



Normal mode calculation and infrared spectroscopy of proteins in water solution: Relationship between amide I transition dipole strength and secondary structure

Francesca Ripanti^a, Nicole Luchetti^b, Alessandro Nucara^{a,*}, Velia Minicozzi^b, Almerinda Di Venere^c, Alessandra Filabozzi^b, Marina Carbonaro^d

^a Department of Physics, Sapienza University of Rome, P.le A. Moro 5, 00185 Rome, Italy

^b Department of Physics and INFN, Tor Vergata University of Rome, Via della Ricerca Scientifica 1, 00133 Rome, Italy

^c Department of Experimental Medicine, Tor Vergata University of Rome, Via Montpellier 1, 00133 Rome, Italy

^d Council for Agricultural Research and Economics (CREA), Research Centre for Food and Nutrition, Via Ardeatina 546, 00178 Rome, Italy

ARTICLE INFO

Keywords:
Infrared spectroscopy
Transition dipole moment
Amide I

ABSTRACT

Dipole Strength (DS) of the amides has gained a renewed interest in chemical physics since it provides an important tool to disclose the on-site vibrational energy distributions. Apart from earlier experimental efforts on polypeptides, little is still known about DS in complex proteins. We accurately measured the Fourier Transform Infrared absorption spectra of nine proteins in water solution obtaining their Molar Extinction Coefficient in the amide I and II spectral region. Our results show that the amide I DS value depends on the protein secondary structure, being that of the α -rich and unstructured proteins lower by a factor of 2 than that of the β -rich proteins. The average DS for amino acids in α and β secondary structures confirms this finding. Normal Mode calculation and Molecular Dynamics were performed and used as tools for data analysis and interpretation. The present outcomes corroborate the hypothesis that antiparallel β -sheet environment is more prone to delocalize the on-site C=O stretching vibration through coupling mechanisms between carbonyl groups, whereas α -helix structures are energetically less stable to permit vibrational mode delocalization.

1. Introduction

One of the most intriguing issue in biomolecule science is the understanding of mechanisms for protein misfolding and aggregation, problems that generate concern for the variety and complexity of the different scenarios [1,2]. Since the early time, Fourier Transform InfraRed (FTIR) spectroscopy has been widely exploited to investigate protein aggregation pathways, by virtue of its capability to provide information on the secondary and tertiary structures and to disclose textured/unstructured aggregate kinetics [3–6]. FTIR experiments mainly focus on the amide I (C=O stretching mode of the carbonyl bond) and amide II (mainly N–H bending vibrations) bands, but only the former is considered reliable for the study of protein structural changes. Generally, the secondary structure is inferred upon deconvolution of the amide I band and by evaluation of the relative component intensity [7,8]. The major secondary structures, i.e. α -helix, unstructured, and antiparallel β -structures, provide absorption features centered tens

cm^{-1} apart, thus demanding narrowing methods for their recognition. Moreover, other secondary structures as turn and unordered structures might contribute to amide I absorption, but, in most cases, they are hardly distinguishable from the principal ones. Spectral deconvolution and band narrowing suffer for arbitrariness in the number and in the intensity of the contributions, preventing in most cases unambiguous assignment [9,10]. Besides, the assumption of an identical infrared response for each secondary structure is quite arbitrary, signaling the difficulty to assess the strength of the C=O vibrational mode in the different amino acid arrangements.

According to the Born-Oppenheimer approximation, the Transition Dipole Moment (TDM) of an oscillator at frequency ω and normal coordinate Q is proportional to $\frac{1}{\omega^{1/2}} \langle \Psi_0 | \frac{\partial \vec{\mu}}{\partial Q} \Psi_0 \rangle$, where $\vec{\mu}$ is the molecule dipole moment and Ψ_0 its electronic ground state. This expression, primarily employed in theoretical contexts, is useless in spectroscopic investigations, since only the Dipole Strength (DS) of the vibration modes

* Corresponding author.

E-mail address: alessandro.nucara@roma1.infn.it (A. Nucara).

<https://doi.org/10.1016/j.ijbiomac.2021.06.092>

Received 4 April 2021; Received in revised form 11 June 2021; Accepted 13 June 2021

Available online 19 June 2021

0141-8130/© 2021 Elsevier B.V. All rights reserved.

is achievable from the integrated intensity of the spectral bands. Several analytical expressions are available in the literature to evaluate the dipole strength from the experimental data, once the Molar Extinction Coefficient (MEC) is exactly estimated: therefore, efforts in FTIR experiments are nowadays promoting quantitative measurements of the absorption band intensities. Actually, to the best of our knowledge, only few studies report on the intensity of protein secondary structures and there is not general agreement on the expected DS values for the α -helix, β -sheet, and unordered moieties [11,12]. Early studies on the amide intensities date back to the nineties, especially devoted to polypeptide chains and small proteins [13,14]. These seminal experiments provided the amide I intensity of peptides arranged in α -helix, β -sheet, and unordered structures, and yielded identical values for α and β structures (DS for single amino acids (aa) of $0.09 \text{ D}^2/\text{aa}$), whereas the average intensity of unordered structure was foreseen 1.7 times smaller. However, literature papers of the same period reported rather contradictory results, forcing to prudently extend these outcomes to complex systems. A systematic study on globular proteins was carried out in Ref. [12]: therein, different DS values for β -sheet, α -helix, and unordered structures were inferred, with relative intensities in the ratio 1: 0.7: 0.5, respectively.

Studies on DS of protein C=O gained a renewed interest with the advent of Density Functional Theory (DFT) and two-dimensional (2D) spectroscopy. Among many theoretical studies, Wang [15] assessed substantial independence of TDM from the secondary structure and suggested a DS value of $0.1 \text{ D}^2/\text{aa}$, in a fair agreement with those reported in Ref. [16]. Recently, a comprehensive DFT calculation in small polypeptides [17] found dipole strengths of α and β structures lower than those previously cited (0.048 and $0.06\text{--}0.07 \text{ D}^2/\text{aa}$, respectively). 2D infrared spectroscopy showed its capability to infer on the excitonic nature of C=O stretching in peptides and small amyloid oligomers [18,19], in line with exciton theory [20]. Non-linear IR results outlined the spread of the local stretching vibration over neighboring carbonyl groups, which returns much higher values of the DS for helix structures ($0.26 \text{ D}^2/\text{aa}$) [18]. In the same work, 2D spectra provided DS values of unstructured polypeptides in line with the earlier findings ($0.1 \text{ D}^2/\text{aa}$).

