

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/49658901>

# Two-dimensional near-ultraviolet spectroscopy of aromatic residues in amyloid fibrils: A first principles study

ARTICLE *in* PHYSICAL CHEMISTRY CHEMICAL PHYSICS · FEBRUARY 2011

Impact Factor: 4.49 · DOI: 10.1039/c0cp02047h · Source: PubMed

---

CITATIONS

19

---

READS

21

2 AUTHORS, INCLUDING:



Shaul Mukamel

University of California, Irvine

852 PUBLICATIONS 23,706 CITATIONS

SEE PROFILE

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 2394–2400

www.rsc.org/pccp

PAPER

## Two-dimensional near-ultraviolet spectroscopy of aromatic residues in amyloid fibrils: a first principles study

Jun Jiang\* and Shaul Mukamel\*

Received 5th October 2010, Accepted 9th November 2010

DOI: 10.1039/c0cp02047h

We report a first principles study of two dimensional electronic spectroscopy of aromatic side chain transitions in the 32-residue  $\beta$ -amyloid ( $A\beta_{9-40}$ ) fibrils in the near ultraviolet (250–300 nm). An efficient exciton Hamiltonian with electrostatic fluctuations (EHEF) algorithm is used to compute the electronic excitations in the presence of environmental fluctuations. The through-space inter- and intra-molecular interactions are calculated with high level quantum mechanics (QM) approaches, and interfaced with molecular mechanics (MM) simulations. Distinct two dimensional near ultraviolet (2DNUV) spectroscopic signatures are identified for different aromatic transitions, and the couplings between them. 2DNUV signals associated with the transition couplings are shown to be very sensitive to the change of residue-residue interactions induced by residue mutations. Our simulations suggest that 2DNUV spectra could provide a useful local probe for the structure and kinetics of fibrils.

### 1. Introduction

Many important biological functions of proteins depend crucially on their fluctuating structures.<sup>1,2</sup> Optical spectroscopy is a widely used tool for identifying structural details and monitoring the transformations of biological complexes.<sup>3–5</sup> Traditional one-dimensional (1D) Ultraviolet (UV) techniques such as Linear absorption (LA) and circular dichroism (CD) are routinely used for probing proteins. Proteins are made of amino acid side chains connected by the peptide backbone. The  $n\pi^*$  and  $\pi\pi^*$  transitions of the peptide backbone in the far-ultraviolet (FUV) (190–250 nm) are used for estimating protein secondary structure content.<sup>6,7</sup> Other important transitions correspond to the aromatic side chains, which absorb in the near-ultraviolet (NUV) (> 250 nm) regime. These are easier to characterize since aromatic amino acids are relatively rare: Only three amino acids contain aromatic side chains: phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp). FUV spectroscopy gives a global picture of secondary structure, while NUV spectroscopy provides a more detailed local probe.<sup>8</sup>

Multidimensional spectroscopy uses sequences of laser pulses to excite the vibrational or electronic degrees of freedom and watch for correlated events which take place during two (or more) controlled time intervals. The resulting correlation plots provide a multidimensional view of the relevant molecular motions, structures, and dynamical events.<sup>9–13</sup> The development of intense, coherent, high repetition rate ultrafast sources in the

UV<sup>3,14–17</sup> should allow to extend multidimensional spectroscopy into the NUV.<sup>18,19</sup> Two dimensional UV (2DUV) spectroscopy has a great potential for protein structure refinement,<sup>20</sup> and 2DNUV could offer an additional powerful high resolution local observation window.

Accurate microscopic simulation algorithms are essential for designing and exploring the possible applications of 2DUV techniques. The UV spectra are strongly affected by environmental fluctuations of the backbone, side chains and the solvent. Direct *Ab initio* simulations of UV signals for large proteins with hundreds of chromophores are too expensive. The efficient map method widely used in two dimensional IR (2DIR) simulations<sup>9,21,22</sup> is not applicable to 2DUV. Since vibrational chromophores are highly localized, the IR transition can be modeled as point dipoles, and the electrostatic fluctuations may be simulated by sampling the electric fields at a few points. UV simulations are much more challenging. The transition charge densities are delocalized and proper description of the electrostatic interactions involves the computation of extended charge space distributions. Motions of proteins and the surrounding water lead to fluctuations of the electrostatic environment, which affect the intra- and inter-molecular interactions, excitation energies and the local Hamiltonian. A proper description environmental fluctuations which is crucial for the prediction of spectra requires thousands of MD snapshots.<sup>9</sup> Since repeated first principles calculations are prohibitively expensive, it is necessary to adopt a QM/MM approach.

We have developed the EHEF (exciton Hamiltonian with electrostatic fluctuations) algorithm for first principles simulations of 2DUV spectra of proteins. In this algorithm, proteins

Chemistry Department, University of California Irvine, USA.  
E-mail: jiangj1@uci.edu, smukamel@uci.edu; Fax: 1-949 824 8571;  
Tel: 1-949 824 6164

are divided into fragments of amino acids and peptide bond units. Electronic excitations of isolated aromatic chromophores are treated at a high *Ab initio* level, the complete active space self-consistent field (CASSCF).<sup>23</sup> The surrounding protein amino acids and the solvent are treated by density function theory (DFT).<sup>24</sup> The inter- and intra-molecular interactions are calculated using a grid technique and effective charges fitted from the electrostatic potential. EHEF also provides a convenient interface for quantum mechanics (QM) and molecular mechanics (MM) calculations. The utilization of atomic frames in EHEF makes it possible to compute Coulomb interactions in all molecular dynamic (MD) snapshots, taking into account of environmental fluctuations at the QM level, yet avoiding repeated QM calculations.

Here we employ this protocol previously applied in the FUV region<sup>7</sup> to simulate the 2DNUV spectra of the 32-residue  $\beta$ -amyloid (A $\beta_{9-40}$ ) fibril. Misfolded proteins which form amyloid fibrils are associated with more than 20 neurodegenerative diseases,<sup>25,26</sup> and their kinetic pathways and toxicity are under active study.<sup>27,28</sup> In the case of Alzheimer's disease, fibrils are composed of  $\beta$ -amyloid (A $\beta$ ) peptides ranging from 39 to 42 residues, rich in  $\beta$ -sheet secondary structure. Our simulations demonstrate that 2DNUV techniques should provide a novel tool for visualizing molecular interactions, identifying the local structures, and monitoring transformations in proteins.

