This enables measurements of tiny CPL signals undetectable by other means.

Traditionally, the vibrational ROA spectroscopy detects a small difference in Raman scattering intensities of the right- and left-circularly polarized light.[16] It is sensitive to fine structural variations in chiral molecules and has been applied to a wide range of molecules including proteins, nucleic acids, and mono- and polysaccharides.[17−21] The ROA spectrum itself can thus be used as an extremely useful characteristic of the sugar. However, the ROA signal is often difficult to measure as well, because a typical circular intensity difference (CID, ratio of the ROA and Raman signal, i.e., the ROA analogy of g)[22] is very small, typically around 10^{-4}, and the Raman scattering itself is rather weak. High sugar concentrations are needed for a meaningful analysis.[23]

The CPL component of the ROA spectrum of europium−sugar conjugates measured together with the "true" vibrational ROA signal of pure sugars thus provides a welcome sensitivity enhancement of the spectroscopic detection. The interaction between the lanthanide or its complex and the monosaccharide provides an additional specificity about the sugar skeleton. The fluorescence bands are usually easily recognizable among the vibrational Raman and ROA bands in the spectrum, because of their higher intensity and stability of the lanthanide transition energies, only weakly dependent on the environment.[12] As previously discussed, the physical origin (fluorescence or Raman scattering) of the observed bands can also be unambiguously determined using multiple laser excitation wavelengths or by measuring the degree of circularity.[14]

In the present study, europium(III) in the form of chloride and two complexes stable in an aqueous environment are used to investigate the interaction with four common monosaccharides. The Eu^{3+} ion in particular provides a rich fluorescence spectrum within the wavelength range of the ROA spectrometer (about 532−610 nm). For mannose and fructose providing the strongest spectral responses, we correlate the CPL data to fluorescence decay times and paramagnetic nuclear magnetic resonance (NMR) shifts caused by the binding. A custom-made setup is used to record the fluorescence kinetics, as it is too slow (in the microsecond range) to be measurable on standard fluorescence spectrometers. The Eu−sugar interactions clearly bring about kinetics changes that are unreported so far to the best of our knowledge.

CPL induced in europium and other lanthanide complexes has been previously observed as a result of interaction with amino acids.[24−26] With the sugars, however, the interaction is much more specific. Early CPL studies were hampered by the limited sensitivity of available spectrometers[11] which restricted the number of systems that could be studied. For the amino acids, the induction of chirality was explained by a perturbed equilibrium of two enantiometric forms of the lanthanide (III) complexes. The ROA/CPL technique applied for sugars reveals greater variability and complexity of induced CPL spectra. At least to some extent, this could be explained by the multivalence modes possible for various sugar forms and rationalized by computational models involving density functional theory (DFT), molecular dynamics (MD), and the crystal field theory.

METHODS

The NaEuEDTA and Na_2EuDEPA complexes (Figure 1) were obtained by a reaction of europium oxide with 2.05 equiv of

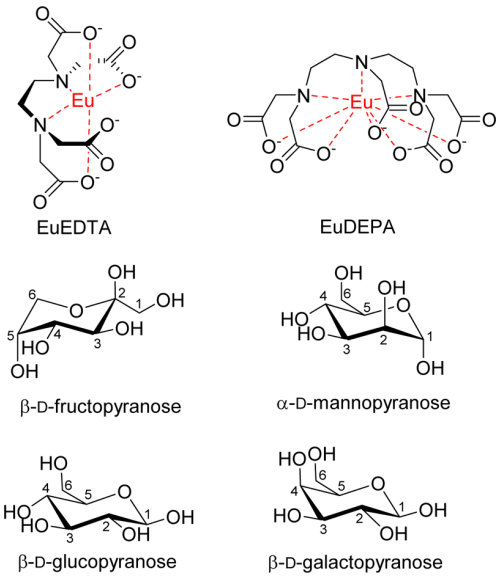


Figure 1. Structures of the two europium complexes and investigated monosaccharides. Monosaccharide forms most abundant in water solutions are depicted.

ethylenediaminetetraacetic (EDTA) and diethylenetriaminepentaacetic (DEPA) acid, respectively, kept in water at 70 °C for 4 h. The solution was then cooled down to room temperature, and the pH was adjusted to 7.0 by 1 M sodium carbonate solution. Water solutions of $EuCl_3$, NaEuEDTA, and Na_2EuDEPA complexes in 4 mM concentrations, and sugars (concentrations of 400−800 mM) were prepared, and their Raman and (back-)scattered circular polarized (SCP) ROA spectra were acquired on a BioTools spectrometer using 532 nm laser excitation, resolution of 7 cm^{-1}, laser power at the sample of 120−900 mW, and acquisition times of 1−16 h. Water background was subtracted from the Raman spectra; the water 1650 cm^{-1} band was also used to normalize the Raman intensities.

Fluorescence lifetimes were measured by a custom-build spectroscope (Charles University)[27] using epifluorescence illumination and collection of signal with an objective lens 4×/0.13 with a working distance of 17 mm. The continuous diode laser at 405 nm was modulated by a quartz acousto-optic modulator to provide square pulses with a repetition rate of 800 Hz and a 25% duty cycle. Typically, the pulse duration was 312.5 μs and the edge smearing was about 0.1 μs. Excitation power density in a solution measured inside a cuvette was about 0.8 W/cm^2. The fluorescence signal was verified to be a linear function of the excitation power. The signal was focused by a tube lens with focal length of 18 cm on an entrance slit of a grating spectrometer and detected by the Hamamatsu H11526-20-NF photomultiplier in the photon counting mode. Counts were treated with a multichannel scaler card MSA-300 (Becker & Hickl) set to 1200 points with 1 μs step. The signal was acquired during 1×10^5, 2×10^5, or 3×10^5 cycles. A second

DOI: 10.1021/acs.analchem.6b02505
Anal. Chem. 2016, 88, 8878−8885

output port of the spectrometer was equipped with a LN-cooled CCD camera which detected the fluorescence spectra.

