

# Computing Infrared Spectra of Proteins Using the Exciton Model

Fouad S. Hussein<sup>[a]</sup>, David Robinson<sup>[a]</sup>, Neil T. Hunt<sup>[b]</sup>, Anthony W. Parker<sup>[c]</sup> and Jonathan D. Hirst<sup>\*[a]</sup>

The ability to compute from first principles the infrared spectrum of a protein in solution phase representing a biological system would provide a useful connection to atomistic models of protein structure and dynamics. Indeed, such calculations are a vital complement to 2DIR experimental measurements, allowing the observed signals to be interpreted in terms of detailed structural and dynamical information. In this article, we have studied nine structurally and spectroscopically well-characterized proteins, representing a range of structural types. We have simulated the equilibrium conformational dynamics in an explicit point charge water model. Using the resulting trajectories based on MD simulations, we have computed the one and two dimensional infrared spectra in the Amide I region, using an exciton approach,

in which a local mode basis of carbonyl stretches is considered. The role of solvent in shifting the Amide I band (by 30 to 50  $\text{cm}^{-1}$ ) is clearly evident. Similarly, the conformational dynamics contribute to the broadening of peaks in the spectrum. The inhomogeneous broadening in both the 1D and 2D spectra reflects the significant conformational diversity observed in the simulations. Through the computed 2D cross-peak spectra, we show how different pulse schemes can provide additional information on the coupled vibrations. © 2016 The Authors. Journal of Computational Chemistry Published by Wiley Periodicals, Inc.

DOI: 10.1002/jcc.24674

## Introduction

Understanding the three-dimensional structure of a protein is often a challenging task but is an undertaking that can yield deep insights into biological functions, ranging from membrane signaling to catalysis to charge transfer as well as dynamic scaffolding, mechanical and electrical transduction. Approaches such as X-ray crystallography and nuclear magnetic resonance can provide atomistic detail, while optical spectroscopy in the ultra-violet and infrared (IR) regions can provide useful qualitative information. Efforts to understand and interpret the characteristic spectroscopic features of proteins have been ongoing for many decades. In the IR, the Amide I region lies between 1600 and 1700  $\text{cm}^{-1}$ ,<sup>[1–3]</sup> and is sensitive to the backbone conformation of a protein. This region has been extensively used to probe protein structure and dynamics, as it can provide useful information with respect to protein folding, misfolding, and unfolding.<sup>[3–6]</sup> However, the spectra often show convoluted and overlapping bands that can be challenging to decipher. There are distinctive spectral characteristics arising from  $\alpha$ -helices,  $\beta$ -sheets, and random coil structures.<sup>[7–12]</sup>  $\alpha$ -helices exhibit a band between 1650 and 1658  $\text{cm}^{-1}$ .<sup>[3]</sup> Bands near 1663  $\text{cm}^{-1}$  have been associated with  $3_{10}$  helices,<sup>[3,7,8]</sup> while  $\beta$ -sheets exhibit bands between 1620 and 1640  $\text{cm}^{-1}$  as well as 1690 and 1695  $\text{cm}^{-1}$ .<sup>[3–7]</sup>

The characteristic vibrations of polypeptides in general consist of nine types: Amide A, B, and I–VII modes. The Amide I and II are interesting from a structural perspective as they give rise to two broad bands associated with the protein backbone. The former is the primary focus of this article. The Amide I

mode is characterized by the C=O stretch which accounts for about 80% of the vibration, and the wagging and bending of the N–H bond which accounts for the remaining 20%.

Torii and Tasumi<sup>[13]</sup> used *ab initio* calculations to investigate the A (singly degenerate) and E (doubly degenerate) components of the Amide I bands in the Raman and IR spectra of peptides. For helical conformations, the A component of the Raman band is intense and corresponds to the carbonyl groups vibrating in-phase; the E component of the IR band is less intense whereby an out-of-phase vibrational combination leads to a net transition dipole moment perpendicular to the helix axis. For  $\beta$ -sheets, the splitting of the characteristic intense peak at  $\sim 1620 \text{ cm}^{-1}$  and the weaker peak at  $\sim 1690 \text{ cm}^{-1}$  is directly proportional to the number of strands (up to a certain amount) in the sheet.<sup>[14]</sup> For larger  $\beta$ -sheets, the absorption becomes independent of the size of the sheet.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

[a] F. S. Hussein, D. Robinson, J. D. Hirst

School of Chemistry, University of Nottingham, Nottingham, NG7 2RD, United Kingdom  
E-mail: jonathan.hirst@nottingham.ac.uk

[b] N. T. Hunt

Department of Physics, University of Strathclyde, SUPA, 107 Rottenrow East, Glasgow G4 0NG, Scotland, United Kingdom

[c] A. W. Parker

STFC Rutherford Appleton Laboratory, Central Laser Facility, Harwell Campus, Didcot OX11 0QX, United Kingdom  
Contract grant sponsor: Saudi Cultural Bureau (to F.S.H.)

© 2016 The Authors. Journal of Computational Chemistry Published by Wiley Periodicals, Inc.

In the past 20 years or so, sophisticated experimental techniques have been developed to allow collection of IR spectra in two dimensions using both time (photon echo<sup>[15]</sup>) and frequency domain (double resonance<sup>[16]</sup>), methodologies.<sup>[16–23]</sup> Irrespective of the experimental approach, a 2D-IR signal arises from a sequence of three laser-sample interactions and the resulting spectrum is a correlation map of excitation and detection frequencies. This leads to the spreading of the molecular response over a second frequency axis, allowing resolution of features that are obscured by overlapping peaks in a traditional IR spectrum.

In a 2D-IR spectrum, diagonal peaks represent signals featuring excitation (pump) and detection (probe) at the same frequency and these are analogous to the features observed in a 1D-IR spectrum. Additional information arises from the 2D line-shape of these features, which reflects the temporal evolution of the local environment of a given oscillator. Off-diagonal peaks arise when the excitation and detection frequencies differ and these provide insights into vibrational couplings, energy transfer or chemical exchange processes. The shape of these off-diagonal peaks can also be influenced by coupling between vibrational modes.

2D-IR has been increasingly applied to protein samples and a wide range of applications have been reported. These have included the spectroscopy and dynamics of disordered polypeptides,<sup>[24–26]</sup> picosecond protein conformational dynamics,<sup>[27–30]</sup> amyloid fibril formation<sup>[31–34]</sup> and the structure of transmembrane proteins.<sup>[35–37]</sup> These have been extensively reviewed elsewhere.<sup>[38–41]</sup>

Fully *ab initio* or density functional theory (DFT) calculations of the vibrational frequencies of large polypeptides are currently prohibitively demanding of computational resources. Thus, more approximate approaches are adopted. Krimm and Bandekar<sup>[2,3]</sup> recognized that the nature of the Amide I band is influenced by the interactions of carbonyl vibrations via electrostatics and they constructed the Transition Dipole Coupling (TDC) model, which captures an essential part of the inter-peptide couplings. The model has formed the basis for interpreting the Amide I bands of polypeptides in linear absorption spectra<sup>[42]</sup> and has been extended to the analysis of 2D-IR spectra. Hamm and Woutersen<sup>[43]</sup> suggested a Transition Charge Coupling model that included higher order multipole contributions, and further improved on the TDC results. Although the model was, in general, consistent with DFT studies, it could not describe through bond coupling.

The influence of the molecular environment on individual local modes has attracted significant attention.<sup>[44–54]</sup> Ham and Cho<sup>[47]</sup> provided a framework for considering the influence of the electrostatic environment, with the development of coupling and frequency maps derived from *ab initio* calculations on model peptides such as *N*-methylacetamide (NMA) and dipeptides. Both the coupling and frequency shift maps are dependent on the main chain dihedral angle of the dipeptide. These frequency maps can include the effects of water surrounding the chromophores as well as other components such as ions or lipids. Many of these maps have been developed over the past decade, some derived from *ab initio*

calculations<sup>[50,55–59]</sup> and some, such as Skinner's,<sup>[60]</sup> derived empirically. These maps have been widely adopted to calculate short-range interactions, and are used in conjunction with the TDC model for the long-range interactions.

Early<sup>[20,61]</sup> calculations of Amide I bands used models based on simple geometric properties relating to the nature of hydrogen bonding. Karjalainen et al.<sup>[54]</sup> studied a set of 44 proteins, calculating Amide I spectra by empirically optimizing parameters in several terms accounting for the effects of solvent, the local environment, and inter-peptide hydrogen bonding. Their work showed how the shift in frequency is strongly dependent on the number of hydrogen bonds to the amide oxygen atom or the amide NH group. This empirical approach contrasts with the more sophisticated models based on different electrostatic properties such as the electric field, electric field gradient, and the electrostatic potential.

Ganim and Tokmakoff<sup>[62]</sup> examined the influence of conformational fluctuations on computed Amide I bands in 1D and 2D IR spectra for three small proteins using molecular dynamics (MD) simulations. The fluctuations of both the solvent and solute influenced the calculated IR lineshapes. They reported that the computed lineshapes were broader than the experiment. Choi et al.<sup>[63]</sup> presented computational (semi-empirical and MD) simulations and theoretically predicted the IR, 2D IR, electronic and vibrational dichroism spectra of ubiquitin. In their simulations, the backbone atoms were constrained to keep the conformations close to those obtained from semi-empirical geometry optimizations. They found that hydration had a significant effect on the computed IR spectra, contributing to the computed red shift of the Amide I bond of different structural components. Recently, Jansen and coworkers have benchmarked several approaches to computing the Amide I band and 2D IR of proteins from MD simulations.<sup>[64,65]</sup> Up to four proteins were studied using several combinations of force fields for the MD simulations, electrostatic mappings and couplings. Skinner's empirical frequency map<sup>[60]</sup> with the TDC model was reported to do well in conjunction with the OPLS-AA force field. However, it was concluded that there is still considerable scope for understanding and improving modeling approaches. Our study provides some additional complementary insight into the current state of the art.

We develop further insight into the relationship between protein conformation and spectra. Using the electrostatic potential to compute the coupling and frequency maps, we consider a range of well-studied, typical proteins and investigate how the environment can affect protein IR spectra in the Amide I region. We model the shift of spectral peaks resulting from the fluctuating surrounding environment with the aid of MD simulations. To investigate the effect of solvent and conformational dynamics on the Amide I spectra, we use a test-bed of eight globular proteins<sup>[42]</sup> plus ubiquitin.<sup>[63]</sup> The former were previously used<sup>[42]</sup> to study the effect that irregularities and distortions in structural components had on the Amide I bands. Following Grechko and Zanni,<sup>[66]</sup> we investigate the influence of various structural and dynamical aspects on the location and intensities of bands of 2D signals.