## 2. The simulation protocol

### 2.1 The exciton Hamiltonian

The NUV transitions will be modeled using the Frenkel exciton Hamiltonian,<sup>29–32</sup> which is constructed as described in ref. 7:

$$\hat{H} = \sum_{me} \varepsilon_{me} \hat{B}_{me}^\dagger \hat{B}_{me} + \sum_{me, nf}^{m \neq n} J_{me, nf} \hat{B}_{me}^\dagger \hat{B}_{nf} \quad (1)$$

where  $me$  is the  $e$  electronic transition on the  $m$ -th aromatic chromophore.  $\hat{B}_{me}^\dagger$  is the creation Pauli operator which promotes the chromophore  $m$  into the excited state  $e$ , and  $\hat{B}_{me}$  is the corresponding annihilation operator.

The excited state energy  $\varepsilon_{me}$  can be calculated as the summation of the excited energy of the isolated chromophore and its interactions with local electrostatic fields:

$$\varepsilon_{me} = \varepsilon_{me,0} + \sum_{k \neq m}^k \frac{1}{4\pi\epsilon} \times \int \int d\mathbf{r}_m d\mathbf{r}_k \frac{[\rho_m^{ee}(\mathbf{r}_m) - \rho_m^{gg}(\mathbf{r}_m)][\rho_k^{gg}(\mathbf{r}_k)]}{|\mathbf{r}_m - \mathbf{r}_k|} \quad (2)$$

in which the first term  $\varepsilon_{me,0}$  is the excited state energy of the isolated chromophore, and the second term describes the inter- and intra-molecular electrostatic corrections. Here  $k$  runs over all the atomic or molecular sites surrounding the excited chromophore,  $\mathbf{r}_m$  and  $\mathbf{r}_k$  are the positions coordinates, and  $\rho_m^{gg}$  and  $\rho_k^{gg}(\rho_m^{ee})$  represent the ground (excited) state charge density.  $\epsilon$  is the dielectric constant. We have used the vacuum dielectric constant ( $\epsilon = 1$ ). The variation of the dielectric constant on the molecular length scale was not taken into account in our simulations. This will be studied in the future.

The resonant coupling between singly-excited states  $me$  and  $nf$  is given by  $\langle 0 | \hat{B}_{me} \hat{H} \hat{B}_{nf}^\dagger | 0 \rangle = J_{me, nf}$  where  $|0\rangle$  is the ground state.

$$J_{me, nf} = \frac{1}{4\pi\epsilon} \int \int d\mathbf{r}_m d\mathbf{r}_n \frac{\rho_{me}^{eg}(\mathbf{r}_m) \rho_{nf}^{gf}(\mathbf{r}_n)}{|\mathbf{r}_m - \mathbf{r}_n|} \quad (3)$$

here  $\rho_{me}^{eg}(\mathbf{r}_n)$  and  $\rho_{nf}^{gf}(\mathbf{r}_n)$  are the transition charge densities.

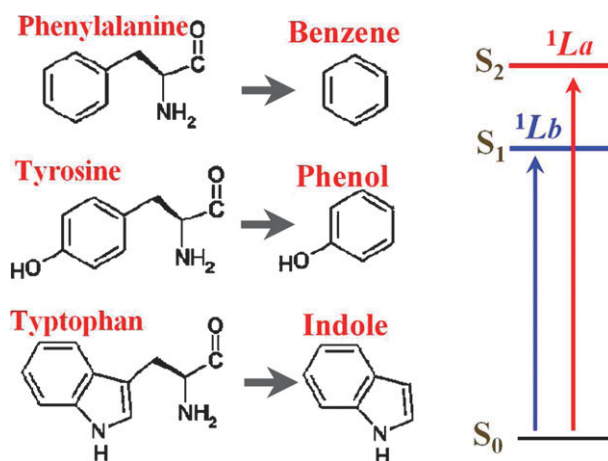
$\varepsilon_{me}$  and  $J_{me, nf}$  are calculated using our EHEF algorithm at a high *Ab initio* level, as described below. The electronic excitation Hamiltonian is then constructed with the computed parameters using the matrix method as implemented in the DichroCalc program.<sup>33,34</sup>

### 2.2 Electronic excitations of isolated aromatic chromophores

The NUV chromophores in the three natural aromatic amino acids, Phe, Tyr, and Trp, are benzene, phenol, and indole, respectively, as shown in Fig. 1. Each of these chromophores has four valence electronic excitations in the  $> 250$  nm region, namely  $^1L_b$ ,  $^1L_a$ ,  $^1B_a$ ,  $^1B_b$  in Platt's notation.<sup>35</sup> The lower states  $^1L_b$  and  $^1L_a$  are responsible for the main NUV peaks. The CASSCF method<sup>23</sup> has been successfully employed for electronic excitations of aromatic chromophores.<sup>36</sup> we have calculated the electronic states of isolated benzene, phenol and indole at the self-consistent reaction field (CASSCF/SCRF) and multi-configurational second-order perturbation theory (CASPT2)<sup>37</sup> level using the MOLCAS<sup>38</sup> package. These result in the excited state energies  $\varepsilon_{me,0}$  listed in Table 1, as well as the corresponding electronic densities ( $\rho_m^{ee}$  and  $\rho_m^{gg}$ ) and transition densities  $\rho_m^{eg}$ .  $g$  is the ground state and  $e$  runs over the  $^1L_b$ ,  $^1L_a$ ,  $^1B_a$ ,  $^1B_b$  states.

### 2.3 The EHEF (exciton Hamiltonian with electrostatic fluctuations) algorithm

As pointed out by Kurapkat *et al.*, interactions of chromophores with local electrostatic fields coming from the rest of the protein and the surrounding solvent shift the excited state energies.<sup>39</sup> The CASSCF calculations of isolated benzene, phenol, and indole must therefore be corrected to include these effects. A chromophore interacts with surroundings through



**Fig. 1** The three aromatic amino acids in proteins, phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp), contain the chromophores benzene, phenol, and indole, respectively. The  $^1L_b$  and  $^1L_a$  excitations dominate the NUV signals in the wavelength region above 250 nm.