NMR spectra were recorded at room temperature on a Bruker AVANCE III spectrometer operating at 500.0 MHz (^{1}H) and 125.7 MHz (^{13}C). About 16 mg of the monosaccharide was dissolved in 0.4 mL of D_2O and titrated by adding five times of 100 μL of $EuCl_3 \cdot 6H_2O$ (30 mg/0.6 mL) or EuEDTA (30 mg/0.6 mL) D_2O solutions. The spectra were referenced to C_6D_6, which was kept in a sealed capillary placed in the sample cuvette. 1D and 2D correlation NMR experiments (COSY, HSQC, HMBC) were combined to assign the signals.

The Gaussian[28] program was used to provide model geometries. X-ray geometries[29,30] of the NaEuEDTA and Na_2EuDEPA complexes were used as starting geometries and optimized by energy minimization using the B3LYP[31] functional, 6-311++G** basis set (the MWB28[32] pseudopotential and basis set for Eu), and polarizable continuum model (PCM)[33] for the water environment.

The semiempirical crystal field theory[12,34] was used to assign europium bands that are due to *f*-shell transitions and to approximately simulate the effect of the ligands. An adapted version of the Lanthanide[35] program was used in the calculation. As usual, the ligands were approximated by charge density computed on a grid using the Gaussian program, and the resultant electrostatic potential was used to perturb free ion energies and wave functions.[36−38]

For fructose and mannose, possible geometries and association energies of their complexes with the Eu^{3+} ion were also estimated using molecular dynamics simulations within the Amber program package.[39] In vacuum, a systematic search for the best binding sites was performed by minimizing the energy of a complex with the Eu^{3+} ion and the sugar, separately for α and β-anomers, and the furanose and pyranose fructose forms. About 200 positions of europium around the sugar were tested as the initial geometries; the minimization was performed with the GLYCAM06 force field[40] for the sugars; Eu^{3+} force field parameters were taken from ref 41. For all minima, the complexes were put into a cubic box ($(20\ Å)^3$) filled with 255 water molecules and molecular dynamics was run. For an equilibration phase (500 ps), the complex atoms were fixed and only water was allowed to relax, using the *nVT* ensemble, temperature of 300 K, and 1 fs integration time; then, the geometry was minimized again without any constraints. Free energies of the europium−sugar complex formation were estimated using the weighted histogram analysis method (WHAM).[42] Three characteristic Eu···O distances of the minimized structures were incremented by 0.25 Å, and the histograms were collected at 20 points, each of them containing 1 000 000 MD steps; the free energy profiles were obtained using the "Wham" script.[43]

RESULTS AND DISCUSSION

ROA and CPL Spectra. The spectra of the $EuCl_3$ and NaEuEDTA and Na_2EuDEPA aqueous solutions mixed with fructose, mannose, glucose, and galactose are plotted in Figure 2. Within the 200−1500 cm^{-1} interval, "ordinary" monosaccharide vibrational ROA spectra are apparent, as analyzed in other studies.[20,44−47] The ratio of the ROA and Raman signals (CID, circular intensity difference) is rather weak, and a relatively high noise level is present. Corresponding Raman spectra and ROA peak positions are plotted in Figure S1.

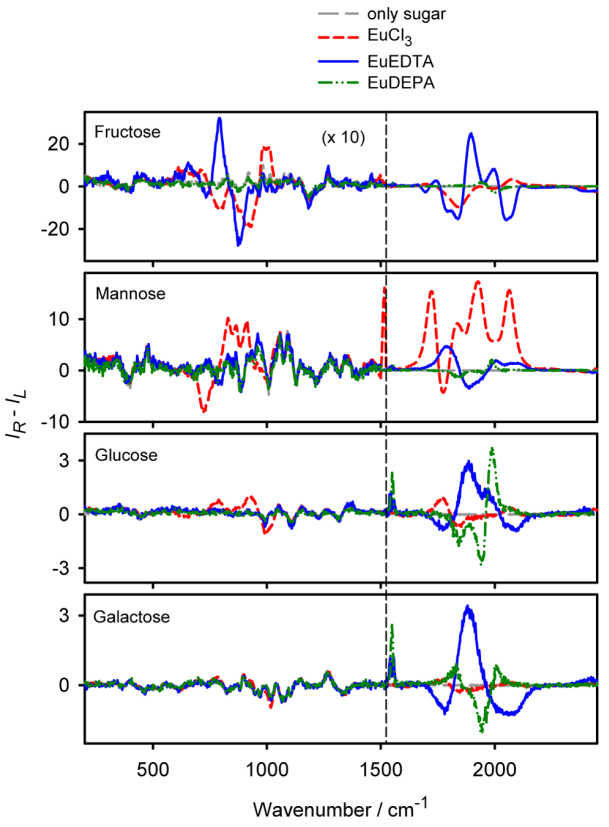


Figure 2. ROA spectra of $EuCl_3$, NaEuEDTA, and Na_2EuDEPA solutions in the presence of four monosaccharides exhibit a strong circularly polarized fluorescence component (right-hand side), very specific for each of the studied sugars.