The above scenario provides a rather contradictory picture: on one hand, identical values of DS/aa are inferred for proteins and polypeptides arranged in different structures; on the other side, differences between secondary structures are found in 2D and in linear IR spectroscopy. It is well recognized [21] that carbonyl groups are coupled via Transition Dipole Coupling (TDC), which delocalizes the local modes over neighboring peptides [20] in form of Frenkel-like excitons. The TDC intensity among neighboring peptides varies from 20 to 7 cm^{-1} [22], depending on the secondary structure arrangement, the intra- and inter-strand potential, the molecular weight (MW), and the secondary structure length and packing. However, Wang [23] and Grechko [18,19] demonstrated that, whenever the spread Δ of the oscillators energy significantly exceeds the coupling strength, delocalization drops out and it is not possible to infer on the local or excitonic nature of the vibration, as schematically represented in Fig. 1. This argument provides a rule of thumb to assess the intra-molecular order-disorder degree, since the exciton collapse/delocalization is a well-known topic in molecular physics [24,25].

According to this background, our work is aimed at disclosing the DS value of several globular proteins in aqueous solution. We focused on two groups of proteins: those mainly folded in α -helix and those with a majority of antiparallel β -sheet secondary structures. Moreover, we also investigated one intrinsically unordered structure for completeness. To achieve reliable DS values, proteins were dissolved in H_2O solution, avoiding band-shift effects observed in D_2O [26]. The simultaneous detection of amide I and II bands, the latter not detectable in D_2O protein solution, permits to evaluate the intensity ratio of the two amides and returns a criterium for the exactness of the background correction.

Molecular Dynamics (MD) and Normal Mode (NM) calculations were also performed with a twofold aim: i) to characterize the secondary structure of the investigated proteins, and ii) to evaluate the molecule infrared active normal modes. The former aim was prompted by the multiplicity of secondary structure percentages available in literature, a variety that depends on the employed experimental technique and on the protein status itself. Since this information is crucial for the data interpretation, we believe it is more appropriate to refer to widespread

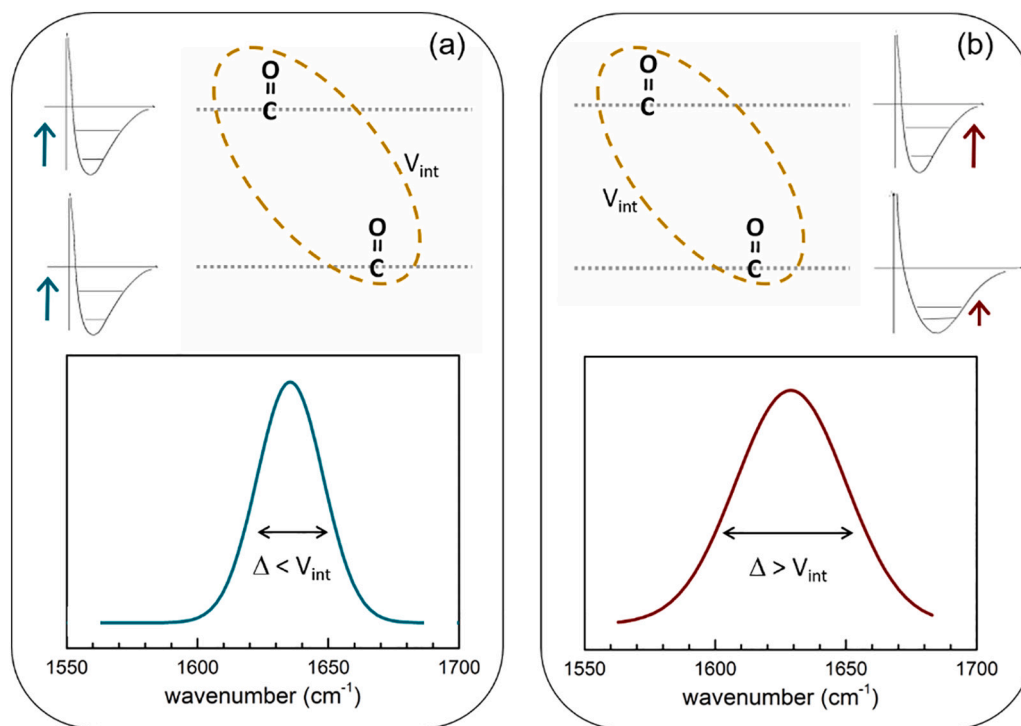


Fig. 1. Cartoon showing the competing condition for exciton collapse/formation in a chain: (a) identical C=O oscillators results in a small inhomogeneous broadening Δ and vibrational excitons formation; (b) large Δ width from different on-site potential does not allow delocalization.

established models. Item ii) provided the vibrational modes distribution in the limit of incoherent C=O oscillators, to be considered as a useful tool for experimental data comparison.

Our results show that value DS of the single C=O is different in the α - and β -rich proteins, implying that the secondary structure is a relevant factor for the tuning of TDC/ Δ competing effects, which might be at the basis of the different DS values reported in the literature.

2. Materials and methods

All proteins were supplied by Merck KGaA (Darmstadt, Germany). Protein solutions were prepared dissolving powder of lyophilized protein in Milli-Q water. We measured the pH value of any solutions to verify that the native folding of the protein and its oligomeric structure were preserved. The protein concentration was monitored by the absorbance at $\lambda = 280$ nm of the centrifuged solution using the values of the molar extinction coefficient at this wavelength (ϵ_{280}) available in the literature. It is worth mentioning that the spread of ϵ_{280} value provides a relative uncertainty on the concentration of 2–10%, an error that consequently affects the MEC values. The proteins studied in this work are listed in Table 1.

FTIR measurements were performed with a Michelson interferometer (Bruker IFS66V/S) equipped with a DTGS detector and KBr beamsplitter. The apodized spectral resolution was 2 cm^{-1} and the interferometer was purged with dried air to minimize artefacts from atmospheric absorption. It is worth noticing that even most intense amide I and II absorptions decrease the reference signal by less than 2% in the region of the absorption band of water ($1500\text{--}1700\text{ cm}^{-1}$), thus the reproducibility of the experimental conditions was carefully checked. To achieve data reliability, a custom-made spectroscopic cell was specifically realized for this experiment, consisting of a polyethylene piston drilled along the axes, which can be moved inside a steel cylinder. The space between the cylinder base and the piston, both sealed with CaF_2 windows, comprises the volume for the sample solutions. Typically, the sample volume used in this experiment was $30\ \mu\text{L}$. Piston translation was micrometrical controlled by an Arduino microprocessor [27] achieving the accuracy of $1.4\ \mu\text{m}/\text{step}$. Arduino also interfaces the cell and the interferometer with an automated procedure. Data acquisition was performed in two phases: in the first step, spectra of the buffer (as reference) were acquired by changing the optical path

Table 1

List of the protein studied in this work, with the purity provided by the supplier, the molecular weight, the total number of amino acids, the measured concentration and pH values.