**Table 1** Computed excited state energies  $\varepsilon_{me,0}$  for the isolated benzene, phenol and indole

$\varepsilon_{me,0}/\text{cm}^{-1}$	$^1L_b$	$^1L_a$	$^1B_a$	$^1B_b$
Benzene	38 005	47 953	52 324	52 531
Phenol	36 492	46 205	50 930	51 968
Indole	35 396	38 053	51 030	48 955

long range Coulombic interactions, which results in the inter- and intra-molecular electrostatic corrections in eqn (2).

The proteins are divided into fragments of amino acids and peptide bond units. Their ground states are obtained from the first principles calculations. The peptide units of the protein backbone are modeled by CASSCF/SCRF and CASPT2 calculations of an isolated *N*-methylacetamide (NMA) molecule. The gas-phase electronic structures and charge densities of amino acids and water molecules were obtained with the hybrid DFT B3LYP/6-311++G\*\* method implemented in the GAUSSIAN03 package.<sup>40</sup>

The direct implementation of eqn (2) and (3) is too expensive for the present applications. Since the chromophores interact with many surrounding atoms or molecules, evaluating the factor  $[\rho_m^{ee}(\mathbf{r}_m) - \rho_m^{gg}(\mathbf{r}_m)]\rho_k^{gg}(\mathbf{r}_k)$  or  $\rho_m^{eg}(\mathbf{r}_m)\rho_n^{gf}(\mathbf{r}_n)$  will require a huge number of integrations of four basis wavefunctions over space. This is one of the most serious bottlenecks in standard quantum chemistry calculations. Some empirical methods have been developed in order to reduce computational cost: (1) the dipole approximation,<sup>41,42</sup> which describes the environmental influences as the interaction between electric dipole of the chromophore and the electric field induced by surroundings. This has a limited accuracy. (2) the map method,<sup>9,21,22</sup> which uses empirically fitted parameters to calculate the excited state energy as a function of electric fields at some selected positions or some special geometric variables. It has a limited predictive power because of the fitting procedure.<sup>7</sup>

An alternative approach, free from empirical parameters, is to calculate charge densities by coarse-graining the electronic wavefunction in space. Krueger *et al.*<sup>43</sup> had developed the transition density cube (TDC) method, in which 3-D space is divided into many small volume elements. Intra- and inter-molecular interactions were calculated directly from the Coulomb interactions between charges of the cubes in each molecule. The excited state energy and resonant coupling are then given by

$$\varepsilon_{me} = \varepsilon_{me,0} + \sum_{m_\mu} \sum_{k_\nu}^{k \neq m} \frac{1}{4\pi\epsilon} \frac{(Q_{m_\mu}^{ee} - Q_{m_\mu}^{gg})Q_{k_\nu}^{gg}}{|\mathbf{r}_{m_\mu} - \mathbf{r}_{k_\nu}|} \quad (4)$$

and

$$J_{me,nf} = \sum_{m_\mu} \sum_{n_\nu}^{n \neq m} \frac{1}{4\pi\epsilon} \frac{Q_{m_\mu}^{eg} Q_{n_\nu}^{gf}}{|\mathbf{r}_{m_\mu} - \mathbf{r}_{n_\nu}|} \quad (5)$$

where the indices  $\mu(\nu)$  run over the cubes with permanent or transition charge  $Q_{m_\mu}$  ( $Q_{k_\nu}$  or  $Q_{n_\nu}$ ) at position  $\mathbf{r}_{m_\mu}$  ( $\mathbf{r}_{k_\nu}$  or  $\mathbf{r}_{n_\nu}$ ). This grid method is accurate, but very expensive, since it requires a large number of cubes to maintain the precision (normally  $\approx 500\,000$  cubes for a system with  $\approx 50$  atoms). Madjet *et al.*<sup>42</sup> have developed a more efficient way for taking the charge distribution into account. In their TrEsp (transition charge from electrostatic potential) code, electrostatic

potentials on sample points are computed at the quantum chemistry level. Partial charges are then assigned to atomic positions by fitting to potentials. TrEsp has been successfully employed in the study of photosynthetic complexes. However, charges distributed only at the atomic positions cannot represent the electronic cloud when the molecular orbitals are delocalized.

EHEF<sup>7</sup> offers a good balance of accuracy and cost by combining the advantages of the TDC and TrEsp methods. The algorithm consists of five steps:

1. We define an atomic frame  $A_{ij}(\mathbf{R}_{ij}, \theta, \gamma)$  for each pair of atoms  $A_i$  and  $A_j$  in the molecule, as shown in Fig. 2(a) and (b). Here  $\mathbf{R}_{ij} = \mathbf{r}_{A_i} - \mathbf{r}_{A_j}$  is the atom-atom vector, and the angle  $\theta$  and dihedral angle  $\gamma$  are relative to the whole molecule.

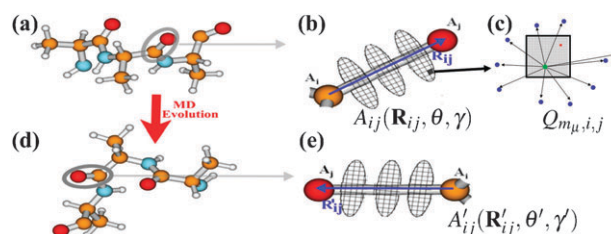
2. For each atomic frame, we construct a set of ( $\sim$ ten) planar disks perpendicular to  $\mathbf{R}_{ij}$ , whose centers lie on  $\mathbf{R}_{ij}$ , located in the range  $\pm 1.5|\mathbf{R}_{ij}|$ . Each disk is divided into a limited number of (typically 20 to 50) grids, as shown in Fig. 2(b). The grids are stored with internal coordinates relative to the atomic frame.

3. Each grid in the atomic frame is divided into a large number of (typically 1000) small sub-grids, as shown in Fig. 2(c). Using standard quantum chemistry methods, we compute the electronic charge by integrating charge density (both permanent and transition densities) over space.<sup>7</sup>  $q_{m_\mu,ij}$  represents the transition charges of sub-grid  $\mu$  in the  $A_{ij}$  atomic frame of the  $m$ th molecule, as shown in Fig. 2.