An addition of the europium compounds is occasionally accompanied by the appearance of new bands within the 700−1000 cm^{-1} interval, intensity and CID of which is comparable with that of pure sugars. Much stronger bands appear within 1500−2450 cm^{-1}. These can be assigned to europium CPL and are more than ten times stronger than the vibrational ROA signal; it is thus much easier to measure them, and the signal-to-noise ratio is higher. Note that the noise level of both the Raman and ROA signal is proportional to the square root of the Raman counts on the detector.[16] Thus, the stronger signal of Raman/luminescence scattering also improves the accuracy of the ROA/CPL component. This is critical, for example, in a quantitative analysis for sugar mixtures.[20,23]

A closer look reveals remarkable specificity and significant differences among both the europium compounds and the sugars. For example, fructose induces a strong multiband pattern in the EuEDTA ion with peaks at 1696(−), 1746 (+), 1780 (−), 1839(−), 1897 (+), 1999 (+), 2052 (−), 2133 (+), and 2416 (−) cm^{-1}. The last band (2416 cm^{-1}) is close to the operational limit of the spectrometer, and its intensity might be attenuated by the limited sensitivity of the CCD detector.[48] For $EuCl_3$, the fructose ROA/CPL spectral pattern is much simpler and the signal is weaker (dominated by 1823 (−) and 2055 (+) cm^{-1} bands) than for EuEDTA. For mannose, the situation is rather opposite; i.e., there is a strong, about six-band signal with $EuCl_3$ and a weaker 1808(+)/1888(−) cm^{-1} "couplet" (two strong close bands of similar intensities but of opposite signs)

DOI: 10.1021/acs.analchem.6b02505
Anal. Chem. 2016, 88, 8878−8885

dominating the EuEDTA spectrum. The third EuDEPA complex gives much weaker CPL for both sugars.

Glucose and galactose are rather similar in that their CPL is about 10× weaker than for fructose and mannose, although the luminescence is still significantly stronger than the vibrational ROA. Also, in terms of CID, the highest values are provided by mannose/$EuCl_3$ ($\sim 3 \times 10^{-3}$) and fructose/EuEDTA or $EuCl_3$ complexes ($\sim 1 \times 10^{-3}$), while CID values for glucose and galactose are smaller than 2×10^{-4}. However, the "performance" of the EuDEPA complex for glucose and galactose is better than for fructose and mannose, as it provides CPL intensities comparable to EuEDTA.

Typical Raman spectra for fructose are plotted in Figure 3; spectra of the other sugars are quite similar and can be found in

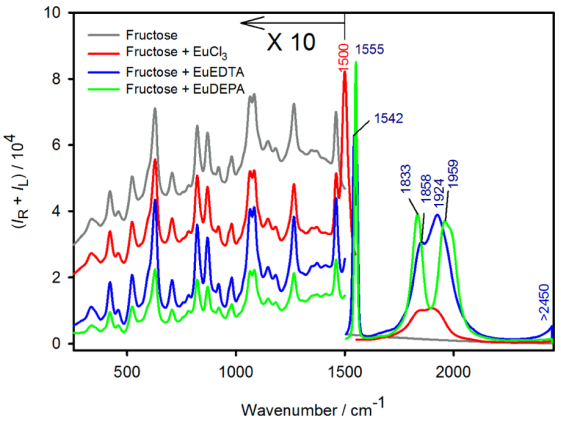


Figure 3. Raman spectra of $EuCl_3$, EuEDTA, and EuDEPA solutions in the presence of fructose. The Raman/luminescence spectra are not as sensitive to the sugar type as CPL.

the Supporting Information. As for ROA, the 200−1500 cm^{-1} region is dominated by a relatively weak vibrational Raman scattering of the sugar, whereas above 1500 cm^{-1}, the spectrum mostly comprises the total luminescence of europium. The Raman spectrum obviously also comprises the luminescence, and intensities of both are not much affected by the presence of monosaccharides. This reflects the generally lower sensitivity of unpolarized spectra to structural changes[16] and weak binding interactions between the europium ions/complexes and the sugars. On the other hand, $EuCl_3$, EuEDTA, and EuDEPA do exhibit specific luminescence. For example, $EuCl_3$ provides the weakest signal around 1900 and 1500 cm^{-1}; the latter transition is also shifted to higher wavenumbers for the other two complexes. EuDEPA gives the most characteristic split luminescence bands at 1833/1959 cm^{-1}.

Ligand Filed Theory Simulations. Luminescence spectra of the europium(III) ion in $Eu(H_2O)_9$ cluster and EuDEPA and EuEDTA optimized geometries, as simulated by the ligand (crystal) field theory, are plotted in Figure 4. The accuracy of the semiempirical approach is limited; for example, the experimental bands observed within 1860−1930 cm^{-1} are predicted at 1630−1790 cm^{-1} etc., and even bigger error is expected for the intensities. However, the model provides a solid basis for the band assignment. It is based on energy levels of free Eu^{3+} ion, because even in crystals and complexes, the orbital (L), spin (S), and total (J) quantum numbers are not quenched.[49] Using the usual notation $^{2S+1}L_J$, we can thus distinguish the $^5D_1 \rightarrow {}^7F_2$ (experimentally at 650−1010 cm^{-1}/

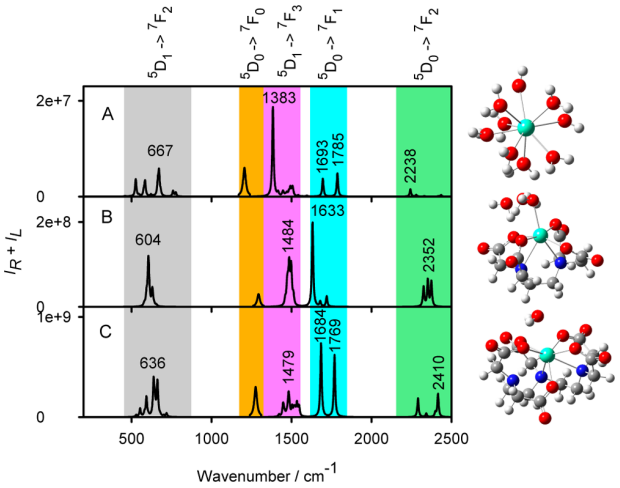


Figure 4. Simulated Raman/luminescence spectra of a $Eu(H_2O)_9$ cluster, EuEDTA, and EuDEPA ions. The $^5D_0 \rightarrow {}^7F_0$ bands were multiplied by 100 to be visible. The ligand (crystal)-field theory enables one to identify observed transitions.