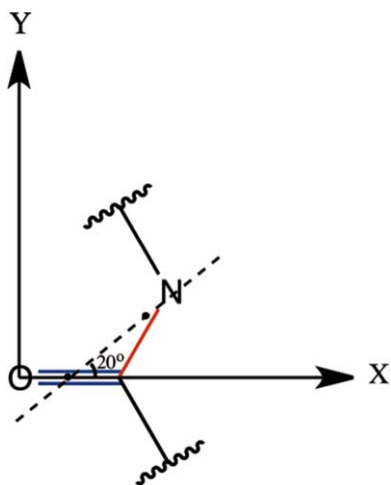


Figure 1. The location and orientation of the transition dipole. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

## Methodology

### Exciton Hamiltonian

Exciton theory<sup>[67]</sup> provides a framework for considering a large polymeric system. The vibrational exciton one-quantum Hamiltonian is constructed based on a system of coupled local modes:

$$\hat{H} = H_0 + F - E_0 \quad (1)$$

in which  $\hat{H}$  is the Hamiltonian,  $H_0$  is the Hamiltonian of  $N$  non-interacting peptide groups,  $F$  is the inter-peptide potential, and  $E_0$  is the ground state energy. Hence, the Hamiltonian matrix consists of three types of element: the diagonal elements which correspond to the harmonic (central) frequency, the off-diagonal nearest neighbor coupling (NNC) constants, and the other off-diagonal elements which describe the through-space interaction between local mode vibrations. The TDC approximation<sup>[2,3]</sup> calculates the latter elements as:

$$f_{ij} = \frac{0.1}{\epsilon} \times \frac{\vec{\mu}_i \cdot \vec{\mu}_j - 3(\vec{\mu}_i \cdot \vec{\eta}_{ij})(\vec{\mu}_j \cdot \vec{\eta}_{ij})}{r_{ij}^3} \quad (2)$$

where  $f_{ij}$  is the TDC,  $\epsilon$  is the dielectric constant,  $\vec{\mu}_i$  is the transition dipole moment for the Amide I mode located on peptide  $i$ ,  $\vec{\mu}_j$  is the transition dipole moment for peptide  $j$ ,  $r_{ij}$  is the separation of the dipoles between peptides  $i$  and  $j$ ,  $\eta_{ij}$  is the vector defining the separation between the  $i^{\text{th}}$  and  $j^{\text{th}}$  peptide.

Following Torii and Tasumi,<sup>[42]</sup> the TDC was computed using a transition dipole (Fig. 1) placed 0.868 Å away from the amide carbonyl bond, and oriented 20° toward the amide nitrogen along the OCN plane. Its magnitude was 3.7 D Å<sup>-1</sup> amu<sup>-1/2</sup>.

The nearest neighbor off-diagonal Hamiltonian matrix elements are assigned from a NNC look-up table or coupling map<sup>[57]</sup> that consists of force constants calculated *ab initio* for all combinations of main-chain dihedral angles (in increments of 30°) for a di-peptide.

Table 1. The backbone partial charges from the CHARMM force field,<sup>[69]</sup> assigned to each of the atoms in a peptide unit.

Atom	Partial charge (e)	$l_j$ (e)
C	0.51	0.00160
O	-0.51	-0.00554
N	-0.47	0.00479
H	0.31	-0.00086

Their respective linear expansion coefficients from the Cho<sup>[47]</sup> map were used to calculate the shift in central frequency.

We now turn to the calculation (using a modified version of the dichrocalc software<sup>[68]</sup>) of the diagonal elements of the Hamiltonian and the change in frequency for each local mode due to the surrounding electrostatic environment. The electrostatic potential at site  $i$  is computed from a set of atom-centered partial charges<sup>[47]</sup>:

$$\phi_i = \sum_{j=1}^N \frac{c_j}{4\pi\epsilon_0 r_{i,j}} \quad (3)$$

where  $j$  is the index that runs over all  $N$  partial atomic charges,  $c_j$ , in the system. The atoms in the peptide group where the potential is calculated are excluded from the summation. The atomic partial charges for backbone atoms and side chains groups were taken from CHARMM36 force field.<sup>[69]</sup> Explicit solvent (e.g., water) and hetero-atoms are thus readily (and have been) included in the calculations. The interplay between the force field used to sample conformational dynamics and models used to construct the exciton Hamiltonian is complex.<sup>[64,65]</sup> Previous studies have considered CHARMM22<sup>[70,71]</sup> amongst other force fields. In our work, we use the CHARMM36 force field, where there is evidence<sup>[69,72]</sup> that changes in the internal parameters describing the peptide backbone give a better representation of the structure and dynamics than CHARMM22 as assessed through validation against various experimental (in many cases NMR) observables.

The electrostatic potential is used in conjunction with a combination of so called linear expansion coefficients to give the shifted frequency:

$$\omega_k = \omega_0 + \sum_{j=1}^4 l_j \phi_{kj} \quad (4)$$

where  $\omega_k$  is the shifted frequency of peptide  $k$ ,  $\omega_0$  is the central frequency (discussed later). Index  $j$  runs across the four atoms at which there are partial charges (Table 1) in each peptide  $k$ . The linear expansion coefficients  $l_j$  (Table 1) are derived from the following equation:

$$l_j = \frac{g^l}{4\pi c M_l^2 (\omega_l^0)^3} \left( \frac{\partial c_l}{\partial Q_l} \right)_0^{\text{eff}} \quad (5)$$

where  $g^l$  is the cubic anharmonic coefficient for the Amide I mode,  $c$  is the speed of light,  $M_l$  is the reduced mass,  $\omega_l^0$  is the angular frequency, and  $\left( \frac{\partial c_l}{\partial Q_l} \right)_0^{\text{eff}}$  is the effective transition charge in units of e Å<sup>-1</sup>.

Table 2. Proteins studied with their PDB codes.

Protein	PDB Code	Class	% $\beta$ -strands	% $\beta$ -turns	% $\alpha$ -helix	$3_{10}$ - helices	% Coil	No. solvent molecules
$\alpha$ -Lactalbumin	1ALC	$\alpha + \beta$	7	30	30	14	19	8053
Carbonmonoxy-Myoglobin	1MBC	$\alpha$	0	7	74	4	15	6860
Concanavalin A	3CNA	$\beta$	42	44	0	0	14	11227
Egg White Lysozyme	2LYM	$\alpha + \beta$	6	40	30	10	14	8412
$\beta$ -Trypsinogen	2PTN	$\alpha + \beta$	32	44	8	3	13	7641
Carboxypeptidase A	5CPA	$\alpha + \beta$	16	30	35	2	17	10862
Ribonuclease A	7RSA	$\alpha + \beta$	27	24	18	3	28	6379
Ubiquitin	1UBQ	$\alpha + \beta$	32	28	16	8	16	5978
Flavodoxin	5NLL	$\alpha + \beta$	21	19	37	4	19	6782

Their secondary structure content computed using PROMOTIF.<sup>[92]</sup> The number of water molecules used in the MD simulations is provided.

The central frequency is usually chosen between 1650 and 1710  $\text{cm}^{-1}$ .<sup>[16,73,74]</sup> We adopted a value of 1680  $\text{cm}^{-1}$ , which gives computed spectra consistent with the range observed experimentally for the Amide I region. The difference between this value and the gas phase value for NMA of 1717  $\text{cm}^{-1}$  suggests that the electrostatic effect (as modeled here) does not fully account for the solvent-induced shift. For proline residues, which do not have an amide hydrogen atom the frequency shift was not explicitly calculated; instead we simply used a fixed frequency of 1653  $\text{cm}^{-1}$ .<sup>[75]</sup> Side chains that are known to absorb in the Amide I region such as those present in glutamine and asparagine have not been considered as chromophore groups; they are treated as side chains contributing to the electrostatic potential instead. The 1D absorption line spectra were convoluted with a Gaussian full width at half maximum bandwidth of 4  $\text{cm}^{-1}$ . This convolution accounts for broadening due to mechanisms not captured explicitly by the MD simulation. We illustrate how isotopic labeling of the carbonyl groups in different secondary structure elements in ubiquitin could be used to deconvolute the distinct contributions of helix and sheet to the 2D signal. The shift due to  $^{13}\text{C}^{18}\text{O}$  isotope labeling lies between 60 and 75  $\text{cm}^{-1}$ .<sup>[50,76–78]</sup> We adopted a value of 65  $\text{cm}^{-1}$  for residues belonging to secondary structure types of interest. To calculate the 2D spectra, the two-quantum Hamiltonian is constructed from the one-quantum Hamiltonian matrix elements as follows<sup>[62]</sup>:

$$H_{m,n}^{II} = H_{m,m} + H_{n,n} - \delta_{m,n} \Delta \quad (6)$$

$$H_{mm,nk}^{II} = \sqrt{2}(H_{m,k} \delta_{m,n} + H_{m,n} \delta_{m,k}) \quad (7)$$

$mm \neq nk, m \neq k$

$$H_{mn,nk}^{II} = H_{m,k}(\delta_{m,n} + \delta_{n,k}) \quad (8)$$

$m \neq k$

where  $\delta_{m,n}$  is Kronecker's delta,  $H^{II}$  is the two-quantum Hamiltonian,  $H$  is the one-quantum Hamiltonian with site indices  $m$ ,  $n$ , and  $k$ .  $\Delta$  is the difference between the fundamental and overtone absorption frequencies, also known as the anharmonicity. The Hamiltonian operator for singly and doubly excited states can be expressed as<sup>[79]</sup>:

$$\hat{H} = \sum_{n=1}^N \varepsilon_n |n\rangle \langle n| + \sum_{m,n=1}^N J_{mn} |m\rangle \langle n| + \sum_{m,n=1}^N (\varepsilon_m + \varepsilon_n - \Delta \delta_{mn}) |mn\rangle \langle mn| + \sum_{m,n=1}^N \sum_{j,k=1}^N J_{mn,jk} |mn\rangle \langle jk| \quad (9)$$

$m, n \neq j, k$

where  $J$  is the coupling constant between the singly excited  $|m\rangle, |n\rangle$  states or doubly excited  $|mn\rangle$  and  $|jk\rangle$  states.  $\varepsilon$  is the site energy of the relevant state.  $N$  is the number of sites. The local transition dipoles  $\vec{\mu}_i$ , are likewise constructed from the one-quantum transition dipole moments to produce two-quantum transition dipole moments using the following expression:

$$\vec{\mu}_{m,n} = \sqrt{2} \vec{\mu}_m \delta_{m,n} + \vec{\mu}_n (1 - \delta_{m,n}) \quad (10)$$

with  $m$  and  $n$  again being the site indices. The two-exciton Hamiltonian matrix is diagonalized to produce a set of  $\frac{N^2+N}{2}$  energies that are used to compute the non-linear response. For a protein of  $N$  oscillators, there are  $N^2$  two-quantum states and the number of interactions (or matrix elements in the Hamiltonian) grows as  $N^4$ , which makes the calculations significantly more demanding than for 1D-IR.