4. Random sample points are taken around each grid of the atomic frame in Fig. 2(c), and electrostatic potentials induced by the charges of small sub-grids are calculated. With the TrEsp technique,<sup>7,42</sup> the sample electrostatic potentials are used to fit out a fictitious charge  $Q_{m_\mu,ij}$  for the grid  $\mu$  in the  $A_{ij}$  atomic frame of the  $m$ th molecule.

5. The grid charges are finally used to compute the Coulomb interactions and electric dipoles. Eqn (4) and (5) are used to calculate the corrections of excitation energies and resonant couplings, respectively.

Most localized charges are included as atomic partial charges. Delocalized charges vary slowly in space, and only require a sparse grid (typically  $\approx 10\,000$  cubes for a system of 50 atoms), which is much smaller than that in the TDC method. The fictitious grid charges are fitted from the electrostatic potentials generated by a large number of sub-grids (normally  $\approx 10^7$  cubes for a system with  $\approx 50$  atoms), which retains the same accuracy as the TDC method. Since the



**Fig. 2** (a) Protein fragment with standard geometry. (b) The atomic frame  $A_{ij}(\mathbf{R}_{ij}, \theta, \gamma)$ . (c) The grid in the atomic frame is divided into a large number of small sub-grids. A fictitious point charge  $Q_{m_\mu,ij}$  is generated by fitting the sample electrostatic potentials induced by charges of sub-grids. (d) The fragment in a new MD snapshot. (e) Update the atomic frame to  $A'_{ij}(\mathbf{R}'_{ij}, \theta', \gamma')$ .



grid spacing is much smaller than interatomic distances, the computed excited state energies are as accurate as the full integration of eqn (2).

The utilization of the atomic frame in EHEF provides a convenient interface for QM and MM calculations. Proteins are not rigid and the interatomic distances and angles vary during the MD evolutions. We had introduced the following key approximation: we assume that the charge distributions in the atomic frame for each pair of atoms do not vary significantly during MM simulations, and can be held fixed. First principles (QM) calculations are carried out only on a limited number of standard geometries. As long as the chemical structure does not vary strongly, the charge grids computed for atomic frames of the standard geometries can be used for all MD snapshots. As shown in Fig. 2(d) and (e), the atomic frame  $A_{ij}(\mathbf{R}_{ij}, \theta, \gamma)$  of a standard geometry is updated to  $A'_{ij}(\mathbf{R}'_{ij}, \theta', \gamma')$  for a new MD snapshot. Since the charge grids of the standard geometry are stored with internal coordinates relative to the atomic frame, the update of the atomic frame automatically generates a new set of charge grids, which accurately reflect the electronic properties of the new MD snapshot. Using the charge distribution, we have calculated the interactions in eqn (4), and updated the excitation energies for each MD snapshot. Environmental fluctuations are taken into account at the QM level, with very few QM calculations.

## 2.4 2D signals

The Frenkel Hamiltonian matrix was diagonalized using the SPECTRON code.<sup>44</sup> 2D photon echo signals were calculated using the protocol described in ref. 20 and 45. We assume four short Gaussian laser pulses with central frequency  $37\,000\text{ cm}^{-1}$  (270 nm) and full width at half-maximum (FWHM)  $3754\text{ cm}^{-1}$  (corresponds to a Fourier limited pulse of 2.8 fs), which span the absorption band in the NUV region. We have used these broadband pulses in order to reveal most detailed information. Narrower pulses will miss some of the features. These could be recovered using two color measurements. The four pulses in chronological order have wavevectors  $\mathbf{k}_1$ ,  $\mathbf{k}_2$ ,  $\mathbf{k}_3$ , and  $\mathbf{k}_4$ , with  $\mathbf{k}_4 = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$ . The absorption change of the  $\mathbf{k}_4$  beam is recorded as a function of the three consecutive delay times:  $t_1$ ,  $t_2$ , and  $t_3$ . We have used the nonlinear exciton equations (NEE) approach. The scattering matrix of exciton quasi-particles is constructed as described in ref. 44. Calculations were performed for both the ordinary non chiral ( $xxxx$ ) and chirality-induced ( $xxxy$ ) polarization configurations. In the later, the  $\mathbf{k}_1$  pulse is polarized along the  $y$  axis and the other three pulses are  $x$  polarized. 2D signals are calculated by two-dimensional Fourier transform  $t_1 \rightarrow \Omega_1$  and  $t_3 \rightarrow \Omega_3$  with  $t_2$  set to zero.

2D spectra will be plotted using non-linear scale which reveals both the strong and weak signals,

$$\text{arcsinh}(cS) = \ln(cS + \sqrt{1 + c^2 S^2}), \quad (6)$$

The signal  $S$  is multiplied by a scale factor  $c$  to make it close to 1, so that weak amplitudes are amplified: for  $cS < 1$  the scale is linear,  $\text{arcsinh}(cS) \approx cS$ , and for larger  $cS$  it becomes logarithmic,  $\text{arcsinh}(cS) \approx (S|S|^{-1})\ln(2|cS|)$ .

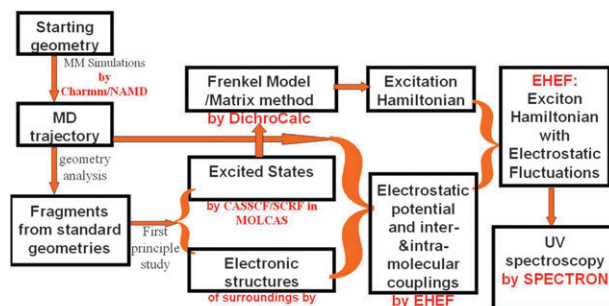


Fig. 3 Computation flow chart of 2DNUV simulations.