calculated at 570−680 cm^{-1}), $^5D_1 \rightarrow {}^7F_3$ (1500−1550/1460−1510), $^5D_0 \rightarrow {}^7F_1$ (1860−1930/1630−1790), and $^5D_0 \rightarrow {}^7F_2$ (>2400/2300−2400) regions, in agreement with europium energies observed in other systems.[12] Even some experimentally observed intensity trends are predicted by this model, such as the lower intensity of the $^5D_0 \rightarrow {}^7F_1$ transitions of the hydrated Eu^{3+} ion compared to EuEDTA and EuDEPA, smaller signal of $^5D_1 \rightarrow {}^7F_3$ fluorescence in EuDEPA than in EuEDTA, and the split and shift of the EuDEPA $^5D_0 \rightarrow {}^7F_1$ bulk intensity toward higher wavenumbers compared to EuEDTA.

On a *qualitative* level, the crystal field theory can thus be used to simulate the CPL intensities stemming from the Eu^{3+} ion and reveal the chirality-induction mechanism. The spectra of europium complexes with the EDTA and DEPA ions and with α- and β-mannose are plotted in Figure 5. The mannose

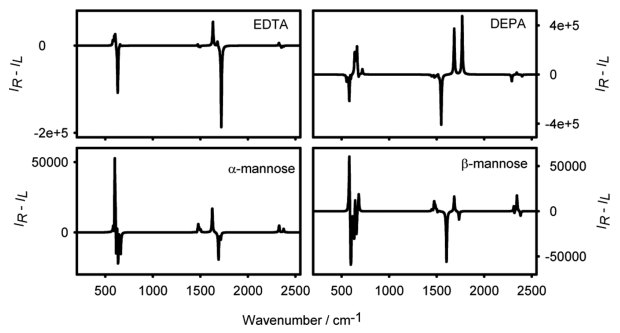


Figure 5. Simulated ROA/CPL spectra of Eu^{3+} complexes with EDTA, DEPA, and α- and β-mannose. The ligand-field theory qualitatively describes the experimental observations.

complexes were chosen as their geometry maximizes the number of Eu···O interactions. Note that the EuEDTA and DEPA complexes are chiral; in solution, they exist in an equilibrium of the "Δ" and "Λ" enantiomeric forms.[29,30]

The simulations are well consistent with the experiment in that the predicted CID ratios (2×10^{-4} to 1×10^{-3}) agree with the experimental range of dissymmetry factors found for the EDTA and DEPA complexes and bare Eu^{3+} ion. The CPL/total

DOI: 10.1021/acs.analchem.6b02505
Anal. Chem. 2016, 88, 8878−8885

Table 1. Fluorescence Decay Times (t_1 and t_2, in μs) and Relative Amplitudes (A_1 and A_2, in %) of Three Bands (around 592, 615, and 697 nm, Exact Peak Positions λ_{max} in nm) as Obtained by a Two-Exponential Fit $I = I_0 + A_1 \exp(-t/t_1) + A_2 \exp(-t/t_2)$[a]

	~592 nm					~615 nm					~697 nm				
	λ_{max}	t_1	t_2	A_1	A_2	λ_{max}	t_1	t_2	A_1	A_2	λ_{max}	t_1	t_2	A_1	A_2
$EuCl_3$	591.6	106	1430	85	15	615.8	100	1697	79	21	697.7	98	1933	77	23
$EuCl_3$+F[b]	591.6	101	384	73	27	615.8	98	474	75	25	696.9	82	866	69	31
$EuCl_3$+M[b]	591.2	105	459	75	25	615.8	103	462	69	32	695.9	113	516	71	29
EuEDTA	592.5	210	732	56	44	615.5	249	613	53	37	697.5	147	818	42	58
EuEDTA+F[b]	592.8	57	355	12	88	615.4	70	343	8	92	698.3	874	481	29	71
EuEDTA+M[b]	592.9	130	475	33	67	615.8	18	304	7	93	698.2	58	434	33	67
EuDEPA	594.5	110	745	11	89	615.4	158	708	8	92	694.8	97	796	12	88
EuDEPA+F[b]	594.2	41	603	17	83	615.2	35	572	9	91	694.8	37	563	21	79
EuDEPA+M[b]	594.2	104	766	13	87	615.5	135	717	9	91	695.1	89	817	14	86

[a]The excitation wavelength was 405 nm. [b]F, fructose; M, mannose.

fluorescence ratio is predicted much more reliably than the actual intensities, because absolute values of the transition moments are not known.[36,37] The plausible mechanisms of the chirality transfer involve the previously suggested perturbation of the $\Delta \leftrightarrow \Lambda$ equilibrium by a preferential binding to the sugar[11] but also a direct Eu^{3+}–sugar interaction. The comparable CPL intensities for all the systems in Figure 5 show that both mechanisms are possible, which is also consistent with the observations of high induced chirality, in both the complexes and bare (hydrated) Eu^{3+} ion (Figure 3). However, the actual mode of interaction of EuEDTA, DEPA, and Eu^{3+} with the sugars may all be similar also because the europium ion makes relatively stable aggregates with water. To some extent, its first solvation sphere thus behaves as a complex, too.[50,51]

Fluorescence Decay Times. The fluorescence decay times were determined for the most strongly interacting sugars, fructose and mannose. The times are summarized in Table 1 for the fluorescence at 590 nm. They are relevant for the most pronounced ROA/CPL signal at ~1850 cm^{-1} and confirm specificity of the interactions. All decay curves could be well fitted by a double-exponential function. Compared to typical organic molecules, rather long fluorescence times are observed, unique for the lanthanide electronic system and mostly involving the *f*-shell europium electronic levels.[12]

For example, the addition of mannose or fructose to $EuCl_3$ and EDTA shortens the decay times (in particular t_2), and the amplitude of the longer-time component (A_2) rises. This effect is significantly stronger for fructose than for mannose. The fluorescence kinetics of the DEPA complex is rather unperturbed by mannose, but there is some effect of fructose in shortening the shorter time t_1 from 110 to 41 μs and rising its amplitude A_2. This corresponds to the stronger chiroptical response of EDTA, as shown in Figure 2. The kinetic data including different fluorescence peaks are summarized in Figure 6 revealing similar sensitivity and specificity to the lanthanide–sugar interactions across the entire spectrum.