Protein (acronym)	Purity	MW (kDa)	N_{tot}	C (μM)	pH
Bovine Serum Albumin (BSA)	$\geq 98\%$	66.4	607	484 ± 30 246 ± 5 96 ± 5 50 ± 4	7
Cytochrome C (CytC)	$\geq 99\%$	12.4	105	1200 ± 100	6.5
Human Insulin (HI)	$\geq 98\%$	5.7	51	1800 ± 200	3
Concanavalin A (ConA)	spectroscopic grade	104	948	184 ± 30	6
β -Lactoglobulin (BLG)	$\geq 90\%$	18.3	178	393 ± 60	4.5
Soybean Trypsin Inhibitor (STI)	$\geq 95\%$	20	216	820 ± 26	7.5
Hemoglobin (Hb)	$\geq 96\%$	64.5	587	177 ± 29	6
Avidin (AVI)	$\geq 98\%$	67	532	214 ± 30	6.5
β -Casein (β -Cas)	$\geq 98\%$	24	209	1300 ± 200	6

length within the range $3\text{--}20\ \mu\text{m}$ at the stepping resolution of the cell; in the next step, after cell cleaning, the same procedure was identically repeated with the protein solution. Pairs of reference and sample spectra, which met the criteria established in Ref. [28], were used for the calculation of the absorbance $A(\omega)$ once corrected to the water vapor [29]. An example of compatible reference-sample spectra pair is shown in Fig. 2a. In order to accurately evaluate the optical path length, we measured the intensity of the bending absorption lines of the reference at 1645 cm^{-1} and we compared this value with the theoretical one reported in Ref. [30]. The accuracy uncertainty on the path length determination does not exceed 5% (see Fig. 2b).

The efficiency of the experimental setup and procedure was confirmed by the BSA absorption spectra reported in Fig. 2c. Therein, the shape of the spectrum is well reproduced at any protein concentrations and absorption structures are detected over a spectral region as large as 500 cm^{-1} . Alongside the most intense amide I and amide II bands, the spectra clearly show the amide III bands in the $1200\text{--}1350\text{ cm}^{-1}$ region as well as the C–H bending mode at 1447 cm^{-1} . In addition, the 1400 cm^{-1} band of ionized residues of glutamic and aspartic acids [31] is evident. The ratio between amide I and amide II intensities provides a marker to estimate the exactness of the spectrum shape: in spectra of Fig. 2c this ratio holds 1.5, in complete agreement with the values reported for not-saturated spectra of BSA [32].

MEC was obtained by inversion of the Lambert-Beer law, $A(\omega) = \epsilon(\omega)Cd$, being d the optical path length and C the protein molar concentration. Uncertainty on the molar extinction coefficient, resulting from the concentration and the optical path errors, is estimated around 15%. In order to retain only the spectral contributions of the peptide backbone, we subtracted absorptions arising from the side chains of selected amino acids, i.e. aspartic and glutamic acids (absorption in the range $1570\text{--}1600\text{ cm}^{-1}$), asparagine and glutamine ($1670\text{--}1680\text{ cm}^{-1}$), and lysine (1635 cm^{-1}). Spectra of these amino acid residues have been retrieved from literature [33]. The spectrum of ConA, before and after correction for the side chains, is reported in Fig. 3: changes are evident mainly in the $1570\text{--}1600\text{ cm}^{-1}$ region, where contributions from intermolecular aggregates are also expected, providing evidence of a negligible amount of aggregates in all samples here investigated.

Molecular Dynamics (MD) simulations were performed with the GROMACS package [34]. The initial coordinates of protein atoms were obtained from Protein Data Bank (pdb-ids are indicated in Table 2). Each protein was placed in a cubic box of dimensions such that nearby images lay $10\ \text{\AA}$ away. The box was filled with TIP3P water molecules and an appropriate number of counter-ions to make the whole system neutral. Among several methods used for calculations [35–37], the CHARMM22* force field was here used since it is supposed to reproduce quite well the transitions between α -helix and β -sheet secondary structures and to allow a rather accurate helix-coil balance thanks to a term added to the dihedral potential (the so called CMAP term) [38]. Moreover, CHARMM22* force field includes a cross-term accounting for angle bending using 1,3 nonbonded interactions, the Urey-Bradley term [39–41] that allows a more accurate evaluation of molecules vibrations with respect to other commonly used force fields (see Fig. S3 in Supporting Information, SI).

The equilibration strategy adopted was fully standard [42] consisting firstly of steepest descent and conjugated gradient minimization followed by 10 ns in the NVT- and 10 ns in the NPT-ensembles. After equilibration, 3×500 ns production simulations were performed in the NPT-ensemble. The temperature was kept fixed at 300 K, using the v-rescale thermostat [43] with a coupling time of 0.1 ps, and the pressure at 1 bar, using the Parrinello-Rahman barostat [44,45] with a coupling time of 2 ps and an isothermal compressibility of $4.5 \cdot 10^{-5}\text{ bar}^{-1}$. The Particle Mesh Ewald algorithm [46] was employed to treat Coulomb interaction. A time step of 2 fs and a non-bond pair list cut-off of 1.0 nm were used. The list was updated every 10 steps. The analysis of the numerical data obtained in the simulation was carried out by GROMACS and VMD [47] tools according to needs.