## 2.5 Computation flow chart

The computation flow chart is summarized in Fig. 3. Initial protein geometries are taken from experiments or the protein data bank. MM simulations of proteins in water are carried out with the CHARMM22 force field<sup>46</sup> using the software package NAMD.<sup>47</sup> Geometric variations along the MD trajectory were examined, and a set of standard geometries are selected. QM calculations are performed on the fragments (amino acids, peptide bond units, and water) at standard geometries, to yield the necessary information about ground and excited states. The EHEF algorithm is used to calculate the electrostatic potential, inter- and intra-molecular interactions. Based on the Frenkel model, the matrix method in DichroCalc program<sup>33,34</sup> uses the parameters from EHEF to construct effective exciton Hamiltonian with electrostatic fluctuations. UV spectra are finally simulated using the SPECTRON code.<sup>44</sup>

## 3. 2DNUV Spectra of $\beta$ -amyloid ( $A\beta_{9-40}$ ) fibrils

### 3.1 $\beta$ -Amyloid ( $A\beta_{9-40}$ ) fibrils

Tycko and co-workers had used NMR constraints to construct a detailed molecular model of the 32-residue  $\beta$ -amyloid ( $A\beta_{9-40}$ ),<sup>48</sup> which is characterized by cross- $\beta$  structural motifs. Fig. 4(a) shows the structure of a single  $A\beta_{9-40}$  peptide, an asymmetric “U” structure composed of two  $\beta$ -strands as N-terminal (red) and C-terminal (blue), and a turn (green) which links the strands.  $A\beta_{9-40}$  has three aromatic side chains: Tyr10, Phe19, Phe20, all on the N-terminal. Mutations of the native  $A\beta_{9-40}$  peptide were obtained with the VMD tools.<sup>49</sup>

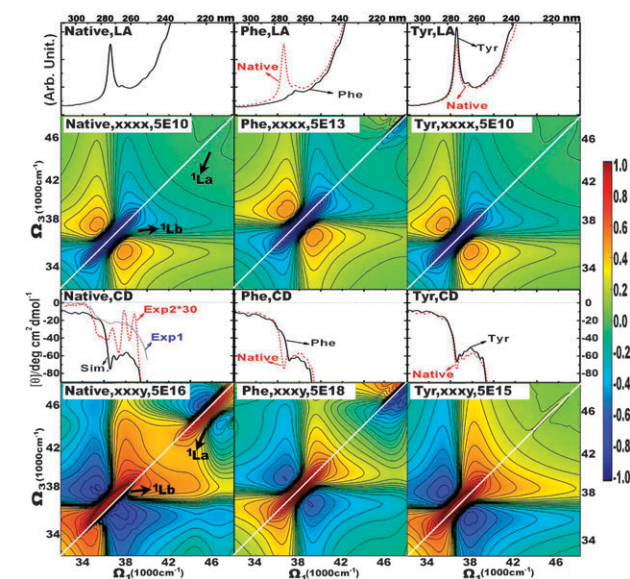
MM simulations were carried out on the STAG(+2) structure composed of twelve native or mutated  $A\beta_{9-40}$  monomers, as shown in Fig. 4(b). The two types of  $\beta$ -strands



Fig. 4 (a)  $A\beta_{1-40}$  sequence and the structure of a single  $A\beta_{9-40}$  peptide, with N-terminal strand (red), C-terminal strand (blue), and turn region (green). It has three aromatic side chains, Tyr10, Phe19, Phe20, all located on the N-terminal. (b) Model structure of amyloid fibril made of twelve  $A\beta_{9-40}$  molecules. Red and blue indicate two parallel  $\beta$ -sheets. (c) The model periodic fibril structure made of nine peptides used in the UV spectroscopy simulations.

form separate, parallel  $\beta$ -sheets in amyloid fibrils. We ran a 1 fs time-step 200 ps simulation of the amyloid fibrils in vacuum to obtain an initial periodic structure. Cubic periodic boundary conditions were used. Long-range electrostatic interactions were computed using the particle-mesh Ewald (PME)<sup>50</sup> approach and a real space 12 Å cutoff for nonbonded interactions. Langevin dynamics with a 1 ps damping coefficient were used to achieve a constant temperature. The amyloid fibrils were then inserted in a TIP3P water box<sup>51</sup> containing 9236 water molecules, which were not allowed to penetrate inside the fibrils. 36 chloride ions were added to neutralize the box. Using a Nosé-Hoover Langevin piston<sup>52</sup> with a decay period of 200 fs and a 100 fs damping time. The system was equilibrated at 1 atm constant pressure for 500 ps. After equilibration, we performed 5 ns NVT simulations at the room temperature.

Ensembles of MD geometric snapshots were harvested for the *Ab initio* studies. The aromatic side chains are located at the N-terminal (red in Fig. 4(b)), which is on the outer interface of the fibrils. We therefore focused on the upper stack of the fibrils, and neglected the influences from the bottom peptides. Since fibrils are composed of a large number of periodically arrayed peptides, we have only harvested the central A $\beta_{9-40}$  peptide of the upper stack from the MD ensembles. This peptide was repeated to build a set of fibrils with nine periodically arrayed peptides, as shown in Fig. 4(c). 1D (2D) UV simulations used 1000 (500) MD snapshots of this nine-peptide amyloid fibril. Altogether our simulations included 27 aromatic chromophores (nine peptides, each containing one Tyr and two Phe).



**Fig. 5** Left column: spectra of the native A $\beta_{9-40}$  fibril. From top to bottom: LA, non-chiral (xxxx) 2DNUV signal, CD, chiral (xxxy) 2DNUV signal. 1D and 2D signals are averaged over 1000 and 500 MD snapshots, respectively. Blue dotted and red dashed lines in the CD of the fibrils are experimental spectra exp1<sup>53</sup> and exp2 (multiplied by 30),<sup>54</sup> respectively. 2DNUV signals are displayed with a nonlinear scale (eqn (6)), the scale factor  $c$  is given on the top, red signals are positive and blue signals are negative, and a white diagonal line is drawn to reveal the symmetry of signals. 2DNUV scale bar is plotted at the right edge. Middle and right columns: same as the left column but only for the Phe and Tyr residues, respectively.