NMR Chemical Shift Changes. The chemical shifts induced in D_2O solutions of mannose and fructose by $EuCl_3$ and EuEDTA also indicate selective interactions of the europium ion and complexes with the monosaccharides, even though they are not as specific as for the fluorescence. They suggest complex binding with multiple binding sites and a rather weak interaction. In general, the addition of a paramagnetic ion into the sample was accompanied by both

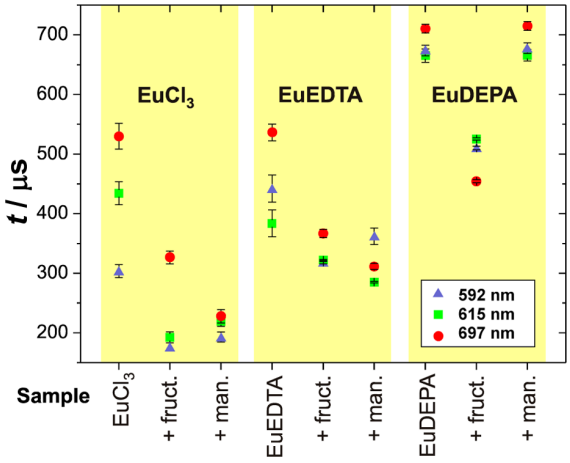


Figure 6. Average fluorescence decay times ($t = A_1 t_1 + A_2 t_2$) at three different wavelengths for the three europium compounds, with and without sugars.

line broadening (due to enhanced "T_2" relaxation) and chemical shift changes.

Assignment of the proton spectra was impossible because the monosaccharide solutions contained two (α/β mannose anomers) or four (α/β anomers for each furanose and pyranose fructose form) sugar isomers, most of the ^1H signals clustered in a very narrow chemical shift range (3.2–4.0 ppm), and after the addition of europium compounds the lines became too broad to be assignable. Therefore, we focused on ^{13}C NMR spectra, where the signals were well separated in most cases, and the line broadening did not prevent signal assignment and interpretation. The signals were referenced to the C_6D_6 external standard (sealed in a capillary), i.e., europium-free.

Mannose exists in water solution as a mixture of α- and β-D-mannopyranose in a ratio of about 2:1. Both the $EuCl_3$ and EuEDTA solutions caused a downfield shift (higher chemical shift values) of all mannose ^{13}C signals. The chemical shift change is almost uniform across all carbon atoms (1 ppm for 2:1 ratio of mannose–$EuCl_3$ and about 0.5 ppm for the same stoichiometric mixture of mannose–EuEDTA), which may suggest that there is not a single strongly preferred geometry of the metal–sugar interaction. A typical dependence of relative chemical shifts on europium concentration is exemplified in Figure 7 for the C3 carbon atom. A closer look, however,

DOI: 10.1021/acs.analchem.6b02505
Anal. Chem. 2016, 88, 8878–8885

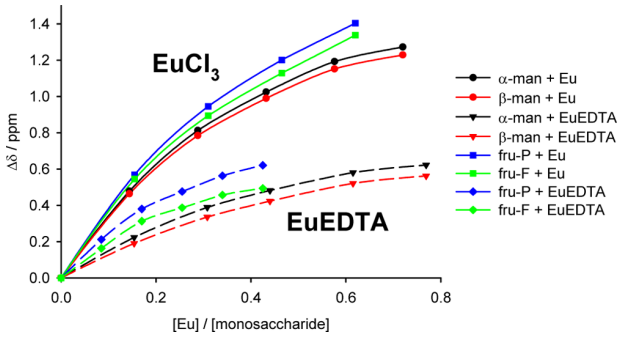


Figure 7. Changes of ^{13}C chemical shifts ($\Delta\delta$) of carbon C3 in mannose and fructose, upon addition of $EuCl_3$ and EuEDTA: α-man, α-D-mannopyranose; β-man, β-D-mannopyranose; fru-P, β-D-fructopyranose; fru-F, β-D-fructofuranose. The variations are small relative to the overall shift but specific for a particular sugar or carbon type.

reveals that the chemical shift changes are not completely uniform and slight differences between individual carbon atoms exists. Interestingly, such differences between individual carbon atoms are much more pronounced for the interaction with EuEDTA complex, which may indicate that the specificity of the EuEDTA–sugar interaction is greater than for $EuCl_3$. It is true that the EuEDTA average chemical shift changes are about two times smaller than those caused by $EuCl_3$, but this is likely to be caused by the EDTA ligands, at least partially or fully preventing a direct europium–sugar binding.