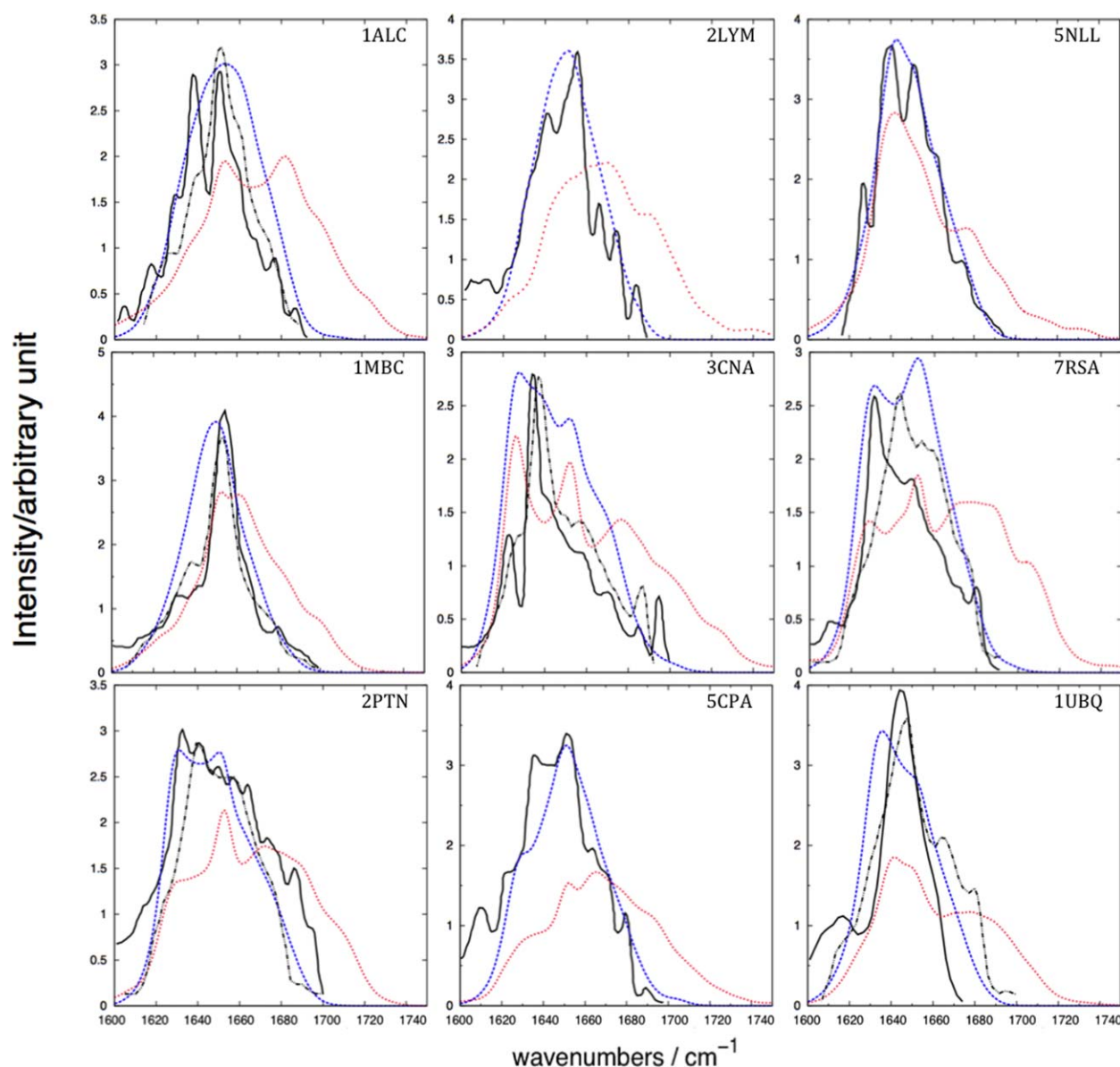
The third order non-linear polarization  $P^{(3)}$  is a convolution<sup>[80–84]</sup> of the third order non-linear response functions  $R^{(3)}$  and the three electric fields  $E_n$ :

$$P_{(t)}^{(3)} = \int_0^\infty dt_3 \int_0^\infty dt_2 \int_0^\infty dt_1 E_3(t-t_3) E_2(t-t_3-t_2) E_1(t-t_3-t_2-t_1) R^{(3)}(t_3, t_2, t_1) \quad (11)$$

where  $t_n$  refers to the time intervals between laser pulses.  $R^{(3)}$  describes the macroscopic behavior of the system under the effect of the optical fields between the time intervals.

In 2D photon echo experiments, the diagonal peaks appear as positive signals while they appear as negative bleaches in double resonance experiments. The off-diagonal contributions to the 2D signal; however are both positive and negative. We computed the two-quantum Hamiltonian using a modified version of the Zanni and Hamm code<sup>[80]</sup> reading into the *peptide.c* code the one-quantum exciton Hamiltonians constructed for each snapshot, and uses a fixed anharmonicity of





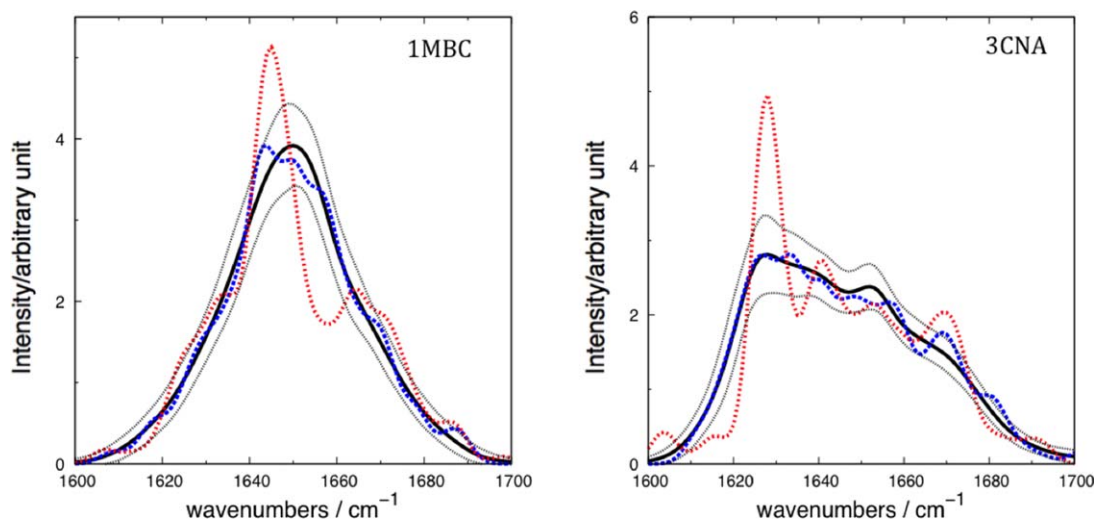
**Figure 2.** Amide I IR spectra of the nine proteins. The solid black line represents the experimental spectra from various sources cited by Torii and Tasumi, *J. Chem. Phys.*, 1992, 96, 3379, reproduced by permission,<sup>[42]</sup> who reported that the spectra were “weakly deconvoluted”. The dashed black represents the experimental (recorded for 3% protein solution in H<sub>2</sub>O) transmission IR spectra taken from Karjalainen et al., *J. Phys. Chem. B*, 2012, 116, 4831, reproduced by permission.<sup>[54]</sup> The dashed blue represents the average computed spectra including solvent, and the dotted red line represents the average computed spectra excluding solvent. See Table 2 for the names of the proteins. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

16 cm<sup>-1</sup>. The Hamiltonian is diagonalized and the corresponding unitary transformation is used to transform the transition dipole matrix. The dipole approximation is used, whereby cross-excitations are not allowed. The 2D signal is evaluated as the sum of the rephasing and nonrephasing components. The computed spectra shown in this article are purely in the frequency domain, and the diagonal and off-diagonal contributions to the 2D signal shown here are positive and negative signals respectively (as in photon echo experiments). Thus the positive signal represents the ground state depletion (bleach) and stimulated emission ( $v=0 \rightarrow 1$ ), while the negative signal corresponds to excited state emission ( $v=1 \rightarrow 2$ ). Polarization conditions have been examined previously by Hochstrasser.<sup>[82]</sup> The peak intensities are

collected by ensemble averaging the lab frame dipole moment components to account for the orientation of residues with respect to the laser polarization. Our 2D signal is computed using the ZZZZ polarization condition:

$$\langle Z_\alpha Z_\beta Z_\gamma Z_\delta \rangle = \frac{1}{15} (\langle \cos \theta_{\alpha\beta} \cos \theta_{\gamma\delta} \rangle + \langle \cos \theta_{\alpha\gamma} \cos \theta_{\beta\delta} \rangle + \langle \cos \theta_{\alpha\delta} \cos \theta_{\beta\gamma} \rangle) \quad (12)$$

where  $\theta_{mn}$  is the angle between transition dipoles  $m$  and  $n$ . In a later section, we show the enhancement of cross-peaks by subtracting two computed spectra: one using the ZZZZ condition and the other using the ZXXZ pulse condition.<sup>[63,80–83]</sup> Similar to eq. (12), the latter pulse condition can be expressed as:



**Figure 3.** Computed spectra for a)  $\alpha$ -helical protein carbonmonoxy-myoglobin (1MBC) and b)  $\beta$ -sheet protein concanavalin A (3CNA), for the snapshots with spectra most similar (dashed blue), and least similar (dotted red) to the computed average spectra (solid black). The thin dotted lines (upper and lower) represent the mean average  $\pm$  standard deviation, respectively. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

$$\begin{aligned} \langle Z_{\alpha} X_{\beta} X_{\gamma} Z_{\delta} \rangle = & -\frac{1}{30} \left( \langle \cos \theta_{\alpha\beta} \cos \theta_{\gamma\delta} \rangle - 4 \langle \cos \theta_{\alpha\gamma} \cos \theta_{\beta\delta} \rangle \right. \\ & \left. + \langle \cos \theta_{\alpha\delta} \cos \theta_{\beta\gamma} \rangle \right) \end{aligned} \quad (13)$$

The enhanced signal is computed using:

$$\text{Sig}_{\text{en}} = \langle \text{ZZZZ} \rangle - 3 \langle \text{ZZXZ} \rangle \quad (14)$$

### Molecular dynamics simulations

Using the NAMD 2.9 molecular dynamics package,<sup>[84]</sup> we performed MD simulations on ubiquitin and the eight proteins studied by Torii and Tasumi.<sup>[42]</sup> The structures were taken from the Protein Data Bank, and the N- and C-termini were capped to give  $(\text{NH}_3^+ - \text{C}_{\alpha}\text{H}_2)$  and  $(-\text{CH}_2 - \text{CO}_2^-)$ , respectively. For the cases of  $\alpha$ -lactalbumin, concanavalin A and myoglobin, the *apo* forms of the proteins were used. Neutralization was achieved by adding 9  $\text{Na}^+$  ions for  $\alpha$ -lactalbumin, 9  $\text{K}^+$  for concanavalin A, 16  $\text{K}^+$  for flavodoxin, 2  $\text{Cl}^-$  for myoglobin, 5  $\text{Cl}^-$  for ribonuclease A, 6  $\text{Cl}^-$  for trypsin, and 8  $\text{Cl}^-$  for lysozyme. The simulations included explicit water to model the influence of conformational dynamics on line broadening, and to investigate the effect of solvation on the Amide I band. Each protein was solvated in a hexagonal prism of TIP3P water molecules,<sup>[85]</sup> and periodic boundary conditions were applied. To account for long-range interactions, the Particle Mesh Ewald method<sup>[86]</sup> was used, and the Lennard-Jones cutoff was 12 Å. Energy minimization was performed for each protein for 30,000 cycles.