### 3.2 1D and 2DNUV spectra of the native fibril

The simulated 1D spectra (LA and CD) of the native A $\beta_{9-40}$  fibril are displayed in Fig. 5. The NUV LA spectrum is dominated by a strong absorption peak at around 36 500 cm<sup>-1</sup> (274 nm), which is ascribed to the <sup>1</sup>L<sub>b</sub> transition of the Tyr10 side chains. The simulated NUV CD spectrum is in good agreement with experiment. The absolute magnitude of simulated CD is close to that of the latest experimental CD spectrum (exp1).<sup>53</sup> The negative feature observed in exp1 from 35 000 cm<sup>-1</sup> (286 nm) to 40 000 cm<sup>-1</sup> (250 nm) is reproduced by the simulations. The CD peaks observed in exp2<sup>54</sup> are sharper than those in exp1 and in our simulated CD. Nevertheless, the primary negative CD valleys in exp2 at 36 800 cm<sup>-1</sup> (272 nm), 37 900 cm<sup>-1</sup> (264 nm), and 38 800 cm<sup>-1</sup> (258 nm) are well reproduced by the simulated ones at 36 900 cm<sup>-1</sup> (271 nm), 38 100 cm<sup>-1</sup> (262 nm), and 38 800 cm<sup>-1</sup> (258 nm), with less than 2 nm shifts.

The simulated 2DNUV spectra are displayed in the left column of Fig. 5. The non-chiral xxxx spectrum is symmetric with respect to the white diagonal line. Since the laser pulses are centered at 37 000 cm<sup>-1</sup>, the xxxx signals are dominated by the <sup>1</sup>L<sub>b</sub> transitions, which are displayed as a negative (blue) diagonal peak centered at 37 000 cm<sup>-1</sup> accompanied by two positive (red) side-bands. A relatively weak negative (blue) <sup>1</sup>L<sub>a</sub> diagonal peak appears at around 46 500 cm<sup>-1</sup>.

The chirality-induced xxxy signal shows a non-symmetric pattern shifted away from the diagonal line. The <sup>1</sup>L<sub>b</sub> and <sup>1</sup>L<sub>a</sub> transitions give the positive (red) diagonal peak with two negative (blue) side-bands centered at 37 000 cm<sup>-1</sup> and 46 500 cm<sup>-1</sup>, respectively. The negative side-band of <sup>1</sup>L<sub>b</sub> in the lower (upper) triangle region consists of one (two) negative peaks. That <sup>1</sup>L<sub>b</sub> side-band in the lower triangle region is very strong and it pushes the positive diagonal peak away from the diagonal line. For the <sup>1</sup>L<sub>a</sub> signals, the strong side-band in the upper triangle region shifts the diagonal peak away from the diagonal line. The two crosspeaks at  $\Omega_1 = 37\,000\text{ cm}^{-1}$ ,  $\Omega_3 = 46\,500\text{ cm}^{-1}$  and  $\Omega_1 = 46\,500\text{ cm}^{-1}$ ,  $\Omega_3 = 37\,000\text{ cm}^{-1}$ , are spectroscopic signatures of couplings between the <sup>1</sup>L<sub>b</sub> and <sup>1</sup>L<sub>a</sub> transitions.

Using our simulations, we can separate the contributions of different aromatic chromophores to the 2DNUV signals and thereby examine the signatures of couplings between chromophores. We have constructed an exciton Hamiltonian which contains transitions only from the Phe side chains (Phe19 and Phe20 in the A $\beta_{9-40}$  peptide), or the Tyr side chains (Tyr10 in the A $\beta_{9-40}$  peptide). The simulated LA, 2DNUV xxxx, CD and 2DNUV xxxy spectra of the Phe (Tyr) transitions are displayed in the middle (right) column in Fig. 5. These spectra are different from those predicted by the full Hamiltonian (left column). Compared to the full LA and CD (1D) spectra, the Phe LA and CD spectra miss the strong signals at around 36 500 cm<sup>-1</sup> (274 nm) and keep the signals from 37 000 cm<sup>-1</sup> (270 nm) to 40 000 cm<sup>-1</sup> (250 nm). Correspondingly, the Tyr LA and CD (1D) spectra show weaker signals from 37 000 cm<sup>-1</sup> (270 nm) to 40 000 cm<sup>-1</sup> (250 nm). The 2DNUV spectra also show prominent features associated with different transitions. Upon the neglect of Tyr transitions, the dominant <sup>1</sup>L<sub>b</sub> signal in the Phe 2DNUV xxxx and xxxy spectra blue-shifts for 1500 cm<sup>-1</sup>. We have used the



scale factor  $c$  in eqn (6) to normalize the strongest 2D signals to be 1. The scale factor is listed on the top of each 2DNUV spectrum. The scale factors show that the Phe 2DNUV  ${}^1L_b$   $xxxx$  and  $xxxy$  diagonal peaks are about 1000 and 100 times weaker than the full-transitions-induced ones, respectively. On the other hand, the Tyr  ${}^1L_b$   $xxxy$  signals become 10 times stronger. These suggest that the couplings between Phe and Tyr transitions suppress the Tyr transitions in the fibrils. Both the Phe and Tyr 2DNUV  $xxxy$  spectra are symmetric around the white diagonal line. The asymmetric pattern in the full spectrum thus reflects the strong couplings between Phe and Tyr transitions.

### 3.3 1D and 2DNUV spectra of mutated fibrils

The mutated E22G  $\beta$ -amyloid fibril has been extensively studied by Sian *et al.*<sup>54</sup> The mutated E22Q A $\beta_{9-40}$  fibrils was obtained by changing the native Glutamic (E) at residue 22 (E22) to Glutamine (Q) in all peptides. It was found to change conformation and fibrillize more rapidly and produce shorter and stubbier fibrils than the native form. In the middle column of Fig. 6, we display the simulated spectra. Good agreement is seen with the experimental CD (exp2)<sup>54</sup> of the E22Q fibril. The simulated CD spectrum shows two negative valleys at 37 200  $\text{cm}^{-1}$  (269 nm) and 38 400  $\text{cm}^{-1}$  (260 nm) corresponding to those at 37 000  $\text{cm}^{-1}$  (270 nm) and 38 000  $\text{cm}^{-1}$  (263 nm) in exp2. The native and E22Q fibrils have virtually identical LA spectra. But the chirality-induced CD spectra show clear differences. With the mutations from the native fibril to E22Q, the negative CD signals around 240–280 nm become much weaker, and some fine spectral features disappear.