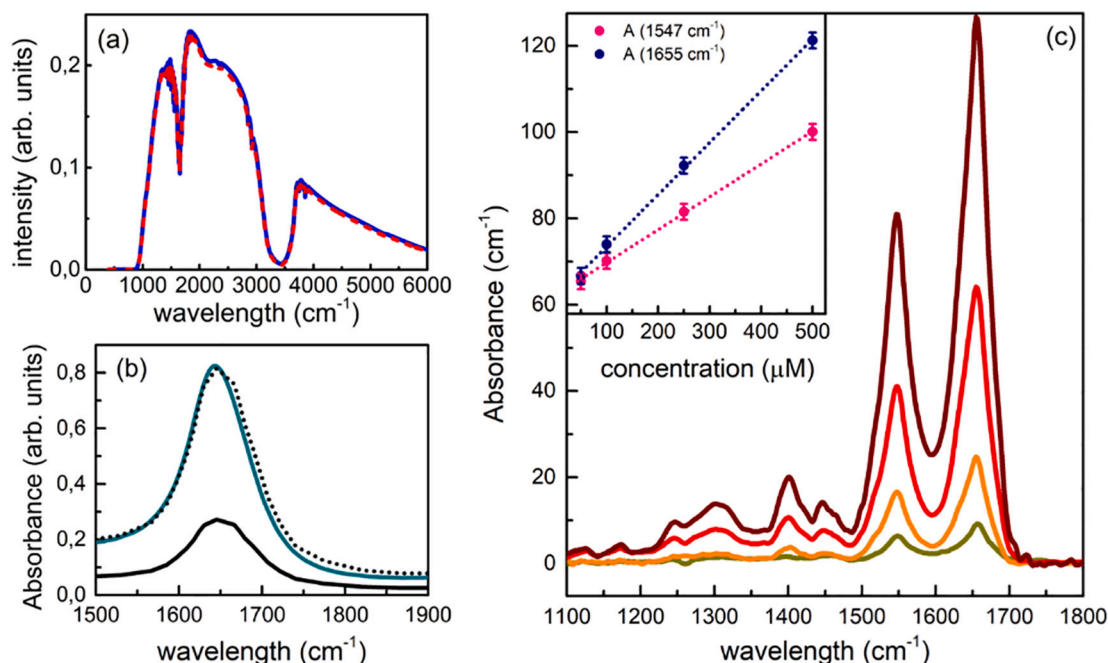


Fig. 2. (a) Example of reference (blue continuous line) and protein solution spectra (dashed red line); (b) H₂O bending line as measured from the reference spectrum (cyan continuous line). The optical path length is provided by a fit to the bending line (dotted line) with the theoretical imaginary refractive index of water reported in Ref. [30] (black line); (c) BSA absorption spectra in the mid-infrared region, collected at the protein concentrations reported in Table 1 (wine: 484 μM, red: 246 μM, orange: 96 μM, green: 50 μM). In the inset of panel c, the value of amide I and amide II maxima are reported vs BSA concentration as a validation of the Lambert-Beer relation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

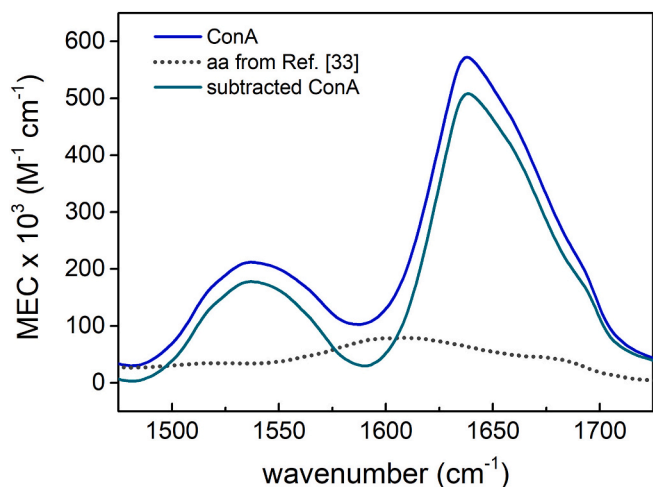


Fig. 3. Molar extinction coefficient of ConA before (blue line) and after (cyan) correction to Gln, Arg, Lys, Asp, and Glu absorptions in the amide I and II region. Dotted line represents the excess of absorption subtracted from the raw spectrum. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Normal Mode (NM) calculations were performed by firstly minimizing each protein structure extracted from MD simulations with the L-BFGS algorithm [48] as implemented in GROMACS (see SI for details). Afterwards, the Hessian matrix was computed and with the help of gm_x nmeig GROMACS routine diagonalized. The same routine allows to perform the sorting of the normal modes according to their frequencies and, for any subset of normal modes, gm_x nmens GROMACS routine calculates an ensemble of structures from which the vibrating bonds may be observed.

Table 2

Proteins, identity number in protein databank, simulated system composition and number of independent simulations (replicas) used in MD calculations.

Protein	pdb-id	Composition	Replicas
BSA	4f5s	BSA + 43,248 H ₂ O + 147 Na ⁺ + 131 Cl ⁻	3 × 500 ns
CytC	3zcf	CytC + 6901 H ₂ O + 20 Na ⁺ + 27 Cl ⁻	3 × 500 ns
HI	1guj	HI + 6366 H ₂ O + 20 Na ⁺ + 18 Cl ⁻	3 × 500 ns
ConA	1vln	ConA + 35,920 H ₂ O + 138 Na ⁺ + 111 Cl ⁻	3 × 500 ns
BLG	5io5	BLG + 9234 H ₂ O + 36 Na ⁺ + 28 Cl ⁻	3 × 500 ns
STI	1avu	STI + 12,780 H ₂ O + 51 Na ⁺ + 40 Cl ⁻	3 × 500 ns
Hb	1g09	Hb + 21,474 H ₂ O + 73 Na ⁺ + 67 Cl ⁻	3 × 500 ns
AVI	1wbi	AVI + 22,347 H ₂ O + 84 Na ⁺ + 68 Cl ⁻	3 × 500 ns

3. Results and discussion

3.1. MD and NM results

The secondary structure percentages of the investigated proteins as obtained from MD calculation are reported in Table 3. The results were clustered in four groups, namely α-helix, antiparallel β-sheet, unordered, and turn structures, where also less recurring conformations as β-parallel or 3₁₀ helix structures were comprised (see SI for further details on calculations and secondary structure data from literature).

NM evaluation was performed on the equilibrium molecular structures calculated by MD. Despite the complexity of the NM analysis, which involves a large number of degrees of freedom, the energy distributions observed in the amide I region result from a limited number of normal coordinates ascribable to vibrational local modes (see SI). As example, the NM energy distributions of BSA and AVI in the amide I spectral region are reported in Fig. 4. Both spectra show bimodal distribution, with the most populated mode at lower (higher) frequency relative to β- (α-) rich proteins. To assign the observed distributions to specific secondary structures, NM calculations were performed on short peptides arranged in α or β conformation and the obtained energy distributions, normalized to the number of amino acids, are reported in the

Table 3

Secondary structure percentages computed on the representative structure of the most populated cluster as explained in SI.

Protein	BSA	CytC	HI	ConA	BLG	STI	Hb	AVI
β-Structures	–	0.02	–	0.47	0.40	0.39	–	0.54
Unordered structures	0.13	0.32	0.31	0.25	0.19	0.30	0.12	0.20
α-Structures	0.68	0.38	0.44	0.03	0.10	0.01	0.70	0.04
Turns	0.19	0.28	0.25	0.25	0.31	0.30	0.18	0.22

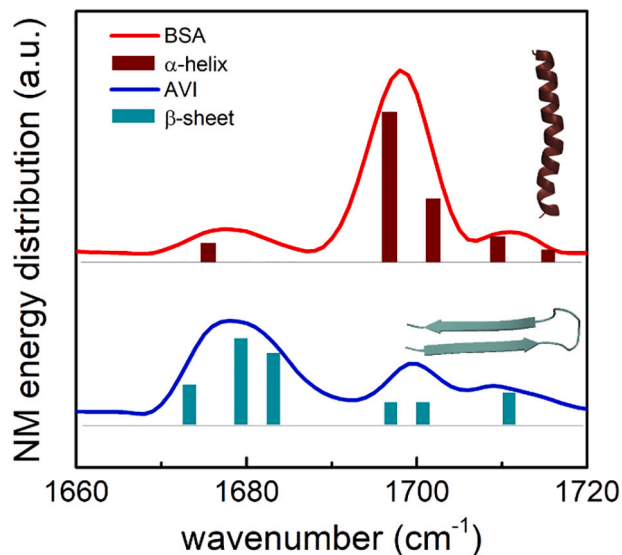


Fig. 4. NM spectra for BSA and AVI plotted together with the histograms (normalized to the number of amino acids) of the energy distribution of a 30 amino acids α -helix peptide and a 24 amino acids antiparallel β -sheet peptide.

same figure as histograms.