Thereafter an equilibration process with an integration time-step of 2 fs ran for 0.5 ns, during which all covalent bonds involving hydrogen were constrained using the SHAKE algorithm.<sup>[87]</sup> Experiments are usually performed in deuterated water. The water molecules in our simulations have rigid O—H bonds. Thus, the TIP3P model, captures hydrogen bonding and electrostatic effects but, neither the influence of the

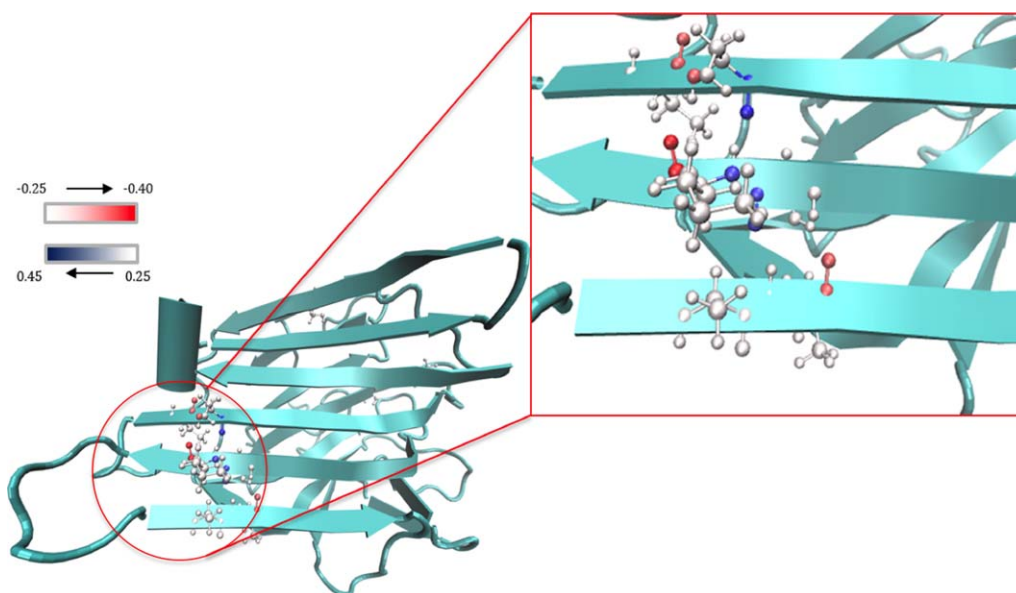
vibrations of water nor the effect of the deuteration on the conformational dynamics of the protein are considered.

Production dynamics were performed for a period of 2 ns in the NPT ensemble using Langevin dynamics and a damping coefficient of  $5 \text{ ps}^{-1}$ . The Nosé–Hoover<sup>[88–90]</sup> and Langevin piston<sup>[91]</sup> periods were set to 100 fs and their time-decay period was set to 50 fs to keep the temperature constant at 300 K, while maintaining pressure at 1 atm. Snapshots were sampled uniformly every picosecond. Trajectory files with and without solvent were saved separately to investigate the effect of solvent on the computed spectra. The 2 ns trajectories are short, but the main purpose is to provide a sample of configurations close to the experimental structures. Our unconstrained simulations will potentially explore a broader and more physical range of equilibrium conformations than the constrained simulations of Choi et al.<sup>[63]</sup>

### Results and Discussion

Experimental transmission IR spectra were taken from the literature<sup>[42,63]</sup> rescaled such that the highest intensity peaks match the computed spectra, and plotted against the computed spectra of the nine proteins (Fig. 2). The experimental conditions used for recording the spectra cited by Torii and Tasumi<sup>[42]</sup> were as follows:  $\alpha$ -lactalbumin and lysozyme were recorded using a 3.5% protein solution in  $\text{D}_2\text{O}$ ; myoglobin and trypsin using a 5% protein solution in  $\text{H}_2\text{O}$ , ribonuclease A using a 10% protein solution in  $\text{H}_2\text{O}$ , while spectra for carboxypeptidase A, concanavalin A, and flavodoxin spectra were all recorded using a 5% protein solution in  $\text{D}_2\text{O}$ . We present our computed spectra without any post-processing to enhance the fine structure. The overall band shape of each of the computed spectra for solvated proteins agrees with the experiment (Fig. 2).

The spectra computed neglecting the solvent exhibit an Amide I band that is broader than the experiment, extending beyond  $1700 \text{ cm}^{-1}$  to around  $1750 \text{ cm}^{-1}$ . We believe that neglecting solvent in the calculations means that the surface



**Figure 4.** Concanavalin A in one of the snapshots of interest. The strongly coupled residues are depicted with an atomic representation. The color bar scale represents the eigenvector coefficients: red indicates a negative coefficient, blue a positive one and the atoms are colored accordingly. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

residues have an environment that is artefactually more different from the buried residues compared to the situation for a solvated system. To investigate the influence of conformational diversity on the computed spectra, we calculated the standard deviation of the computed intensity at each wavenumber over all snapshots (Fig. 3). We also identified the individual snapshots giving rise to the computed spectra that were least and most similar to the mean computed spectra over the 2 ns simulation period (Fig. 3).

The spectra from individual snapshots that are most dissimilar to the ensemble show some of the most intense features. By examining these conformations, we can characterize the extent of (de)localization of the vibrations associated with the most intense features. The squares of the eigenvector coefficients reflect the contributions of the local modes to the transition.<sup>[93,94]</sup> For the most dissimilar snapshot of concanavalin A, only two coefficients had a squared magnitude greater than 0.25, that is, none of the transitions was particularly localized. Of particular note was a pattern observed near the end of the 2 ns trajectory. The intense peak was located between 1622 and 1626  $\text{cm}^{-1}$ . In this spectral region of the simulation, the vibration was delocalized across different residues in parallel strands. This is consistent with the strength of the through-space coupling constants between these residues. Figure 4 shows the location of the vibration in the context of the protein structure. The coupling between the residues fluctuates over the simulation, but certain conformations (Fig. 4) exhibited strong inter-strand coupling between residues perpendicular to the strand orientation of the sheet, which is consistent with the findings of Woys et al.<sup>[95]</sup>

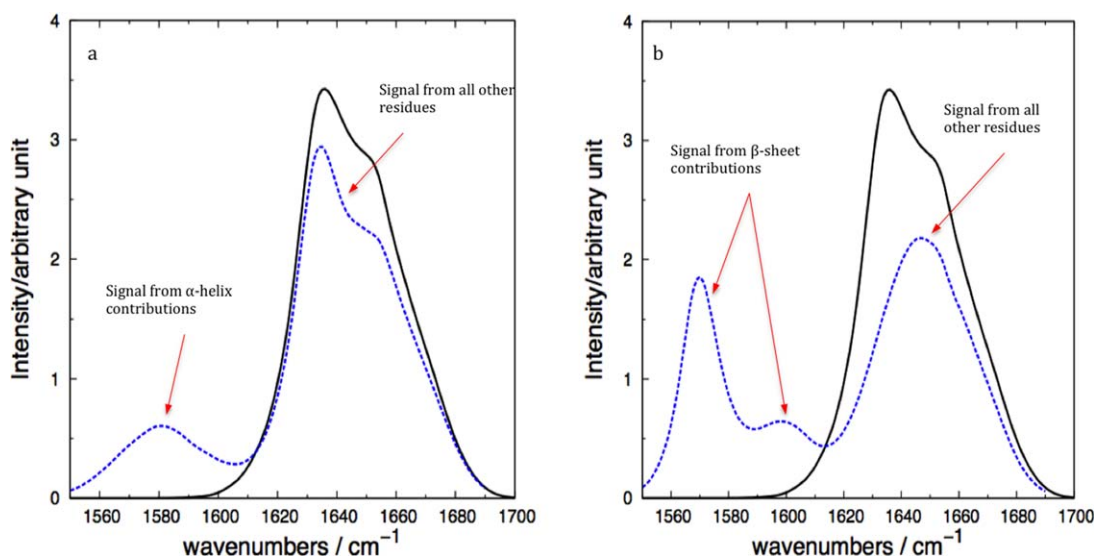
Experimentally,<sup>[96,97]</sup> it is possible using expressed protein ligation and native chemical ligation to isotope label distinct regions in proteins, for example, specific elements of secondary structure. The computational analogue is readily performed and

can help us to understand the various contributions to the spectra and deconvolute the overlapping bands of the 1D signal. We illustrate this for the simulations of solvated ubiquitin. The result of isotope labeling of either residues in  $\alpha$ -helices or in  $\beta$ -sheets (Fig. 5) shows a shift in peaks accordingly. The residues belonging to  $\alpha$ -helices show a contribution to the 1D signal in the form of a single broad peak at 1580  $\text{cm}^{-1}$ , while the contribution from  $\beta$ -sheet residues shows two peaks, one more intense than the other. Splitting the signal by isotope labeling has an effect on the overall band maxima, which will be discussed later.

Karjalainen et al.<sup>[54]</sup> assumed the shift in intrinsic frequency is related to the presence or absence of a hydrogen bond between a carbonyl oxygen of one amide group and an amide hydrogen of another. We investigated the relationship between shift in intrinsic frequency and the presence of hydrogen bonds to solvent. The relationship is evident, but appears to be a weak one (Fig. 6). For example, in the case of the  $\alpha$ -helical carbonmonoxy-myoglobin, the number of hydrogen bonds to solvent molecules shows a modest influence on the extent of the negative shift from the central frequency, as is expected. Concanavalin A, a  $\beta$ -sheet protein, shows a weaker but similar trend. Our modeling of the influence of hydrogen bonds is through the electrostatic potential, which is a non-local approach. It may be that a more explicit treatment of hydrogen bonding would identify a stronger relationship between hydrogen bonding and the frequency shift.

## 2D IR spectra

Conformational dynamics of the protein and solvent are reflected in the 2D lineshapes. We have computed the absorptive 2D IR spectra for ubiquitin, concanavalin A, carbonmonoxy-myoglobin, lysozyme, ribonuclease A, and  $\alpha$ -lactalbumin based on the conformations sampled from the MD trajectories. We



**Figure 5.** Computed 1D spectra of ubiquitin (1UBQ) with  $^{13}\text{C}$ - $^{18}\text{O}$  isotope labeling of the  $\alpha$ -helix residues a), and  $\beta$ -sheet residues b). The solid black line in both panels represents the unlabeled computed spectrum, while the dashed blue line represents the isotope labeling cases. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

chose these proteins as a benchmark due to their different sizes and mixed structural compositions. For ubiquitin, we investigated the convergence of the computed spectra with the number of conformations sampled from the MD trajectory. Spectra computed with 200 snapshots sampled uniformly across the trajectory gave a computed spectrum very similar to that computed with 2000. So the 2D-spectra for the larger proteins were computed with 200 snapshots to limit the computational cost. Thus inhomogeneous broadening in the calculated 2D spectra is a result of the structural changes depicted by the snapshots, and homogeneous broadening was modeled by convoluting with a 2D Lorentzian bandwidth of  $10\text{ cm}^{-1}$ . The vibrational motional narrowing effect, whereby the line width may be overestimated by static structural snapshots,<sup>[98–100]</sup> may influence the computed spectrum. Whilst this is a significant effect for NMA in solution,<sup>[49]</sup> it may be less important for polypeptides, because of the spread of the multiple amide modes of the protein.<sup>[101]</sup> Different features can be seen in the spectra (Fig. 7). The intense peaks which are plotted with pump frequency,  $\omega_1$ , on the horizontal axis and the probe frequency,  $\omega_3$ , on the vertical axis as in Ref. 102 correspond to contributions from the  $\alpha$ -helices [ $\omega_1, \omega_3$ ] = [ $1635, 1635$ ]  $\text{cm}^{-1}$  in the case of ubiquitin. For concanavalin A, the anti-parallel sheet contribution can be seen as a weak peak appearing at [ $1660, 1660$ ]  $\text{cm}^{-1}$ . Two intense signals are also seen between [ $1625, 1625$ ]  $\text{cm}^{-1}$ , both associated with out-of-phase oscillations of carbonyls of the same  $\beta$ -strand.<sup>[74]</sup> A broad peak stretches along the diagonal from [ $1620, 1620$ ]  $\text{cm}^{-1}$  to [ $1670, 1670$ ]  $\text{cm}^{-1}$ . This agrees with experiment<sup>[102]</sup> and is associated with random coils in concanavalin A. Random coils probably contribute to the weaker signal at [ $1660, 1660$ ]  $\text{cm}^{-1}$ .