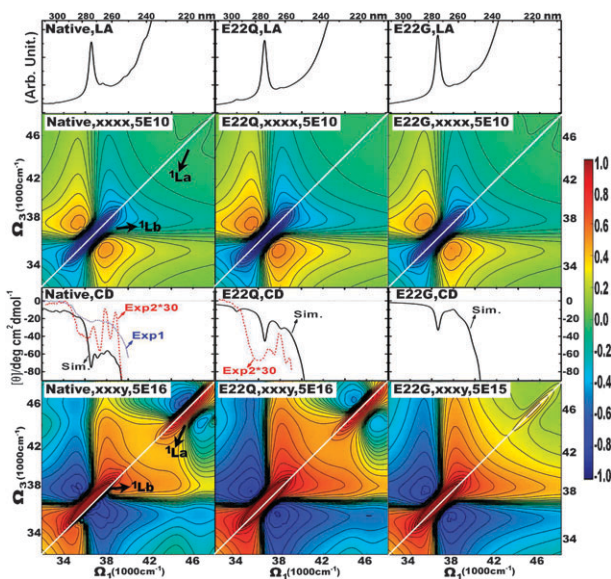
These 1DNUV spectra are highly congested, and differences due to local mutations are not clearly resolved. 2D photon echo signals depend on both singly and doubly excited electronic states, and thus contain richer spectral features associated with the geometric and electronic structures. In

Fig. 6 we compare the 2DNUV spectra of the native and mutated E22Q fibrils. The non-chiral  $xxxx$  spectra are similar for the native and E22Q fibrils. The chirality-induced  $xxxy$  signals, in contrast, show a sensitivity to the fibril mutations. This is illustrated by the simulated spectra of the native and E22Q fibril shown in the bottom row. In the E22Q fibril, the  ${}^1L_b$   $xxxy$  signal at around 37 000  $\text{cm}^{-1}$  is symmetric, with one peak locating at both side-bands of the dominant diagonal peak. While in the native fibril, the spectrum is non-symmetric, and the numbers of negative peaks in both side-bands of the dominant diagonal peak are different. As discussed above, the non-symmetric spectral pattern in the native fibril is resulted from the couplings between chromophores. Therefore, we believe that the E22Q mutation has affected the residue-residue interactions in the native fibril. Correspondingly, we observed weaker crosspeaks for couplings of  ${}^1L_b$  and  ${}^1L_a$  in E22Q than the native fibril, which also implies the change of residue-residue interactions due to mutations.

We have further studied the mutated E22G (the native E22 to G (Glycine)) A $\beta_{9-40}$  fibril. The spectra are displayed in the right column of Fig. 6. The LA spectrum is very similar to the E22Q fibril. Their CD spectrum are also similar, but E22G has less fine spectral features. 2DNUV  $xxxx$  and  $xxxy$  spectra of E22G have essentially the same spectra as E22Q, except the crosspeaks for couplings of  ${}^1L_b$  and  ${}^1L_a$  disappear in E22G, and both of the  ${}^1L_b$  and  ${}^1L_a$  induced diagonal peaks are strictly symmetric. This is because the residue-residue interactions in the E22G fibril is much weaker than those in the native and E22Q fibrils. Thus, the  ${}^1L_b$  transitions are less affected by couplings between chromophores and transitions. It is also the reason that the E22G mutation increases the intensity of the  ${}^1L_b$   $xxxy$  signal by about 10 times than the native and E22Q fibrils.

## 4. Conclusions

EHEF is a first principles simulation protocol for the 2DNUV spectra of proteins. It combines electronic structure computations of isolated protein fragments and solvent molecules, and gives accurate descriptions for aromatic transitions at the first principle level. It also provides an effective interface for QM and MM simulations, that avoids expensive repeated QM calculations and generates the fluctuating Hamiltonian at the QM level for all the MD snapshots. Simulations were carried out for the  $\beta$ -amyloid (A $\beta_{9-40}$ ) fibrils associated with the Alzheimer's disease. The simulated 1DNUV spectra of fibrils in water in the presence of electrostatic fluctuations show good agreements with experiment. 2DNUV spectra are shown to provide a sensitive probe of molecular interactions and local structural details in fibrils. The spectroscopic signatures of  ${}^1L_b$  and  ${}^1L_a$  transitions, and their couplings are identified. Analysis of contributions of different aromatic chromophores to 2DNUV signals demonstrates strong signatures of interactions between chromophores. The 2DNUV chiral  $xxxy$  signals associated with couplings between transitions can reflect the change of residue-residue interactions, so that are sensitive to mutations of residues in the A $\beta_{9-40}$  peptide.



**Fig. 6** Same as Fig. 5 but for the mutated E22Q (middle column) and E22G (right column). The spectra of the native fibril are repeated in the left column for comparison.

## Acknowledgements

We gratefully acknowledge the support of the National Institutes of Health (Grant GM059230), and the National Science Foundation (Grant CHE-0745892). We wish to thank Prof. Jonathan D. Hirst for sharing the DichroCalc code, and Prof. Yi Luo for his advice in developing the EHEF algorithm.