At equilibrium, fructose in water solution is present as a mixture of two predominant forms (70% of β-D-fructopyranose and 21% of β-D-fructofuranose), together with two minor α-forms.[52] As for mannose, an addition either of $EuCl_3$ or EuEDTA caused a downfield shift of all fructose ^{13}C signals. Interestingly, differences between individual chemical shift changes induced by $EuCl_3$ were much higher for the pyranose rather than the furanose form (Figure 8), which indicates that the pyranose–europium interaction may be linked to a more distinct complex structure. The addition of EuEDTA led to significantly broader carbon signals (particularly C2, C4, C5, and C6 in β-D-fructopyranose; C1, C2, and C4 in β-D-fructofuranose) than did similar amounts of added $EuCl_3$, indicating a stronger binding of the former. The NMR spectra

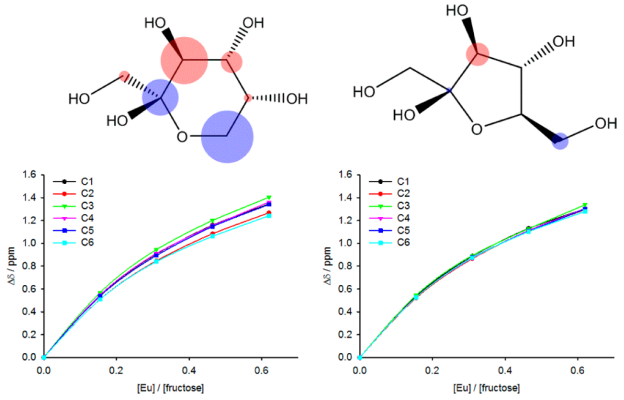


Figure 8. ^{13}C chemical shifts changes ($\Delta\delta$) of carbon atoms in β-D-fructopyranose (left) and in β-D-fructofuranose (right), as caused by the addition of $EuCl_3$. Relative deviations from the average shift change are indicated for all carbons by disks of different diameters (positive, red; negative, blue) in the structures.

thus confirmed the specificity of the europium compounds–sugar interactions and suggested a weak, multisite binding.

MD Modeling of the Europium–Sugar Interactions. Currently, we find it too difficult to reliably simulate all aspects of interactions of the larger EuEDTA and EuDEPA complexes with the monosaccharides. However, for the free europium ion, the MD simulations do reveal the basic binding patterns and energy changes associated with the complexation. The formation free energies of most favored geometries listed in Table 2 suggest that the complexes are rather unstable; the

Table 2. Free Energies of the Sugar–Eu^{3+} Complex Formation, As Calculated Using the Molecular Dynamics and the WHAM Method

sugar	isomer	oxygens bound to Eu	ΔG (kcal/mol)
α-fructopyranose	af1	O_3 O_4	1.0
	af2	O_1 O_3 O_6	1.0
	af3	O_2 O_4 O_5 O_6	0.95
	af4	O_1 O_2 O_6	0.6
β-fructopyranose	bf1	O_1 O_4 O_5 O_6	1.0
	bf2	O_2 O_3 O_6	1.0
	bf3	O_1 O_2 O_6	0.75
α-fructofuranose	cf1	O_1 O_3 O_6	0.9
	cf2	O_2 O_4 O_5	0.8
	cf3	O_1 O_5 O_6	0.7
	cf4	O_1 O_2 O_5	0.6
β-fructofuranose	df1	O_2 O_3 O_6	1.1
	df2	O_1 O_2 O_3	0.9
	df3	O_1 O_2 O_6	0.8
α-mannose	am1	O_1 O_5 O_6	0.9
	am2	O_1 O_2	0.8
	am3	O_2 O_3	0.7
	am4	O_2 O_5 O_6	0.5
β-mannose	bm1	O_1 O_2 O_5 O_6	0.9
	bm2	O_2 O_3	0.9

biggest stabilization energies (~1 kcal/mol) are comparable with the Boltzmann temperature quantum (~0.6 kcal/mol at 300 K). In addition, many approximately equally convenient binding sites are possible at ambient temperature, which is well in agreement with the NMR data discussed above, and binding to more hydrogen atoms often does not yield a more stable complex.

The process of complex formation can also be understood on the whole profiles of the mean force potentials (free energies) obtained by the WHAM method. They are quite similar (Figure S6), and the lowest-energy isomer of the β-fructopyranose/Eu^{3+} complex was selected as an example in Figure 9. Here, we can see a free (IV) and weakly stabilized (~0.2 kcal/mol) preassociation state (III) of the hydrated europium ion surrounded by nine water molecules. The actual binding to the sugar requires a destruction of this hydration shell, which is associated with a relatively high energy (~1 kcal/mol) of the transition state (II). Finally, the most stable complex (I) is stabilized, by about 1 kcal/mol.

Because of the high energy needed to break the europium hydration shell, the bound state (I in Figure 9) could not be obtained from free dynamics in a reasonable time. However, the more weakly associated states (III, essentially a complex of the sugar and $[Eu(H_2O)_9]^{3+}$ ion) are visible in the europium probability plot based on free MD. In the example for β-D-