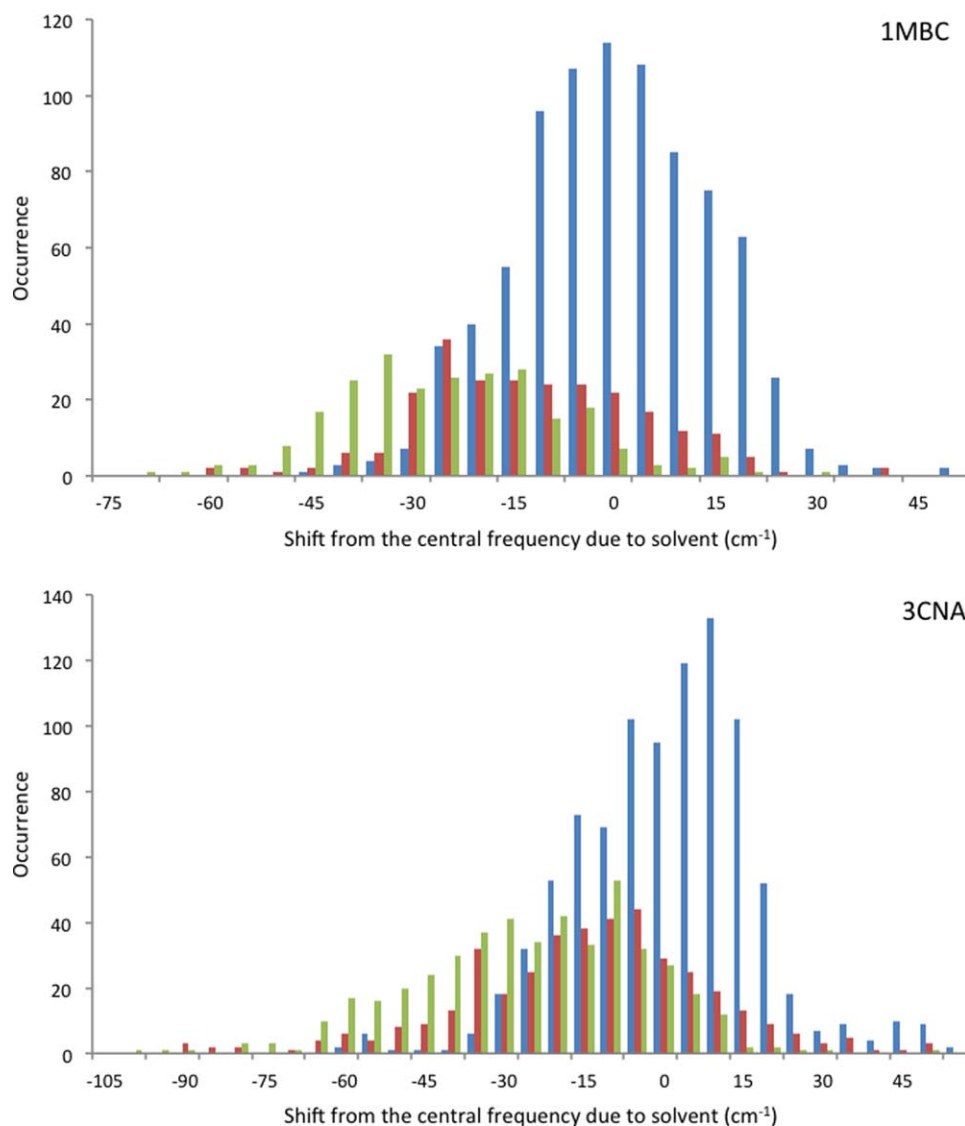
In all spectra shown in Figure 7, the elongation along the diagonal stretches from  $1620$  to  $1670\text{ cm}^{-1}$  and shows that the solvent-exposed residues experience a fluctuating

electrostatic environment, due to conformational disorder and fluctuations of the solvent. The spectrum of carbonmonoxymyoglobin is dominated by contributions from  $\alpha$ -helices. A broad and intense signal stretches from [ $1640, 1640$ ] to [ $1655, 1655$ ]  $\text{cm}^{-1}$ . This is also seen for  $\alpha$ -lactalbumin [ $1630, 1630$ ] to [ $1660, 1660$ ]  $\text{cm}^{-1}$ . Table 3 summarizes the location of the peak centers in the computed and the experimental spectra of Figure 7. Overall, our computed spectra agree well with the experimental spectra. The level of agreement is comparable with recently reported calculations<sup>[64,65]</sup> using different force fields and modeling protocols.

More detailed information can be accessed by enhancing the cross-peaks (Fig. 8). For example, coupled vibrations of different transitions are now revealed compared to the ZZZZ spectra. The cross-polarization  $\langle\text{ZZZZ}\rangle - 3\langle\text{ZXXZ}\rangle$  signal suppresses the diagonals to some extent and enhances other features of the 2D spectra. The off-diagonal peaks are better resolved than their counterparts in Figure 7, allowing the nature of the coupling between different local modes to be explored further. Much weaker peaks appear on the diagonal in Figure 8 compared to the intense diagonal peaks in Figure 7. These weak diagonal peaks (Fig. 8) are now comparable in intensity with the off-diagonal cross-peaks, and these features merge somewhat to give a broad peak which extends parallel to the  $\omega_1$  axis, as can be seen for example in the case of concanavalin A: peak I, Figure 8. The splitting between the positive and the negative signals of a cross-peak is a measurement of the coupling strength and is due to the off-diagonal anharmonicity. For example in concanavalin A, peak I indicates there is a strong coupling between structural elements, while in the rest of the proteins the analogous peak indicates weaker coupling. Table 4 summarizes the cross-peaks highlighted in Figure 8.

In the cross-polarization spectra of  $\alpha$ -lactalbumin and myoglobin (Fig. 8), there are signs of contribution from coupled





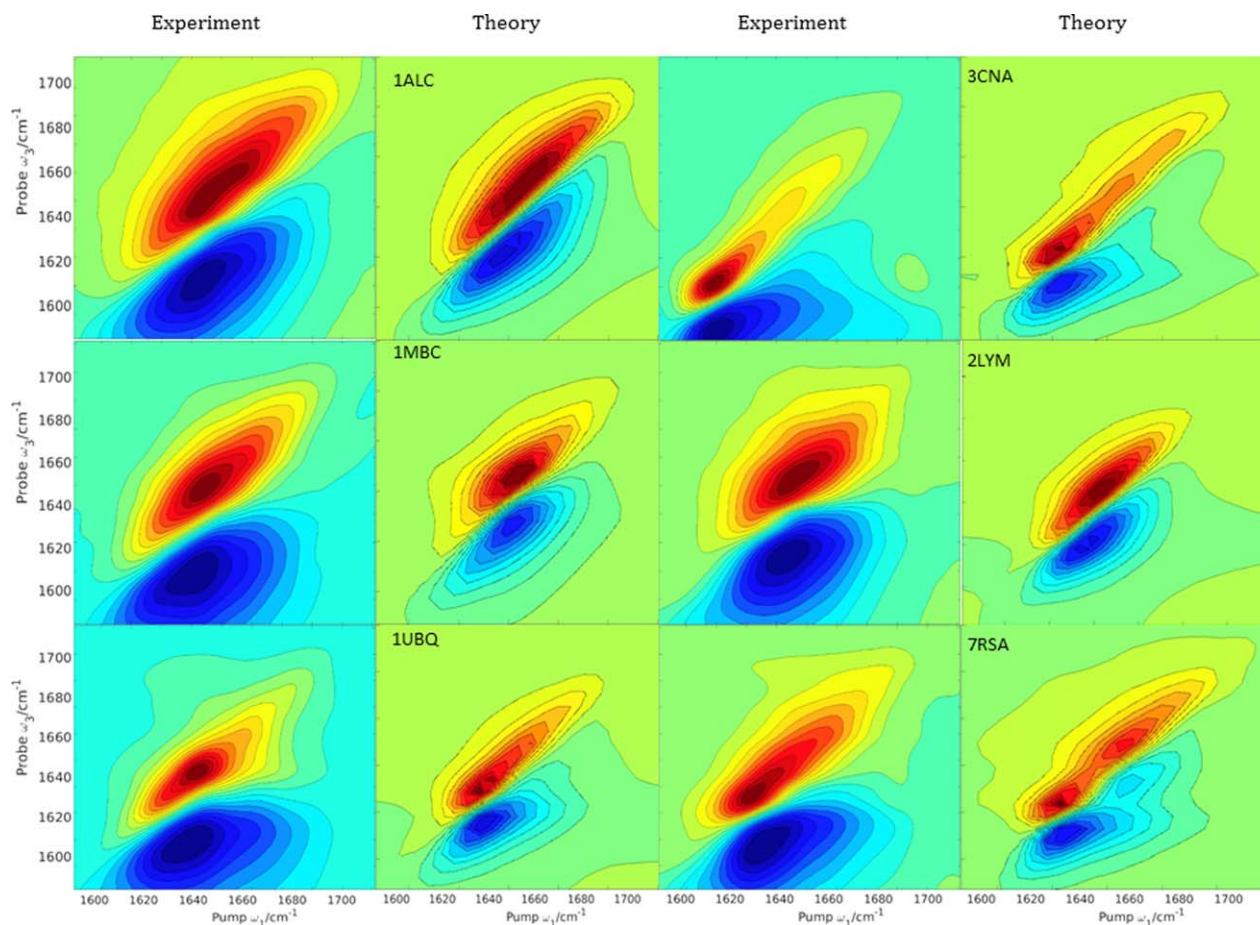
**Figure 6.** Histograms for carbonmonoxy-myoglobin (1MBC) and concanavalin A (3CNA), depicting the relationship between the number of hydrogen bonds to solvent and the shift in intrinsic frequency due to solvent [i.e., eq. (3) evaluated considering just the water molecules as the environment]. The blue line represents the case of no hydrogen bonding occurring between the carbonyl oxygen and solvent molecules, the red represents the case where there is one hydrogen bond, and green represents the case where there are two. The histograms were computed using 10 snapshots, evenly distributed across the first nanosecond of the MD simulation. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

unstructured coils. Unstructured coils appear as broad featureless peaks at  $[\omega_1, \omega_3] = 1650$  or  $1660$   $\text{cm}^{-1}$ ,<sup>[79]</sup> and the cross-polarization spectra reveal broad peaks stretching from  $1660$   $\text{cm}^{-1}$  toward  $1680$   $\text{cm}^{-1}$  due to  $\beta$ -strands coupled with unstructured coils (in the case of  $\alpha$ -lactalbumin: peak IV), and coupled unstructured coils with  $\beta$ -turns (in the case of myoglobin: peak III). The shape of the two cross peaks is different and the finding agrees with a previous study<sup>[103]</sup> to suppress random coil features.

Cross-peaks usually come in pairs, but when one half of the doublet is more intense (as is the case in peak IV for lactalbumin and myoglobin), it generally means that the off-diagonal anharmonicity is weak between the two local sites.<sup>[80]</sup> The splitting between the diagonal and off-diagonal peaks is underestimated in Figure 7, which may suggest that

the value of the anharmonicity used in the calculations should be greater. Choi et al.<sup>[63]</sup> and Chung et al.<sup>[104]</sup> previously investigated ubiquitin in two separate studies. Their spectra exhibited a Z-shape, due to the contribution from anti-parallel  $\beta$ -sheet strands. We find a similar Z-shape in both our ubiquitin and concanavalin A crossed-polarization spectra (Fig. 8).