NM calculations for the short peptides in both a- and b-conformation were also carried out with other commonly used force fields (OPLS [49], Amber99 [50], and Gromos43a1 [51]). The corresponding energy distributions in the amide I spectral region are reported in Fig. S3, compared to those obtained with CHARMM22*. It is noteworthy that integrated intensity of the energy distribution in the 1500–1700 cm^{-1} range linearly scales with the MW confirming that the observed modes refer to local C=O (amide I) and N–H (amide II) vibrations.

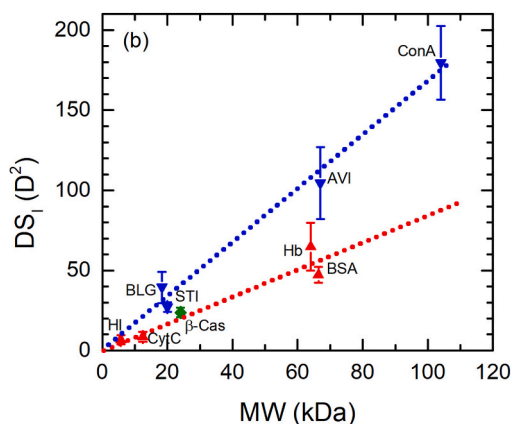
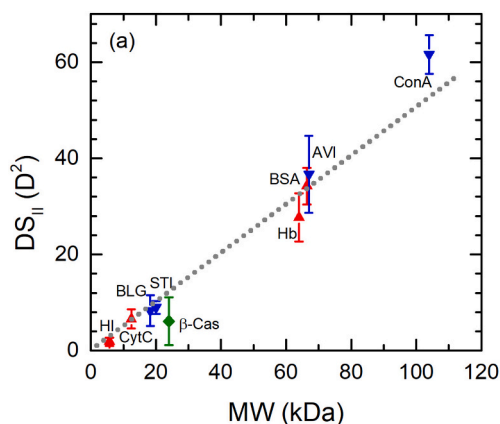


Fig. 5. Experimental dipole strength of the amide II (a) and amide I (b) absorption bands as a function of the protein MW. Linear fits to data are reported as dashed lines. Data relative to α -rich proteins are reported in red, those to β -rich proteins in blue, and that of β -Casein (unordered structure) in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. FTIR results

The experimental DS values of the amide II and amide I bands are reported in Fig. 5 as a function of the protein MW. These values were obtained according to the procedure reported in Ref. [52], where DS is defined by:

$$DS = 9.18 \cdot 10^{-3} \int \frac{\epsilon(\nu) d\nu}{\nu}$$

being $\epsilon(\nu)$ the MEC in $\text{M}^{-1} \text{cm}^{-1}$ and ν the energy in cm^{-1} . The integrals were calculated between 1600 and 1710 cm^{-1} for amide I and between 1500 and 1600 cm^{-1} for amide II.

Amide II band results from the superposition of quite a few local modes, among which the in-plane N–H bending vibration is the most involved. The DS value of this band, reported in Fig. 5a, shows a linear scaling with the MW and, therefore, with the number of oscillators. Data in Fig. 5a return an average value of $0.05 \pm 0.01 \text{ D}^2/\text{aa}$, which is a reliable estimate of the N–H dipole strength.

The linear dependence observed in the amide II DS values is irrespective from the protein secondary structures, suggesting that inhomogeneous broadening of the involved vibrational modes prevails over the dispersion potentials. Moreover, identical N–H couplings in the different secondary structures have been inferred in Ref. [53], anyhow preventing the discernment of excitonic or localized modes for this band.

Similarly, the DS values of the amide I absorption linearly scale with the MW (Fig. 5b), but with different slopes for α - and β -rich proteins, a behavior especially prompted by the values of BSA and Hb. Data in Fig. 5b provide different average DS strengths of C=O stretching in β - and α -rich proteins: $0.19 \pm 0.03 \text{ D}^2/\text{aa}$ and $0.12 \pm 0.03 \text{ D}^2/\text{aa}$, respectively. The former value obtained for β -rich proteins is almost twice than that of previous estimates ($0.10 \text{ D}^2/\text{aa}$), while that of the α proteins agrees with them within errors. Remarkably, DS of β proteins are in line with previous findings that foresaw an IR intensity enhancement of 1.5 in β structures, but much smaller than that of Ref. [18], where a TDC delocalization mechanism over three near neighbor carbonyl groups within helices was considered.

With the aim of accurately investigating the role of the specific secondary structures, we performed fits to the amide I spectra according to the energy distributions disclosed by NM calculation. Experimental spectra and normal mode distributions are reported in Fig. 6, together with the best fit with Gaussian lines centered at the frequencies of the energy distributions. The frequency mismatch between the calculated NM patterns and the experimental spectra was corrected translating the NM distribution by 45 cm^{-1} , to superimpose the most intense contributions of the NM patterns to the second derivative of the spectra. All amide I bands are well reproduced with this procedure, which requires one additional Gaussian for the tail of the amide II band. According to the NM analysis, we considered only the most intense Gaussians for the evaluation of the DS values of the α and β secondary structures.

The dipole strengths of α and β secondary structures, normalized for the number of amino acids in each structure, are reported in Fig. 7. Therein, constant values within the uncertainties are displayed for the two secondary structures, differing each other of about $0.1\text{ D}^2/\text{aa}$. This finding confirms that the protein secondary structure differently influences the intensity of the amide I band, and that the assumption of identical dipole strength for the major secondary structures is improper. A similar analysis was performed for the intrinsically disordered β -Casein, assuming a total unordered secondary structure: the resulting DS/aa value (also reported in Fig. 7) is identical within uncertainty with those obtained for α -rich proteins.