DOI: 10.1021/acs.analchem.6b02505
Anal. Chem. 2016, 88, 8878–8885

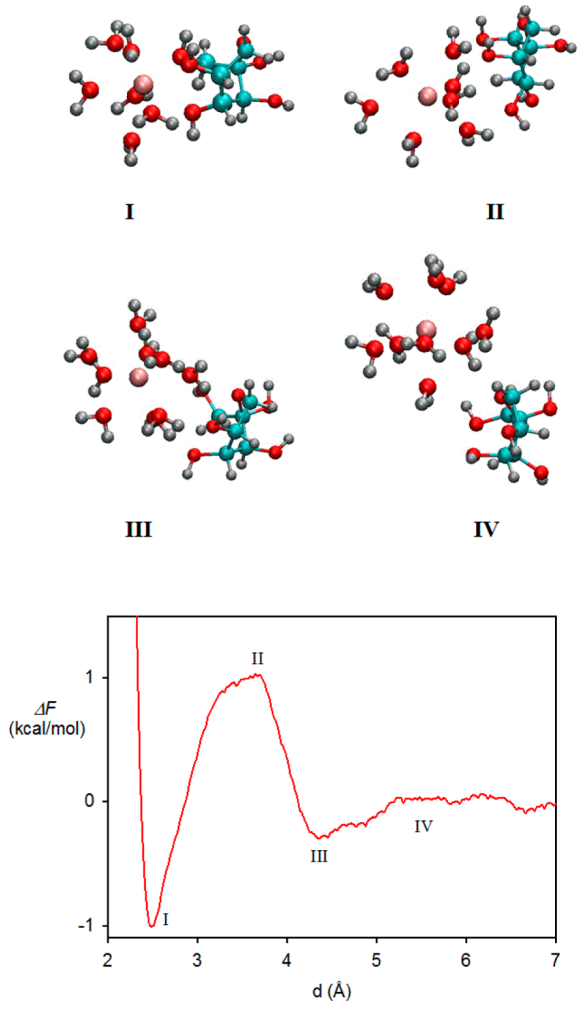


Figure 9. Typical WHAM free energy profile for sugar−europium(III) binding (this one for *β*-fructopyranose) and an example of geometries along the reaction coordinates. The complex formation requires the europium hydration shell to be disturbed, which is associated with a relatively high activation energy.

fructopyranose in Figure 10, we can see that also for this interaction preferential sites exist.

For the direct Eu−sugar complexes (I), the equilibrium Eu···O distance of ∼2.4 Å agrees well with available crystallographic data.[3,4] Geometries of the most stable complexes of various fructose and mannose forms are plotted in Figure 11, and their variability is thus consistent with the rich spectroscopic responses.

Figure 10. Regions of highest density of the Eu^{3+} ion obtained from a 300 ns molecular dynamics run with *β*-D-fructopyranose corresponding to part III in Figure 9.

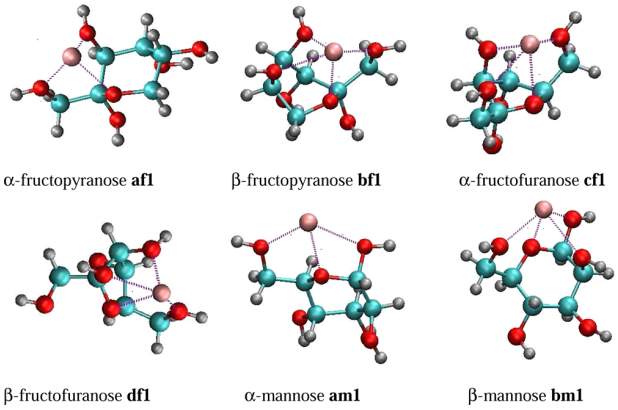


Figure 11. Some lowest-energy conformers of fructose and mannose complexes with Eu^{3+} (cf. Table 2).

CONCLUSIONS

We have explored the interaction of europium compounds with common monosaccharides using ROA/CPL, time-dependent luminescence and NMR spectroscopies. Very specific spectral patterns have been observed for the circularly polarized luminescence, for both the sugar and europium components. The complexation specificity of the spectral response was confirmed by measurement of the fluorescence decay times. In NMR spectra, the paramagnetic lanthanide metal caused nearly uniform chemical shift of the sugar carbon atoms; finer relative shift changes, however, were also very specific to the lanthanide compound and monosaccharide type. The luminescence/CPL spectral bands could be assigned and semiqualitatively modeled using the crystal field theory. Combined, the NMR experiment and molecular dynamics simulations suggest that multiple binding modes for each sugar form are possible, although the link between the actual geometry and detailed spectral data, especially for the interaction of the EDTA and DEPA complexes, still awaits elucidation. The ROA/CPL methodology appears to be a handy tool for studies of structure and interactions of sugars, requiring their identification and detection.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b02505.

Further experimental and computational details (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: bour@uochb.cas.cz.

*E-mail: jan.valenta@mff.cuni.cz.

Notes

The authors declare no competing financial interest.

DOI: 10.1021/acs.analchem.6b02505
Anal. Chem. 2016, 88, 8878−8885

■ ACKNOWLEDGMENTS

The work was supported by the Grant Agency of the Czech Republic (16-05935S, 16-08764Y, 15-11223S, and 15-09072S).