The crossed-polarization spectroscopy is clearly a powerful, experimentally realizable approach, which provides an additional insight into the origins of the 2D signal. Some of these origins can also be explored using different computational strategies. We investigate the contributions to the 2D signal by isotope labeling the  $\alpha$ -helices and  $\beta$ -strand secondary structures of ubiquitin. Figure 9 shows the two cases when either the helices or the sheets are labeled.



**Figure 7.** Experimental<sup>[102]</sup> and computed 2D absorptive spectra using the ZZZZ scheme for:  $\alpha$ -lactalbumin (1ALC), carbonmonoxy-myoglobin (1MBC), ubiquitin (1UBQ), concanavalin A (3CNA), lysozyme (2LYM), and ribonuclease A (7RSA). Time delay ( $t_2$ ) in both the experimental and computed spectra was zero. The contours for the computed spectra are plotted with a 10% intensity of the maximum amplitude with 20 uniformly spread contours from the minimum to the maximum intensity. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

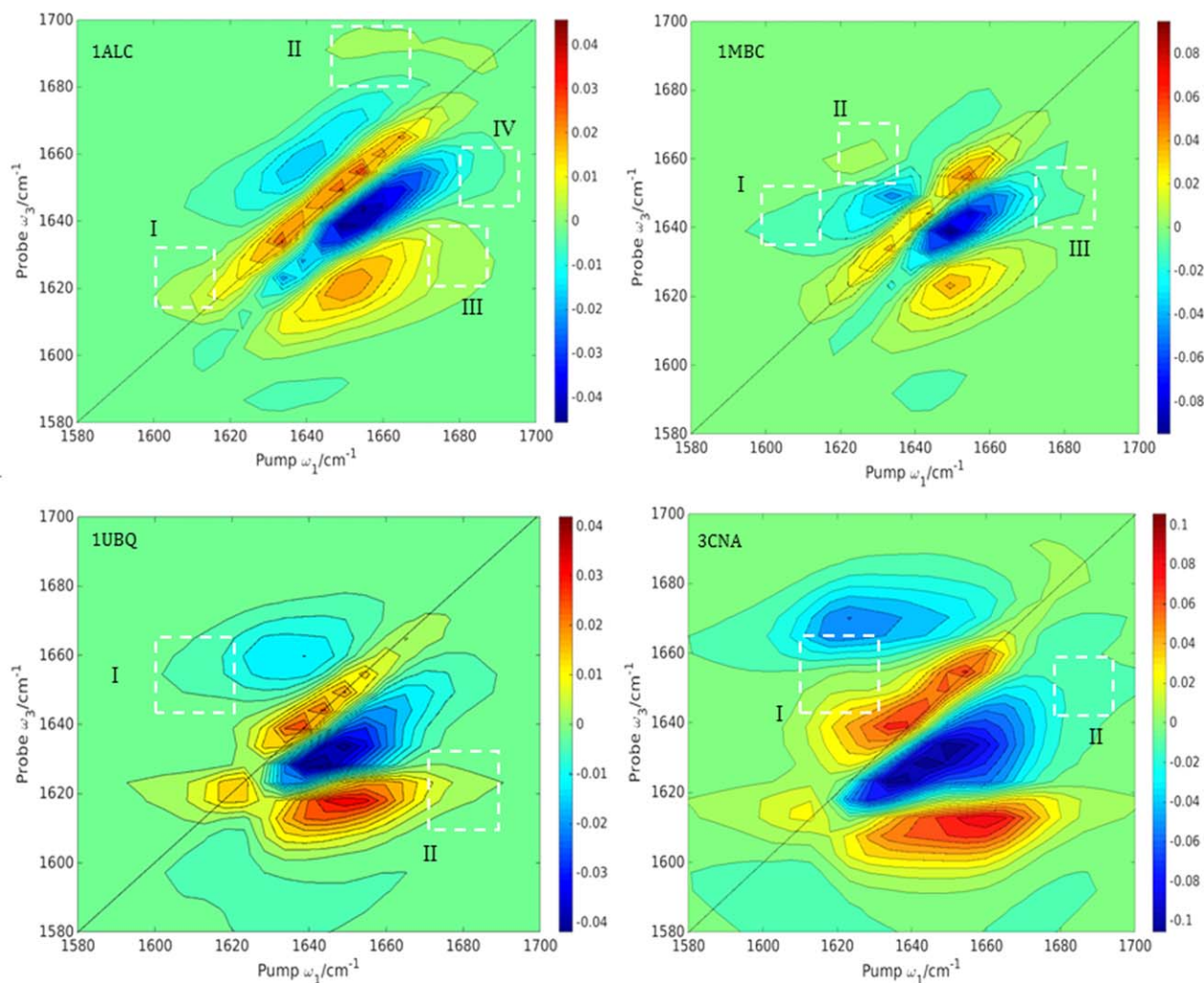
Both isotope labeled components have been shifted to  $\sim 1570\text{ cm}^{-1}$  and are consistent with the computed 1D spectra (Fig. 5). The diagonal peak intensities for the labeled  $\alpha$ -helices are clearly much weaker than those for the rest of the protein. This is more evident in the 2D spectra as the signal is

proportional to the fourth power of the transition dipole moment, whereas in the 1D spectra, the intensity depends on the square of the transition dipole moment. The cross-peak at  $\sim [1660, 1560]\text{ cm}^{-1}$  in Figure 9b indicates coupling between  $\beta$ -sheet and random coil components. The absence of an analogous cross-peak in Figure 9a indicates there is little or no coupling between  $\beta$ -sheet and  $\alpha$ -helices. We have computed the 2D spectra of the  $\alpha$ -helices and  $\beta$ -sheets separately (Fig. 10).

There are 12 residues in  $\alpha$ -helices and 26 residues in  $\beta$ -strands. Two Hamiltonians were constructed in which the size of the one-quantum Hamiltonian was  $12 \times 12$  in the first case and  $26 \times 26$  in the second, with all elements in the environment: main chain, side chains, and solvent molecules contributing to the electrostatic potential and shift in frequency. Figure 10 shows the contributions to the signal from both  $\alpha$ -helices and  $\beta$ -strands. The  $\beta$ -sheet spectrum is about four times more intense than the  $\alpha$ -helices spectrum consistent with the 2:1 ratio observed for the 1D spectrum. The contribution from the  $\alpha$ -helices is a diagonal elongated peak with its center at  $\sim [1640, 1640]\text{ cm}^{-1}$  while the  $\beta$ -strands contribute three peaks: two intense ones at  $\sim [1625, 1625]$  and  $\sim [1638, 1638]\text{ cm}^{-1}$  and a weak one at  $\sim [1660, 1660]\text{ cm}^{-1}$  indicating coupled local modes. There is also a weak cross-peak at  $[1680,$

**Table 3.** Locations of diagonal and off-diagonal peaks in the computed and experimental<sup>[102]</sup> 2D-IR spectra.

Protein	Diagonal peak locations ( $\omega_1 = \omega_3$ )/ $\text{cm}^{-1}$	
	Experiment	Computed
1ALC	1645	1650
1MBC	1650	1647
1UBQ	1640	1635
2LYM	1640	1635
7RSA	1640	1640
3CNA	1620	1625
Protein	Off-diagonal peak locations ( $\omega_1, \omega_3$ )/ $\text{cm}^{-1}$	
	Experiment	Computed
1ALC	(1640, 1610)	(1640, 1630)
1MBC	(1640, 1610)	(1640, 1635)
1UBQ	(1640, 1610)	(1630, 1625)
2LYM	(1640, 1620)	(1630, 1620)
7RSA	(1640, 1620)	(1620, 1620)
3CNA	(1620, 1600)	(1625, 1620)



**Figure 8.** Computed  $\langle ZZZZ \rangle - 3 \langle ZXXZ \rangle$  cross-peak 2D IR absorptive spectra computed for  $\alpha$ -lactalbumin (1ALC), myoglobin (1MBC), ubiquitin (1UBQ), and concanavalin A (3CNA). Cross-peaks are highlighted with white dotted squares. The contours are plotted with a 10% intensity of the maximum amplitude with 20 uniformly spread contours from the minimum to the maximum intensity. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

1620]  $\text{cm}^{-1}$  (Peak I). The presence of Peak I in Figures 10b and Peaks I and II Figure 9a is further indication of the coupling between  $\beta$ -secondary structure components.

### Transition dipole strengths

Non-linear spectroscopies are more sensitive to transition dipole strengths than linear spectroscopies. The 2D intensities

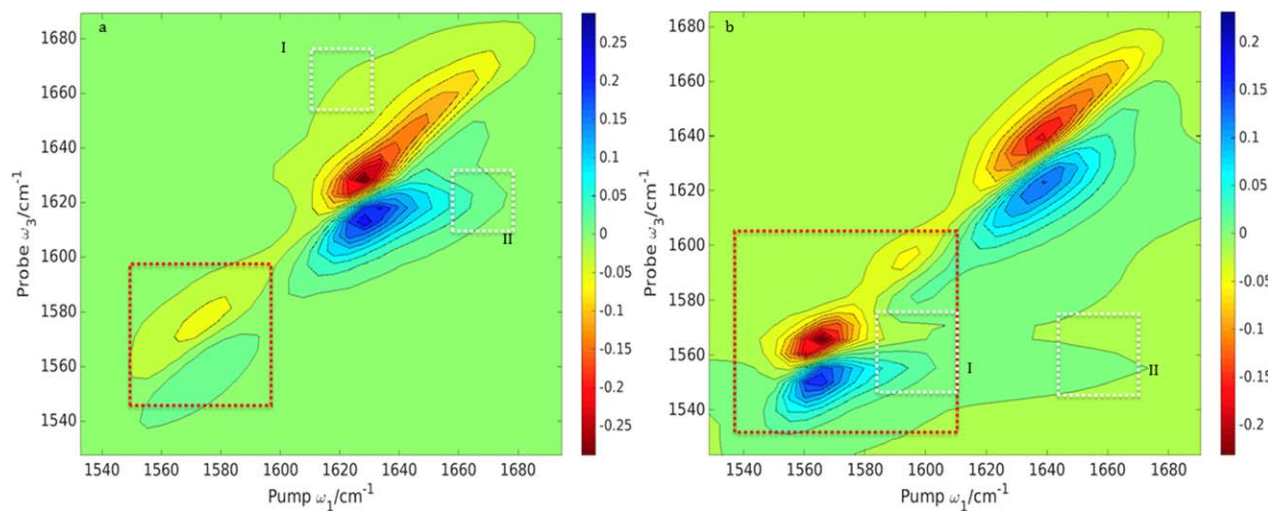
scale as  $|\vec{\mu}|^4$  compared to 1D intensities that scale as  $|\vec{\mu}|^2$ . Grechko and Zanni<sup>[66]</sup> studied the 1D and 2D IR of a model  $\alpha$ -helical system and concluded that the transition dipole strength of the random coil state is 0.12 D<sup>2</sup>, and that a greater magnitude indicates a more ordered system with vibrational excitonic states forming associated with secondary structure. Here, we extend that consideration to six of the proteins, extracting the two-quantum transition dipole moments that contribute to the 2D diagonal signal. In Table 5, we compare the per residue contribution to the 2D intensities,  $|\vec{\mu}|_{\text{residue}}^4$  for the aforementioned proteins. The absolute value of the transition dipole,  $|\vec{\mu}|_{\text{total}}^4$  was computed as a sum over all transition dipole moments. One might anticipate that more ordered structures such as concanavalin A should give more delocalized excitons, which should be manifested in more intense bands.