FTIR results for the amide I modes point out two occurrences: *i*) the DS/aa value of C=O in the α and in unordered structures is identical, and *ii*) the same quantity in the β structures is nearly two times greater. These outcomes give advice that carbonyl vibrations are unevenly affected by environmental condition, being the α -helix C=O groups more prone to incoherently develop a broader vibrational energy distribution. However, it is expected that these vibrational distributions

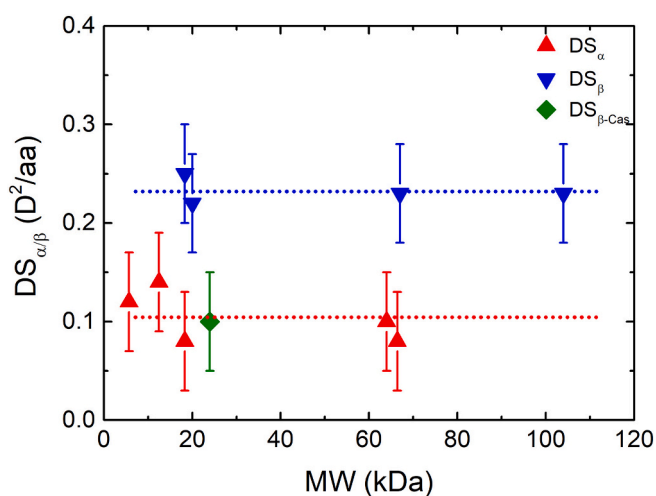


Fig. 7. Experimental dipole strength of the amide I absorption band as a function of the protein MW, normalized to the number of amino acids in each secondary structure, according to the values reported in Tables 1 and 2 ($N_{\alpha/\beta} = N_{\text{tot}} * \%_{\alpha/\beta}$). Constant fits to data are reported as dashed lines. Data relative to α -rich proteins are reported in red, those to β -rich proteins in blue, and that of β -Casein (unordered structure) in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

result not uniquely from the native molecule conformation, as also confirmed by NM calculation, but they are shaped by the interaction of each secondary structures with the surrounding environment. This is also suggested by the energy distribution provided by NM calculation,

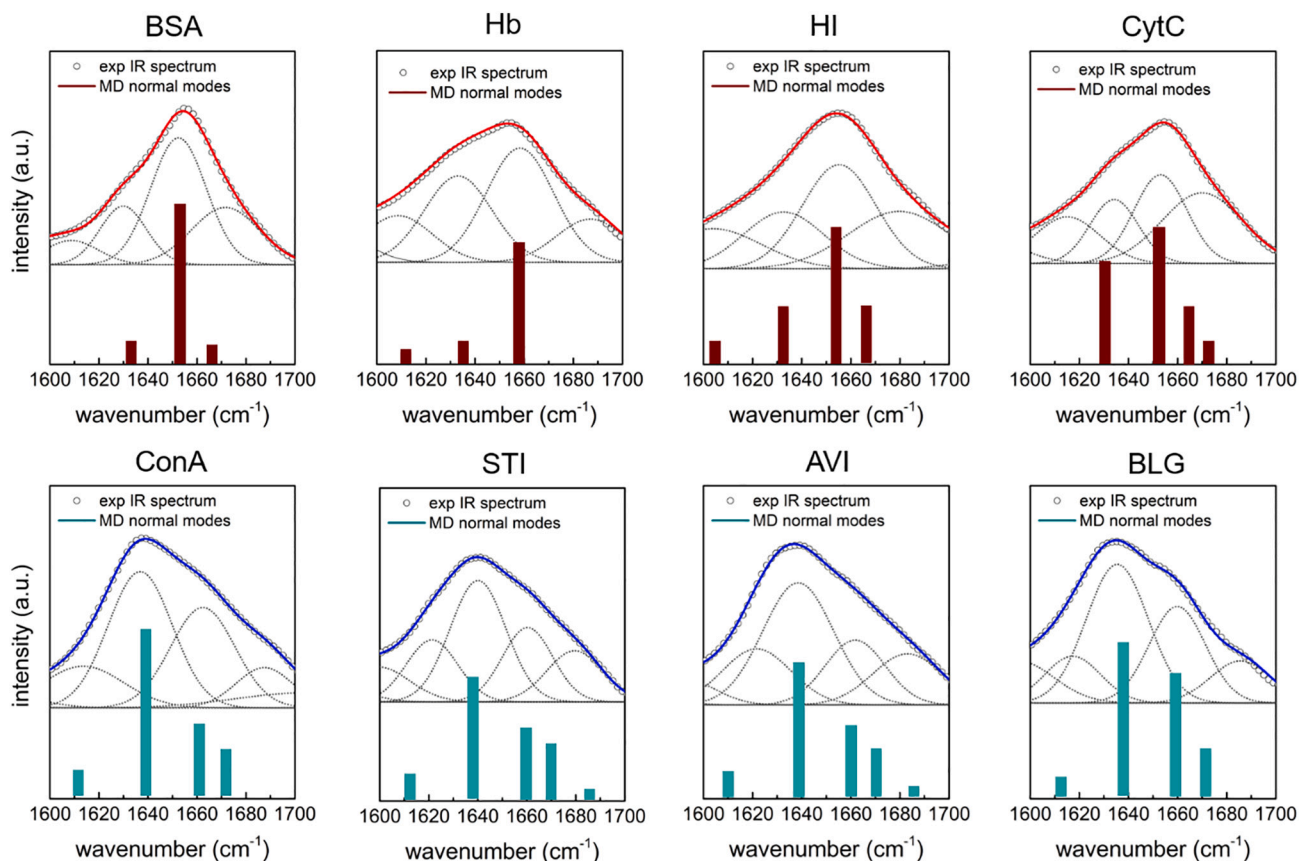


Fig. 6. Fits (continuous lines) of MEC (empty circles) of each protein obtained as the sum of Gaussian contributions (dotted lines) centered at the NM energy distributions, after shift correction.

which returned identical intensity for amide I modes of the α and β structures.

In this context, screening effect on the local dipole moment can be considered. Actually, it is known that solvent screening primarily affects the macroscopic dipole of α ordered peptides in aqueous solution [54,55], whereas the inter-strand carbonyl groups of antiparallel β structures are conformationally less exposed to the solvent. This screening can impact the DS value either through changes in the charge distribution $\left| \frac{\partial \bar{\mu}}{\partial Q} \right|$ or by the on-site shaping of the vibrational energy levels.