■ REFERENCES

(1) Striegler, S.; Dittel, M. *J. Am. Chem. Soc.* **2003**, *125*, 11518.
(2) Davis, A. P.; Wareham, R. S. *Angew. Chem., Int. Ed.* **1999**, *38*, 2978.
(3) Yang, L.; Su, Y.; Xu, Y.; Zhang, S.; Wu, J.; Zhao, K. *J. Inorg. Biochem.* **2004**, *98*, 1251.
(4) Guo, J.; Lu, Y. *J. Carbohydr. Chem.* **2010**, *29*, 10.
(5) Yang, L.; Hua, X.; Xue, J.; Pan, Q.; Yu, L.; Li, W.; Xu, Y.; Zhao, G.; Liu, L.; Liu, K.; Chen, J.; Wu, J. *Inorg. Chem.* **2012**, *51*, 499.
(6) Alptürk, O.; Rusin, O.; Fakayode, S. O.; Wang, W.; Escobedo, J. O.; Warner, I. M.; Crowe, W. E.; Král, V.; Pruet, J. M.; Strongin, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9756.
(7) Battistini, E.; Mortillaro, A.; Aime, S.; Peters, J. A. *Contrast Media Mol. Imaging* **2007**, *2*, 163.
(8) Lu, Y.; Deng, G.; Miao, F.; Li, Z. *Carbohydr. Res.* **2003**, *338*, 2913.
(9) Wang, W.; Gao, X.; Wang, B. *Curr. Org. Chem.* **2002**, *6*, 1285.
(10) Crescenzi, V.; Brittainc, H. G.; Yoshinoc, N.; Okamoto, Y. *J. Polym. Sci., Polym. Phys. Ed.* **1985**, *23*, 437.
(11) Huskowska, E.; Riehl, J. P. *Inorg. Chem.* **1995**, *34*, 5615.
(12) Binnemans, K. *Coord. Chem. Rev.* **2015**, *295*, 1.
(13) Riehl, J. P.; Muller, G. In *Comprehensive chiroptical spectroscopy, volume 1: Instrumentation, methodologies, and theoretical simulations*; Berova, N., Polavarapu, P. L., Nakanishi, K., Woody, R. W., Eds.; John Wiley & Sons: Hoboken, NJ, 2012; Vol. *1*, p 65.
(14) Wu, T.; Kapitán, J.; Mašek, V.; Bouř, P. *Angew. Chem., Int. Ed.* **2015**, *54*, 14933.
(15) Hug, W. *Appl. Spectrosc.* **2003**, *57*, 1.
(16) Nafie, L. *Vibrational optical activity: Principles and applications*; Wiley: Chichester, 2011.
(17) Barron, L. D. *Biomed. Spectrosc. Imaging* **2015**, *4*, 223.
(18) Johannessen, C.; Pendrill, R.; Widmalm, G.; Hecht, L.; Barron, L. D. *Angew. Chem., Int. Ed.* **2011**, *50*, 5349.
(19) Zhu, F.; Isaacs, N. W.; Hecht, L.; Barron, L. D. *J. Am. Chem. Soc.* **2005**, *127*, 6142.
(20) Melcrová, A.; Kessler, J.; Bouř, P.; Kaminský, J. *Phys. Chem. Chem. Phys.* **2016**, *18*, 2130.
(21) Zielinski, F.; Mutter, S. T.; Johannessen, C.; Blanch, E. W.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2015**, *17*, 21799.
(22) Barron, L. D. *Molecular Light Scattering and Optical Activity*; Cambridge University Press: Cambridge, 2004.
(23) Šugar, J.; Bouř, P. *J. Raman Spectrosc.* **2016**, DOI: 10.1002/jrs.4960, in press.
(24) Gawryszewska, P.; Legendziewicz, J.; Ciunik, Z.; Esfandiari, N.; Muller, G.; Piguet, C.; Cantuel, M.; Riehl, J. P. *Chirality* **2006**, *18*, 406.
(25) Moussa, A.; Pham, C.; Bommireddy, S.; Muller, G. *Chirality* **2009**, *21*, 497.
(26) Nguyen, B. T.; Ingram, A. J.; Muller, G. *Chirality* **2016**, *28*, 325.
(27) Valenta, J.; Greben, M. *AIP Adv.* **2015**, *5*, 047131.
(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; *Gaussian 09*, Revision D01 ed.; Gaussian, Inc.: Wallingford CT, 2009.
(29) Gao, J. Q.; Wu, T.; Wang, J.; Bai, Y.; Wang, S. J.; Xu, Y. N.; Li, Y.; Zhang, X. D. *Russ. J. Coord. Chem.* **2012**, *38*, 491.
(30) Mondry, A.; Janicki, R. *Dalton Trans.* **2006**, 4702.
(31) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(32) Figgen, D.; Rauhut, G.; Dolg, M.; Stoll, H. *Chem. Phys.* **2005**, *311*, 227.
(33) Scalmani, G.; Frisch, M. J. *J. Chem. Phys.* **2010**, *132*, 114110.
(34) Carnall, W. T.; Goodman, G. L.; Rajnak, K.; Rana, R. S. *J. Chem. Phys.* **1989**, *90*, 3443.
(35) Edvardsson, S.; Åberg, D. *Comput. Phys. Commun.* **2001**, *133*, 396.
(36) Judd, B. R. *Phys. Rev.* **1962**, *127*, 750.
(37) Ofelt, G. S. *J. Chem. Phys.* **1962**, *37*, 511.
(38) Richardson, F. S.; Faulkner, T. R. *J. Chem. Phys.* **1982**, *76*, 1595.
(39) Case, D. A.; Cheatham, I. T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, J. K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*, 1668.
(40) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Outeiriño, J. G.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622.
(41) Baaden, M.; Burgard, M.; Boehme, C.; Wipff, G. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1317.
(42) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(43) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275.
(44) Barron, L. D.; Gargaro, A. R.; Wen, Z. Q. *Carbohydr. Res.* **1991**, *210*, 39.
(45) Bell, A. F.; Hecht, L.; Barron, L. D. *J. Raman Spectrosc.* **1993**, *24*, 633.
(46) Cheeseman, J. R.; Shaik, M. S.; Popelier, P. L. A.; Blanch, E. W. *J. Am. Chem. Soc.* **2011**, *133*, 4991.
(47) Wen, Z. Q.; Barron, L. D.; Hecht, L. *J. Am. Chem. Soc.* **1993**, *115*, 285.
(48) Profant, V.; Pazderková, M.; Pazderka, T.; Maloň, P.; Baumruk, V. *J. Raman Spectrosc.* **2014**, *45*, 603.
(49) Walrand, C. G.; Binnemans, K. In *Handbook on the physics and chemistry of rare earths*; Gschneider, K. A., Eyring, L., Eds.; Elsevier Science B. V.: Amsterdam, 1996; p 121.
(50) Chaussedent, S.; Monteil, A. *J. Chem. Phys.* **1996**, *105*, 6532.
(51) Clavaguéra, C.; Pollet, R.; Soudan, J. M.; Brenner, V.; Dognon, J. P. *J. Phys. Chem. B* **2005**, *109*, 7614.
(52) Barclay, T.; Ginic-Markovic, M.; Johnston, M. R.; Cooper, P.; Petrovsky, N. *Carbohydr. Res.* **2012**, *347*, 136.

DOI: 10.1021/acs.analchem.6b02505
Anal. Chem. 2016, 88, 8878−8885