The per residue contribution,  $|\vec{\mu}|_{\text{residue}}^2$  which is computed here by taking the square root of  $|\vec{\mu}|_{\text{residue}}^4$ , confirms delocalized excitonic vibrations as expected from ordered

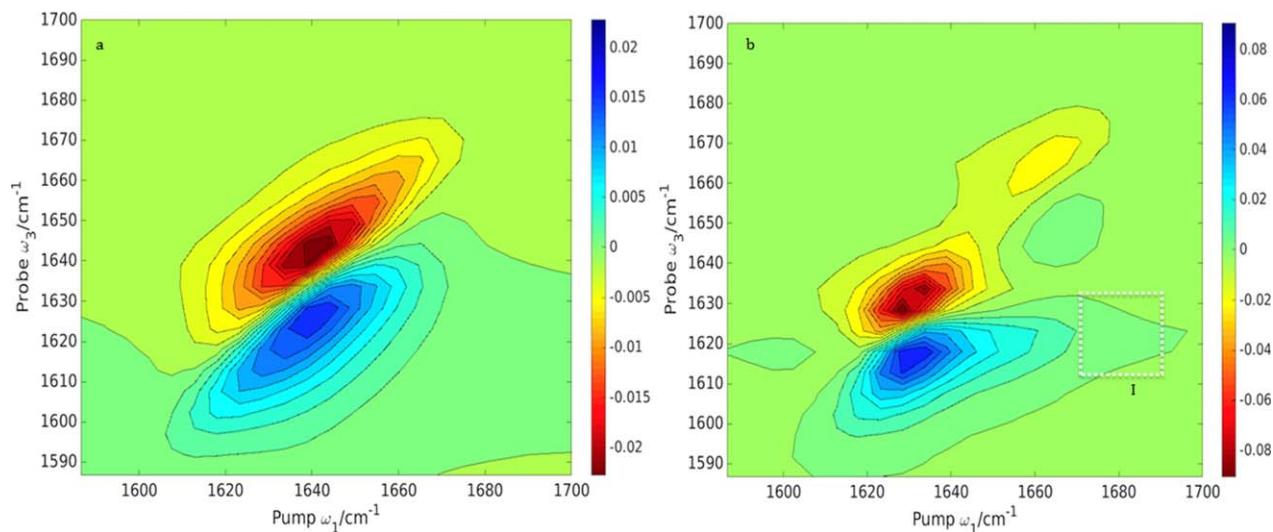
**Table 4.** The cross-peaks in the cross-polarization 2D-IR spectra (Fig. 8) for  $\alpha$ -lactalbumin (1ALC), carbonmonoxy-myoglobin (1MBC), ubiquitin (1UBQ) and concanavalin A (3CNA).

Protein	Peak (strength)				Structural components involved			
1ALC	I (w)	II (w)	III (w)	IV (w)	$\beta$	$\alpha + \beta$	$\beta$	$\beta + \text{coil}$
1MBC	I (w)	II (w)	III (w)		$\alpha$	$\alpha$	$\alpha + \text{coil}$	
1UBQ	I (w)	II (w)			$\alpha$	$\beta$		
3CNA	I (s)	II (w)			$\beta$	$\beta$		
The coupling is characterized as either strong (s) or weak (w).								





**Figure 9.** Computed 2D absorptive spectra using the ZZZZ scheme of cases where the  $\alpha$ -helices a) and  $\beta$ -strands b) of ubiquitin (1UBQ) are isotope labeled. The contours are plotted with a 10% intensity of the maximum amplitude and 20 uniformly spread contours from the minimum to the maximum intensity for the left panel and 20 similar uniformly spread contours for the right panel. The isotope labeled components are highlighted with red dotted squares and the cross peaks with white dotted ones. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**Figure 10.** Computed 2D absorptive spectra using the ZZZZ scheme of  $\alpha$  a) and  $\beta$  b) secondary structure separately of ubiquitin (1UBQ). The contours are plotted with a 10% intensity of the maximum amplitude and 20 uniformly spread contours from the minimum to the maximum intensity for the left panel and 20 similar uniformly spread contours for the right panel. Cross-peaks are highlighted with white dotted squares. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

structured systems, and all the values fall above the threshold observed by Grechko and Zanni.<sup>[66]</sup> However, there is not an obvious relationship between the proportion of coil structure in the protein and the intensity of the 2D diagonal

signal. This discrepancy might be explained by the fact coil structures in folded proteins do not necessarily have the same properties as random coils in unfolded peptides such as the ones studied by Grechko and Zanni.<sup>[66]</sup>

**Table 5.** The studied proteins (with PDB codes), the number of residues, and fraction of coil composition, the two-quantum transition dipole strengths per residue  $|\vec{\mu}|_{\text{total/residue}}^4$ .

Protein	Number of residues	% Coil residues	$ \vec{\mu} _{\text{total/residue}}^4 (D^4)$	$ \vec{\mu} _{\text{total/residue}}^2 (D^2)$
$\alpha$ -lactalbumin (1ALC)	122	19	0.47	0.69
Myoglobin (1MBC)	152	15	0.47	0.69
Lysozyme (2LYM)	128	14	0.40	0.63
Concanavalin A (3CNA)	237	14	0.80	0.89
Ubiquitin (1UBQ)	76	16	0.32	0.57
Ribonuclease A (7RSA)	123	28	0.48	0.69

The fraction of coil composition was computed from PROMOTIF<sup>[92]</sup> as in Table 2.



## Conclusion

Overall, the calculations have quantitatively reproduced 1D-IR spectra for the proteins studied. We have investigated the sensitivity of the Amide I peaks to conformational dynamics. Our motivation was to investigate the effect of conformational dynamics on the overall lineshapes. The mixing of the local modes that underlies the 1D-IR spectra is revealed more explicitly, although not with complete clarity in the 2D-IR spectra. Our study shows that the Cho model<sup>[47]</sup> in combination with MD simulations confirm the importance of conformational dynamics in affecting the overall absorbance. We would expect different maps<sup>[50,55–60]</sup> would yield qualitatively similar findings, although some of the quantitative detail might vary.<sup>[62,105]</sup>

The computed 2DIR spectra for ubiquitin, carbonmonoxy-myoglobin,  $\alpha$ -lactalbumin, concanavalin A, flavodoxin, and ribonuclease A are broadly in agreement with that presented in previous studies.<sup>[62,63,102,104]</sup> Solvation plays a significant role in the nature of the computed spectra. Through calculations of the Amide I band and analysis of the exciton Hamiltonian matrix, it is possible to relate specific conformational features to the IR spectrum. Delving deeper into the dynamics of proteins and studying single snapshots for both concanavalin A and carbonmonoxy-myoglobin provides a stepping stone to understanding the appearance of the 2D spectra. The width of the 1D spectra is reflected in the linewidth of the diagonal on the 2D spectra. The off-diagonal cross peaks were enhanced for ubiquitin, carbonmonoxy-myoglobin,  $\alpha$ -lactalbumin, and concanavalin A by computing the  $\langle ZZZZ \rangle - 3 \langle ZXXZ \rangle$  spectra for each case. We explored whether there is a relationship between the contribution to the 2D diagonal signal and the fraction of random coil, but none was evident. We are currently studying the dynamics and 2D IR spectroscopy of protein-ligand binding, and investigating the behavior of coupled modes during binding and dissociation.

## Acknowledgments

We thank Rachel Hill and Nick Besley (University of Nottingham) for helpful discussions. We are grateful for access to the University of Nottingham High Performance Computing facility and to the Mid-Plus Regional Centre of Excellence for Computational Science, Engineering and Mathematics; this work was supported by the Engineering and Physical Sciences Research Council [grant number EP/K000128/1].

**Keywords:** two-dimensional infrared spectroscopy · protein · molecular dynamics simulation

How to cite this article: F. S. Hussein, D. Robinson, N. T. Hunt, A. W. Parker, J. D. Hirst. *J. Comput. Chem.* **2017**, *38*, 1362–1375. DOI: 10.1002/jcc.24674

[1] A. Elliott, E. J. Ambrose, *Nature* **1950**, *165*, 921.

[2] S. Krimm, *Biopolymers* **1983**, *22*, 217.

[3] S. Krimm, J. Bandekar, *Adv. Prot. Chem.* **1986**, *38*, 181.

[4] F. Siebert, *Method Enzymol.* **1995**, *246*, 501.