Beside the dipole screening, effects of the solvent might be of different nature and could variously affect the TDM; whatever the effects, they are expected to provide a significant increase of the vibrational spread Δ due to their local nature. It is remarkable that the ratio of α and β DS values meets the classical uncoupled (α) and coupled (β) oscillator ratio. This finding fosters the assumption that in the extended β structures vibrational excitons are not completely collapsed. However, it is likely that long-range excitonic order is not achievable even in the sheet structures, and that delocalization occurs only over a few carbonyl groups. In a straightforward Miyazawa model accounting for near neighbor intra-strands C=O coupling, the dipole strength of a two-sites exciton can be expressed as $\mu_0^2(1 + \cos \theta)$, being μ_0 the “bare” TDM of the carbonyl groups and θ the angle between them (see SI for further details). The two in-phase IR actives modes at $\theta \sim 60^\circ$ and $\theta \sim 120^\circ$ provide a two-sites strength of $2\mu_0^2$. This simple model accounts for the different DS/aa value here observed and advises that only short-range excitons can be created between strands.

4. Conclusions

This work was aimed to accurately determine the amplitude of the dipole strength of IR amide I and II bands for several globular proteins. This parameter provides important information on the interaction between carbonyl groups at the microscopic level, since it is involved in models for the local and extended vibrational excitation. Our findings unambiguously confirm that the amplitude of the dipole strength in the C=O groups depends on the protein secondary structure according to the hints of previous literature. Furthermore, we demonstrated that C=O oscillators of the β secondary structures are more prone to develop vibrational excitons, likely limited to few C=O sites. The phenomenological influence of the local energy disorder is addressed as a crucial parameter for the exciton collapse/delocalization. Other than the screening effect of the solvent here suggested, several factors might be accounted in the vibrational energy spread as ionic strength, aggregates formation or chain length. A comprehension of the specific influence of these effects on the dipole strength is not attainable by the present experiment; however, all of them contribute to the local energy disorder of the C=O oscillators and drive the exciton-localized vibration transition. The present results were supported by MD and NM calculations, which provided the basis for the data analysis and spectra deconvolution.

CRedit authorship contribution statement

FR: Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

NL: Data curation, Formal analysis, Methodology, Writing – review & editing.

AN: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing.

VM: Data curation, Formal analysis, Methodology, Writing – review & editing.

ADV: Formal analysis, Funding acquisition, Resources, Writing – review & editing.

AF: Investigation, Methodology, Supervision, Writing – review &

editing.

MC: Funding acquisition, Investigation, Resources, Supervision, Writing – review & editing.

Funding

This work has been funded by Ministry of Agricultural, Alimentary and Forestry Policies (MiPAAF) “PROLEGU” Project 2019-2022 and Beyond Borders Project of Tor Vergata University (CUP: E84I20000560005).

Declaration of competing interest

Authors declare no competing interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2021.06.092>.

References

- [1] T.K. Chaudhuri, S. Paul, Protein-misfolding diseases and chaperone-based therapeutic approaches, *FEBS J.* 273 (2006) 1331–1349.
- [2] F. Chiti, C.M. Dobson, Protein misfolding, functional amyloid, and human disease, *Annu. Rev. Biochem.* 75 (2006) 333–366.
- [3] V. Militello, C. Casarino, A. Emanuele, A. Giostra, F. Pullara, M. Leone, Aggregation kinetics of bovine serum albumin studied by FTIR spectroscopy and light scattering, *Biophys. Chem.* 107 (2004) 175–187.
- [4] G. Zandomenighi, M.R. Krebs, M.G. McCammon, M. Fändrich, FTIR reveals structural differences between native β -sheet proteins and amyloid fibrils, *Protein Sci.* 13 (2004) 3314–3321.
- [5] M. Carbonaro, F. Ripanti, A. Filabozzi, V. Minicozzi, F. Stellato, E. Placidi, S. Morante, A. Di Venere, E. Nicolai, P. Postorino, A. Nucara, Human insulin fibrillogenesis in the presence of epigallocatechin gallate and melatonin: Structural insights from a biophysical approach, *Int. J. Biol. Macromol.* 115 (2018) 1157–1164.
- [6] A. Nucara, M. Carbone, F. Ripanti, R. Manganiello, P. Postorino, M. Carbonaro, Achieving cytochrome c fibril/aggregate control towards micro-platelets and micro-fibers by tuning pH and protein concentration: a combined morphological and spectroscopic analysis, *Int. J. Biol. Macromol.* 138 (2019) 106–115.
- [7] J.L.R. Arrondo, A. Muga, J. Castresana, F.M. Goñi, Quantitative studies of the structure of proteins in solution by Fourier-Transform infrared spectroscopy, *Prog. Biophys. Mol. Biol.* 59 (1993) 23–56.
- [8] H. Susi, D.M. Byler, Resolution-enhanced Fourier Transform infrared spectroscopy of enzymes, *Method Enzymol.* 10 (1986) 290–311.
- [9] W.K. Surewicz, H.H. Mantsch, D. Chapman, Determination of protein secondary structure by Fourier Transform infrared spectroscopy: a critical assessment, *Biochemistry* 32 (1993) 389–394.
- [10] A.P. Fellows, M.T. Casford, P.B. Davies, Spectral analysis and deconvolution of the amide I band of proteins presenting with high-frequency noise and baseline shifts, *Appl. Spectrosc.* 74 (2020) 597–615.
- [11] G. Vedantham, H.G. Sparks, S.U. Sane, S. Tzannis, T.M. Przybycien, A holistic approach for protein secondary structure estimation from Infrared spectra in H₂O solutions, *Anal. Biochem.* 285 (2000) 33–49.
- [12] H.H.J. de Jongh, E. Goormaghtigh, J.M. Ruysschaert, The different molar absorptivities of the secondary structure types in the amide I region: an attenuated total reflection infrared study on globular proteins, *Anal. Biochem.* 242 (1996) 95–103.
- [13] Y.N. Chirgadze, B.V. Shestopalov, S.Y. Venyaminov, Intensities and other spectral parameters of infrared amide bands of polypeptides in the β - and random forms, *Biopolym. Orig. Res. Biomol.* 12 (1973) 1337–1351.
- [14] Y.N. Chirgadze, E.V. Brazhnikov, Intensities and other spectral parameters of infrared amide bands of polypeptides in the α -helical form, *Biopolym. Orig. Res. Biomol.* 13 (1974) 1701–1712.
- [15] J. Wang, Assessment of the amide-I local modes in gamma- and beta-turns of peptides, *Phys. Chem. Chem. Phys.* 11 (2009) 5310–5322.
- [16] S.Y. Venyaminov, N.N. Kalnin, Quantitative IR spectrophotometry of peptide compounds in water (H₂O) solutions. II. Amide absorption bands of polypeptides and fibrous proteins in α -, β -, and random coil conformations, *Biopolym. Orig. Res. Biomol.* 30 (1990) 1259–1271.
- [17] C.M. Baronio, A. Barth, The Amide I spectrum of proteins - optimization of transition dipole coupling parameters using Density Functional Theory calculations, *J. Phys. Chem. B* 124 (2020) 1703–1714.
- [18] M. Grechko, M.T. Zanni, Quantification of transition dipole strengths using 1D and 2D spectroscopy for the identification of molecular structures via exciton delocalization: application to α -helices, *J. Chem. Phys.* 137 (2012) 184202–184211.