## References

- 1 D. Whitford, *Proteins Structure and Function*, Wiley, 2005.
- 2 D. Kern, E. Z. Eisenmesser and M. Wolf-Watz, *Methods Enzymol.*, 2005, **394**, 507–524.
- 3 A. A. Oskouei, O. Bram, A. Cannizzo, F. van Mourik, A. Tortschanoff and M. Chergui, *Chem. Phys.*, 2008, **350**, 104–110.
- 4 M. Chatteraj, B. A. King, G. U. Bublit and S. G. Boxer, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 8362–8367.
- 5 F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi and C. Dobson, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 3590–3594.
- 6 S. Brahm and J. Brahm, *J. Mol. Biol.*, 1980, **138**, 149–178.
- 7 J. Jiang, D. Abramavicius, B. M. Bulheller, J. D. Hirst and S. Mukamel, *J. Phys. Chem. B*, 2010, **114**, 8270–8277.
- 8 D. M. Rogers and J. D. Hirst, *Biochemistry*, 2004, **43**, 11092–11102.
- 9 W. Zhuang, T. Hayashi and S. Mukamel, *Angew. Chem., Int. Ed.*, 2009, **48**, 3750–3781.
- 10 S. Mukamel, Y. Tanimura and P. Hamm, *Acc. Chem. Res.*, 2009, **42**, 1207–1209.
- 11 Y. S. Kim, L. Liu, P. H. Axelsen and R. M. Hochstrasser, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 7720–7725.
- 12 H. S. Chung, Z. Ganim, K. C. Jones and A. Tokmakoff, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 14237–14242.
- 13 S.-H. Shim, R. Gupta, Y. L. Ling, D. B. Strasfeld, D. P. Raleigh and M. T. Zanni, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 6614–6619.
- 14 A. A. Oskouei, O. Bram, A. Cannizzo, F. van Mourik, A. Tortschanoff and M. Chergui, *J. Mol. Liq.*, 2008, **141**, 118–123.
- 15 C. Tseng, S. Matsika and T. Weinacht, *Opt. Express*, 2009, **17**, 18788–18793.
- 16 M. Beutler, M. Ghotbi, F. Noack, D. Brida, C. Manzoni and G. Cerullo, *Opt. Lett.*, 2009, **34**, 710–712.
- 17 *Ultrafast Phenomena XVI*, ed. P. Corkum, S. De Silvestri, K. Nelson, E. Riedle and R. Schoenlein, Springer, 2009.
- 18 P. Nuernberger, R. Selle, F. Langhojer, F. Dimler, S. Fechner, G. Gerber and T. Brixner, *J. Opt. A: Pure Appl. Opt.*, 2009, **11**, 085202.
- 19 U. Selig, *The 5th International Conference on Coherent Multidimensional Spectroscopy*, 2010.
- 20 D. Abramavicius, J. Jiang, B. M. Bulheller, J. D. Hirst and S. Mukamel, *J. Am. Chem. Soc.*, 2010, **132**, 7769–7775.
- 21 T. la Cour Jansen, A. G. Dijkstra, T. M. Watson, J. D. Hirst and J. Knoester, *J. Chem. Phys.*, 2006, **125**, 044312.
- 22 Y.-S. Lin, J. Shorb, P. Mukherjee, M. T. Zanni and J. L. Skinner, *J. Phys. Chem. B*, 2009, **113**, 592–602.
- 23 N. Yamamoto, T. Vreven, M. Robb, M. Frisch and H. Schlegel, *Chem. Phys. Lett.*, 1996, **250**, 373–378.
- 24 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 25 R. Tycko, *Curr. Opin. Struct. Biol.*, 2004, **14**, 96–103.
- 26 N. E. Lester-Coll, E. J. Riviera, S. J. Soscia, K. Doiron, J. R. Wands and S. M. de la Monte, *J. Alzheimers Dis.*, 2006, **9**, 13–33.
- 27 F. Chiti and C. M. Dobson, *Nat. Chem. Biol.*, 2009, **5**, 15–22.
- 28 N. L. Fawzi, E.-H. Yap, Y. Okabe, K. Kohlstedt, S. Brown and T. Head-Gordon, *Acc. Chem. Res.*, 2008, **41**, 1037–1047.
- 29 Y. Frenkel, *Phys. Rev.*, 1931, **37**, 17–44.
- 30 D. Abramavicius, B. Palmieri and S. Mukamel, *Chem. Phys.*, 2009, **357**, 79–84.
- 31 S. Schenkl, F. van Mourik, G. van der Zwan, S. Haacke and M. Chergui, *Science*, 2005, **309**, 917–920.
- 32 J. Leonard, E. Portuondo-Campa, A. Cannizzo, F. van Mourik, G. van der Zwan, J. Tittor, S. Haacke and M. Chergui, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 7718–7723.
- 33 J. D. Hirst, *J. Chem. Phys.*, 1998, **109**, 782–788.
- 34 B. M. Bulheller, A. Rodger and J. D. Hirst, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2020–2035.
- 35 J. R. Platt, *J. Chem. Phys.*, 1949, **17**, 484–495.
- 36 D. M. Rogers and J. D. Hirst, *J. Phys. Chem. A*, 2003, **107**, 11191–11200.
- 37 J. Finley, P. A. Malmqvist, B. O. Roos and L. Serrano-Andres, *Chem. Phys. Lett.*, 1998, **288**, 299–306.
- 38 G. Karlstrom, R. Lindh, P. Malmqvist, B. Roos, U. Ryde, V. Veryazov, P. Widmark, M. Cossi, B. Schimmelpennig, P. Neogrady and L. Seijo, *Comput. Mater. Sci.*, 2003, **28**, 222–239.
- 39 G. Kurapkat, P. Kruger, A. Wolimer, J. Fleischhauer, B. Kramer, A. Zobel, A. Koslowski, H. Botterweck and R. W. Woody, *Biopolymers*, 1997, **41**, 267–287.
- 40 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03 (Revision C.02)*, Gaussian, Inc., Wallingford, CT, 2004.
- 41 H. A. Y. Luo and P. Norman, *J. Chem. Phys.*, 1998, **109**, 3589–3595.
- 42 M. E. Madjet, A. Abdurahaman and T. Renger, *J. Phys. Chem. B*, 2006, **110**, 17268–17281.
- 43 B. P. Krueger, G. D. Scholes and G. R. Fleming, *J. Phys. Chem. B*, 1998, **102**, 9603–9604.
- 44 D. Abramavicius, B. Palmieri, D. V. Voronine, F. Šanda and S. Mukamel, *Chem. Rev.*, 2009, **109**, 2350–2408.
- 45 J. Jiang, D. Abramavicius, C. Falvo, B. M. Bulheller, J. D. Hirst and S. Mukamel, *J. Phys. Chem. B*, 2010, **114**, 12150–12156.
- 46 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, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. Reiher, III, B. Roux, M. Schlenkerich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
- 47 J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. KalÉ and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.
- 48 A. T. Petkova, W. M. Yau and R. Tycko, *Biochemistry*, 2006, **45**, 498–512.
- 49 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
- 50 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
- 51 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 52 S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *J. Chem. Phys.*, 1995, **103**, 4613–4621.
- 53 C.-H. Tang, Y.-H. Zhang, Q.-B. Wen and Q. R. Huang, *J. Agric. Food Chem.*, 2010, **58**, 8061.
- 54 A. K. Sian, E. R. Frears, O. M. A. EL-Agnaf, B. P. Patel, M. F. Manca, G. Siligardi, R. Hussain and B. M. Austen, *Biochem. J.*, 2000, **349**, 299–308.