- [5] J. M. Chalmers, P. R. Griffiths, *Handbook of Vibrational Spectroscopy*; Wiley-Blackwell, New Jersey, USA, **2002**.
- [6] A. K. Dioumaev, *Biochemistry (Moscow)* **2001**, *66*, 1269.
- [7] H. Susi, D. M. Byler, *Method Enzymol.* **1986**, *130*, 290.
- [8] A. Dong, P. Huang, W. S. Caughey, *Biochemistry* **1990**, *29*, 3303.
- [9] W. K. Surewicz, H. H. Mantsch, *Biochim. Biophys. Acta* **1988**, *952*, 115.
- [10] H. Susi, D. M. Byler, *Biochem. Biophys. Res. Commun.* **1983**, *115*, 391.
- [11] S. Y. Venyaminov, N. N. Kalnin, *Biopolymers* **1990**, *30*, 1243.
- [12] N. N. Kalnin, I. A. Baikalov, S. Y. Venyaminov, *Biopolymers* **1990**, *30*, 1273.
- [13] H. Torii, M. Tasumi, *J. Raman Spectrosc.* **1998**, *29*, 81.
- [14] C. Lee, M. H. Cho, *J. Phys. Chem. B* **2004**, *108*, 20397.
- [15] Y. Tanimura, S. Mukamel, *J. Chem. Phys.* **1993**, *99*, 9496.
- [16] P. Hamm, M. H. Lim, R. M. Hochstrasser, *J. Phys. Chem. B* **1998**, *102*, 6123.
- [17] M. C. Asplund, M. T. Zanni, R. M. Hochstrasser, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8219.
- [18] M. T. Zanni, S. Gnanakaran, J. Stenger, R. M. Hochstrasser, *J. Phys. Chem. B* **2001**, *105*, 6520.
- [19] O. Golonzka, M. Khalil, N. Demirdoven, A. Tokmakoff, *Phys. Rev. Lett.* **2001**, *86*, 2154.
- [20] P. Hamm, R. M. Hochstrasser, M. D. Fayer, *Ultrafast Infrared and Raman Spectroscopy*; Dekker, New York, **2001**.
- [21] S. Woutersen, P. Hamm, *J. Chem. Phys.* **2001**, *114*, 2727.
- [22] N. H. Ge, M. T. Zanni, R. M. Hochstrasser, *J. Phys. Chem. A* **2002**, *106*, 962.
- [23] I. V. Rubtsov, J. P. Wang, R. M. Hochstrasser, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5601.
- [24] J. Park, R. M. Hochstrasser, *J. Chem. Phys.* **2006**, *323*, 78.
- [25] Y. S. Kim, J. P. Wang, R. M. Hochstrasser, *J. Phys. Chem. B* **2005**, *109*, 7511.
- [26] C. Fang, J. Wang, Y. S. Kim, A. K. Charnley, W. Barber-Armstrong, A. B. Smith, S. M. Decatur, R. M. Hochstrasser, *J. Phys. Chem. B* **2004**, *108*, 10415.
- [27] H. Ishikawa, K. Kwak, J. K. Chung, S. Kim, M. D. Fayer, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 8619.
- [28] H. Ishikawa, I. J. Finkelstein, S. Kim, K. Kwak, J. K. Chung, K. Wakasugi, A. M. Massari, M. D. Fayer, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16116.
- [29] K. Adamczyk, N. Simpson, G. M. Greetham, A. Gumiero, M. A. Walsh, M. Towrie, A. W. Parker, N. T. Hunt, *Chem. Sci.* **2015**, *6*, 505.
- [30] D. Zimdars, A. Tokmakoff, S. Chen, S. R. Greenfield, M. D. Fayer, T. I. Smith, H. A. Schwettman, *Phys. Rev. Lett.* **1993**, *70*, 2718.
- [31] Y. S. Kim, L. Liu, P. H. Axelsen, R. M. Hochstrasser, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17751.
- [32] C. T. Middleton, P. Marek, P. Cao, C. C. Chiu, S. Singh, A. M. Woys, J. J. Pablo, D. P. Raleigh, M. T. Zanni, *Nat. Chem.* **2012**, *4*, 355.
- [33] D. B. Strasfeld, Y. L. Ling, S. H. Shim, M. T. Zanni, *J. Am. Chem. Soc.* **2008**, *130*, 6698.
- [34] S. H. Shim, D. B. Strasfeld, Y. L. Ling, M. T. Zanni, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14197.
- [35] A. Remorino, I. V. Korendovych, Y. B. Wu, W. F. DeGrado, R. M. Hochstrasser, *Science* **2011**, *332*, 1206.
- [36] A. Ghosh, J. Qiu, W. F. DeGrado, R. M. Hochstrasser, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6115.
- [37] J. Manor, P. Mukherjee, Y. S. Lin, H. Leonov, J. L. Skinner, M. T. Zanni, I. T. Arkin, *Structure* **2009**, *17*, 247.
- [38] Z. Ganim, H. S. Chung, A. W. Smith, L. P. DeFlores, K. C. Jones, A. Tokmakoff, *Acc. Chem. Res.* **2008**, *41*, 432.
- [39] R. M. Hochstrasser, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14190.
- [40] K. Adamczyk, M. Candelaresi, K. Robb, A. Gumiero, M. A. Walsh, A. W. Parker, P. A. Hoskisson, N. P. Tucker, N. T. Hunt, *Meas. Sci. Technol.* **2012**, *23*, 062001.
- [41] N. T. Hunt, *Chem. Soc. Rev.* **2009**, *38*, 1837.
- [42] H. Torii, M. Tasumi, *J. Chem. Phys.* **1992**, *96*, 3379.
- [43] P. Hamm, S. Woutersen, *Bull. Chem. Soc. Jpn.* **2002**, *75*, 985.
- [44] J. W. Brauner, C. R. Flach, R. Mendelsohn, *J. Am. Chem. Soc.* **2005**, *127*, 100.
- [45] S. Y. Cha, S. H. Ham, M. H. Cho, *J. Chem. Phys.* **2002**, *117*, 740.
- [46] J. H. Choi, S. Ham, M. Cho, *J. Chem. Phys.* **2002**, *117*, 6821.
- [47] S. Ham, M. Cho, *J. Chem. Phys.* **2003**, *118*, 6915.
- [48] S. Ham, S. Cha, J. H. Choi, M. Cho, *J. Chem. Phys.* **2003**, *119*, 1451.

- [49] K. Kwac, M. Cho, *J. Chem. Phys.* **2003**, *119*, 2247.
- [50] T. M. Watson, J. D. Hirst, *Phys. Chem. Chem. Phys.* **2004**, *6*, 998.
- [51] T. M. Watson, J. D. Hirst, *Mol. Phys.* **2005**, *103*, 1531.
- [52] P. Bour, T. A. Keiderling, *J. Chem. Phys.* **2003**, *119*, 11253.
- [53] P. Bour, T. A. Keiderling, *J. Phys. Chem. B* **2005**, *109*, 5348.
- [54] E. L. Karjalainen, T. Ersmark, A. Barth, *J. Phys. Chem. B* **2012**, *116*, 4831.
- [55] T. Hayashi, W. Zhuang, S. Mukamel, *J. Phys. Chem. A* **2005**, *109*, 9747.
- [56] T. L. Jansen, J. Knoester, *J. Chem. Phys.* **2006**, *124*, 044502.
- [57] T. L. Jansen, A. G. Dijkstra, T. M. Watson, J. D. Hirst, J. Knoester, *J. Chem. Phys.* **2006**, *125*, 44312.
- [58] H. Torii, T. Tatsumi, T. Kanazawa, M. Tasumi, *J. Phys. Chem. B* **1998**, *102*, 309.
- [59] J. R. Schmidt, S. A. Corcelli, J. L. Skinner, *J. Chem. Phys.* **2004**, *121*, 8887.
- [60] Y. S. Lin, J. M. Shorb, P. Mukherjee, M. T. Zanni, J. L. Skinner, *J. Phys. Chem. B* **2009**, *113*, 592.
- [61] N. G. Mirkin, S. Krimm, *J. Am. Chem. Soc.* **1991**, *113*, 9742.
- [62] Z. Ganim, A. Tokmakoff, *Biophys. J.* **2006**, *91*, 2636.
- [63] J. H. Choi, H. Lee, K. K. Lee, S. Hahn, M. Cho, *J. Chem. Phys.* **2007**, *126*, 045102.
- [64] A. S. Bondarenko, T. L. C. Jansen, *J. Chem. Phys.* **2015**, *142*, 212437.
- [65] A. V. Cunha, A. S. Bondarenko, T. L. C. Jansen, *J. Chem. Theory Comput.* **2016**, *12*, 3982.
- [66] M. Grechko, M. T. Zanni, *J. Chem. Phys.* **2012**, *137*, 184202.
- [67] A. S. Davydov, *The Theory of Molecular Excitons*; McGraw-Hill, New York, **1962**.
- [68] B. M. Bulheller, J. D. Hirst, *Bioinformatics* **2009**, *25*, 539.
- [69] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell, *J. Chem. Theor. Comput.* **2012**, *8*, 3257.
- [70] A. D. MacKerell, M. Feig, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1400.
- [71] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kucsera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kucsera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.
- [72] J. Huang, A. D. MacKerell, *J. Comput. Chem.* **2013**, *34*, 2135.
- [73] D. Abramavicius, W. Zhuang, S. Mukamel, *J. Phys. Chem. B* **2004**, *108*, 18034.
- [74] M. Reppert, A. Tokmakoff, *J. Chem. Phys.* **2013**, *138*, 134116.
- [75] S. Roy, J. Lessing, G. Meisl, Z. Ganim, A. Tokmakoff, J. Knoester, T. L. C. Jansen, *J. Chem. Phys.* **2011**, *135*, 234507.
- [76] C. Fang, J. Wang, A. K. Charnley, W. Barber-Armstrong, A. B. Smith, S. M. Decatur, R. M. Hochstrasser, *Chem. Phys. Lett.* **2003**, *382*, 586.
- [77] L. Wang, C. T. Middleton, M. T. Zanni, J. L. Skinner, *J. Phys. Chem. B* **2011**, *115*, 3713.
- [78] L. Wang, C. T. Middleton, S. Singh, A. S. Reddy, A. M. Woys, D. B. Strasfeld, P. Marek, D. P. Raleigh, J. J. Pablo, M. T. Zanni, J. L. Skinner, *J. Am. Chem. Soc.* **2011**, *133*, 16062.
- [79] M. D. Fayer, *Ultrafast Infrared Vibrational Spectroscopy*; CRC Press, Boca Raton, **2013**.
- [80] M. T. Zanni, P. Hamm, *Concepts and Methods of 2D Infrared Spectroscopy*; Cambridge University Press, Cambridge, **2013**.
- [81] M. Khalil, N. Demirdöven, A. Tokmakoff, *J. Phys. Chem. A* **2003**, *107*, 5258.
- [82] R. M. Hochstrasser, *Chem. Phys.* **2001**, *266*, 273.
- [83] J. Dreyer, A. M. Moran, S. Mukamel, *Bull. Korean Chem. Soc.* **2003**, *24*, 1091.
- [84] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781.
- [85] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
- [86] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089.
- [87] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327.
- [88] G. J. Martyna, D. J. Tobias, M. L. Klein, *J. Chem. Phys.* **1994**, *101*, 4177.
- [89] S. Nose, *J. Chem. Phys.* **1984**, *81*, 511.
- [90] W. G. Hoover, *Phys. Rev. A* **1985**, *31*, 1695.
- [91] S. E. Feller, Y. H. Zhang, R. W. Pastor, B. R. Brooks, *J. Chem. Phys.* **1995**, *103*, 4613.
- [92] E. G. Hutchinson, J. M. Thornton, *Prot. Sci.* **1996**, *5*, 212.
- [93] H. Torii, *J. Phys. Chem. A* **2004**, *108*, 2103.
- [94] M. Lima, R. Chelli, V. V. Volkov, R. Righini, *J. Chem. Phys.* **2009**, *130*, 204518.
- [95] A. M. Woys, A. M. Almeida, L. Wang, C. C. Chiu, M. McGovern, J. J. Pablo, J. L. Skinner, S. H. Gellman, M. T. Zanni, *J. Am. Chem. Soc.* **2012**, *134*, 19118.
- [96] S. D. Moran, S. M. Decatur, M. T. Zanni, *J. Am. Chem. Soc.* **2012**, *134*, 18410.
- [97] L. E. Buchanan, J. K. Carr, A. M. Fluitt, A. J. Hoganson, S. D. Moran, J. J. Pablo, J. L. Skinner, M. T. Zanni, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 5796.
- [98] J. D. Eaves, A. Tokmakoff, P. L. Geissler, *J. Phys. Chem. A* **2005**, *109*, 9424.
- [99] J. D. Smith, R. J. Saykally, P. L. Geissler, *J. Am. Chem. Soc.* **2007**, *129*, 13847.
- [100] B. M. Auer, J. L. Skinner, *J. Chem. Phys.* **2007**, *127*, 104105.
- [101] W. R. W. Welch, J. Kubelka, *J. Phys. Chem. B* **2012**, *116*, 10739. —.
- [102] C. R. Baiz, C. S. Peng, M. E. Reppert, K. C. Jones, A. Tokmakoff, *Analyst* **2012**, *137*, 1793.
- [103] C. T. Middleton, L. E. Buchanan, E. B. Dunkelberger, M. T. Zanni, *J. Phys. Chem. Lett.* **2011**, *2*, 2357.
- [104] H. S. Chung, M. Khalil, A. W. Smith, Z. Ganim, A. Tokmakoff, *Proc. Natl. Acad. Sci. USA* **2004**, *102*, 612.
- [105] A. M. Woys, Y. S. Lin, A. S. Reddy, W. Xiong, J. J. Pablo, J. L. Skinner, M. T. Zanni, *J. Am. Chem. Soc.* **2010**, *132*, 2832.

Received: 1 June 2016

Revised: 22 October 2016

Accepted: 29 October 2016

Published online on 21 November 2016