- [19] E.B. Dunkelberger, M. Grechko, M.T. Zanni, Transition dipoles from 1D and 2D infrared spectroscopy help reveal the secondary structures of proteins: application to amyloids, *J. Phys. Chem. B* 119 (2015) 14065–14075.
- [20] P. Hamm, M. Zanni, *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, 2011.
- [21] A. Barth, C. Zscherp, What vibrations tell about proteins, *Q. Rev. Biophys.* 35 (2002) 369.
- [22] N. Demirdöven, C.M. Cheatum, H.S. Chung, M. Khalil, J. Knoester, A. Tokmakoff, Two-dimensional infrared spectroscopy of antiparallel β -sheet secondary structure, *J. Am. Chem. Soc.* 126 (2004) 7981–7990.
- [23] J. Wang, Anharmonic vibrational signatures of peptides - methods and applications. *Infrared spectroscopy: anharmonicity of biomolecules, crosslinking of biopolymers* 2 (2015).
- [24] F. Bogani, Two-phonon resonances and bound-states in molecular crystals. I. General theory, *J. Phys. C Solid State Phys.* 11 (1978) 1283.
- [25] F. Bogani, Two-phonon resonances and bound-states in molecular crystals. II. Absorption coefficient, *J. Phys. C Solid State Phys.* 11 (1978) 1297.
- [26] F. Caporaletti, M. Carbonaro, P. Maselli, A. Nucara, Hydrogen-deuterium exchange kinetics in β lactoglobulin (-)-epicatechin complexes studied by FTIR spectroscopy, *Int. J. Biol. Macromol.* 104 (2017) 521–526.
- [27] Arduino, Project website. <https://www.arduino.cc/>.
- [28] A. Dong, P. Huang, W.S. Caughey, Protein secondary structures in water from second-derivative amide I infrared spectra, *Biochem J.* 29 (1990) 3303–3308.
- [29] OPUS software version 7.5. <https://www.bruker.com>.
- [30] J.J. Max, C. Chapados, Isotope effects in liquid water by infrared spectroscopy. III. H₂O and D₂O spectra from 6000 to 0 cm⁻¹, *J. Chem. Phys.* 131 (2009) 184505–184518.
- [31] V.N. Uversky, *Methods in Protein Structure and Stability Analysis: Vibrational Spectroscopy*, Nova Publishers, 2007.
- [32] J. Grdadolnik, Saturation effects in FTIR spectroscopy: intensity of Amide I and Amide II bands in protein spectra, *Acta Chim. Slov.* 50 (2003) 777–788.
- [33] M. Wolpert, P. Hellwig, Infrared spectra and molar absorption coefficients of the 20 alpha amino acids in aqueous solutions in the spectral range from 1800 to 500 cm⁻¹, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 64 (2006) 987–1001.
- [34] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1–2 (2015) 19.
- [35] H. Torii, M. Tasumi, Model calculations on the amide-I infrared bands of globular proteins, *J. Chem. Phys.* 96 (1992) 3379–3387.
- [36] J.H. Choi, S. Ham, M. Cho, Inter-peptide interaction and delocalization of amide I vibrational excitons in myoglobin and flavodoxin, *J. Chem. Phys.* 117 (2002) 6821–6832.
- [37] T.A. Keiderling, Structure of condensed phase peptides: insights from vibrational circular dichroism and Raman optical activity techniques, *Chem. Rev.* 120 (2020) 3381–3419.
- [38] S. Piana, K. Lindorff-Larsen, D.E. Shaw, How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* 100 (2011) L47–L49.
- [39] K. Nakamoto, The Urey - Bradley force field: its significance and application, in: J. E. Forrette, E. Lanterman (Eds.), *Developments in Applied Spectroscopy. Developments in Applied Spectroscopy* 3, Springer, Boston, MA, 1964.
- [40] L.H. Ngai, R.H. Mann, A transferable Urey-Bradley force field and the assignments of some mixed halomethane, *J. Macromol. Spectr.* 38 (1971) 322–335.
- [41] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J.D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B* 102 (1998) 3586–3616.
- [42] M. Carbonaro, A. Di Venere, A. Filabozzi, P. Maselli, V. Minicozzi, S. Morante, E. Nicolai, A. Nucara, E. Placidi, F. Stellato, Role of dietary antioxidant (-)-epicatechin in the development of blactoglobulin fibrils, *BBA - Proteins and Proteomics* 2016 (1864) 766–772.
- [43] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* 126 (2007), 014101.
- [44] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, *J. Appl. Phys.* 52 (1981) 7182–7190.
- [45] S. Nosé, M.L. Klein, Constant pressure molecular dynamics for molecular systems, *Mol. Phys.* 50 (1983) 1055–1076.
- [46] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092.
- [47] W. Humphrey, A. Dalke, K. Schulten, VMD-Visual molecular dynamics, *J. Mol. Graph.* 14 (1996) 33–38.
- [48] R.H. Byrd, P. Lu, J. Nocedal, A limited memory algorithm for bound constrained optimization, *SIAM J. Scientific. Comput.* 16 (1995) 1190–1208.
- [49] M.J. Robertson, J. Tirado-Rives, W.L. Jorgensen, Improved peptide and protein torsional energetics with the oPLS-aA force field, *J. Chem. Theory Comput.* 11 (2015) 3499–3509.
- [50] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comp. Chem.* 21 (2000) 1049–1074.
- [51] W.F. van Gunsteren, S.R. Billeter, A.A. Eising, P.H. Hünenberger, P. Krüger, A. E. Mark, W.R.P. Scott, I.G. Tironi, *Biomolecular Simulation: the GROMOS96 Manual and User Guide*, Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996.
- [52] L. Ackels, P. Stawski, K.E. Amunson, J. Kubelka, On the temperature dependence of amide I intensities of peptides in solution, *Vib. Spectrosc.* 50 (2009) 2–9.
- [53] L.P. DeFlores, Z. Ganim, R.A. Nicodemus, A. Tokmakoff, Amide I and II 2D IR spectroscopy provides enhanced protein secondary structural sensitivity, *J. Am. Chem. Soc.* 131 (2009) 3385–3391.
- [54] F. Avbelj, J. Moult, Role of electrostatic screening in determining protein main chain conformational preferences, *Biochemistry* 34 (1995) 755–764.
- [55] D. Sengupta, R.N. Behera, J.C. Smith, G.M. Ullmann, The a helix dipole: screened out? *Structure* 13 (2005) 849–855.