

Simulation of two-dimensional infrared spectroscopy of amyloid fibrils

Wei Zhuang, Darius Abramavicius, Dimitrii V. Voronine, and Shaul Mukamel*

Department of Chemistry, University of California, Irvine, CA 92697

Edited by Robin M. Hochstrasser, University of Pennsylvania, Philadelphia, PA, and approved May 31, 2007 (received for review January 15, 2007)

We propose to use infrared coherent two-dimensional correlation spectroscopy (2DCS) to characterize the fibril structure of A β 42, the dominant composition of Ab deposit, which is crucial for investigating its toxicity and aggregation mechanism. By optimizing the pulse polarization configurations with a genetic algorithm combined with sensitivity analysis, we obtained signals with well resolved cross-peak features attributed to the couplings within and between different structural motifs. These signals may provide new constraints for refining of the currently available NMR structure. Two-dimensional correlation spectroscopy also can differentiate the turn structure of A β 42 and other Ab derivatives.

correlation spectroscopy | four wave mixing

The accumulation of amyloid deposits (1, 2), misfolded peptide aggregates whose dominant component is a 39- to 43-residue A β peptide (3), has been identified as a major feature of the pathogenesis of Alzheimer's disease (AD) (4). Despite their identical residues 1–39 sequence, the various A β peptides have significantly different biochemical properties: The 42-residue derivative, A β 42, has a much stronger tendency to form fibrils *in vitro* (5) as compared with other derivatives. A β 42 also is slightly more hydrophobic compared with shorter analogs because of the additional more-hydrophobic residues at the end of the peptide strand. More importantly, the protease resistance of A β 42 is drastically higher from its analogs (6). The structural basis of these property differences is still unknown. Because of the fibril's noncrystalline, insoluble, and mesoscopically heterogeneous nature, NMR is the primary tool for fibril structure determination (7–10). It provides various structural constraints that, when combined with computational tools, such as geometry optimization and MD simulation, yield the current structural models.

The most recent model of A β 42 structure was proposed by Riek (7) (denoted M42). M42 can be dissected into three motifs; residues 1–16 are randomly coiled, and residues 26–31 are the turn, and the rest form two β -strands. NMR structural information (7, 10) is primarily related to the β -strand. No information is available on the highly irregular coil segment. Because of the lack of structural constraints, the turn structure in this model is obtained by geometry optimization and depends heavily on the computational protocol and the empirical force field.

The present simulation of two-dimensional correlation spectroscopy (2DCS) signals of A β 42 demonstrates that this technique can provide additional constraints for refining the structure and distinguishing the local structures among the models provided by various theoretical protocols consistent with the same NMR experimental constraints. We consider the photon echo technique whereby the fibril is subjected to three very short (≈ 40 fs) infrared pulses propagating along the directions k_1 , k_2 , and k_3 , and the coherent signal is generated in the direction $k_x = -k_1 + k_2 + k_3$. Correlation plots are obtained by a double Fourier transform of the signal with respect to the delay between pulses 1 and 2 and between pulse 3 and the signal. All pulses are resonant with the amide-I band of the peptide bond. Diagonal peaks in these 2DCS signals resemble the linear absorption, whereas the off-diagonal cross-peaks carry valuable signatures of

intermode couplings. The spectral line shapes are indicators of solvent and conformational fluctuations (11–15), and if these excitations are localized within different segments, as is the case for NMR, the cross-peaks may be directly related to distances and dihedral angles. The cross-peak positions and line shapes therefore provide new structural constraints.

The amide-I vibrational mode of fibrils originates from the stretching motion of the C=O bond (coupled to in-phase N—H bending and C—H stretching) (16) and can be used to monitor secondary structure variations (17). The simulated linear absorption of M42 (Fig. 1 *Left*, Abs) shows an intense $\approx 1,635$ cm^{-1} peak (peak a), an $\approx 1,655$ cm^{-1} shoulder (peak b), an $\approx 1,675$ cm^{-1} peak, two additional peaks at $\approx 1,695$ cm^{-1} (peak d) and $\approx 1,715$ cm^{-1} (peak i). Fig. 1 *Left* (NMD) shows the decomposition of the various normal modes into the three structural motifs (β -sheet, turn, and coil). We see that peaks a, b, and c have strong contributions from both β -strand and coil. Peak d has a contribution from turn plus coil, and peak i is purely turn. Fig. 1 *Left* (2D), displays 2DCS signals simulated by assuming that all laser beams and the signal are linearly polarized in the same direction (*xxxx*). The signal is dominated by strong and broad diagonal peaks that resemble the linear absorption, the contributions of all three structural motifs overlap, and no cross-peaks are observed.

The spectral resolution of vibrational spectra of complex molecular assemblies, such as the amyloid fibrils, is much lower compared with NMR. Furthermore, because of the delocalized nature of the normal modes, the infrared peaks usually have contributions from different segments, which complicates their assignment. The participation ratio (PR) (18) provides a convenient measure for the degree of the vibrational eigenstate localization and can be used to assign the spectral features and relate them to the physical properties of the system. The PR of an eigenstate e whose wavefunction is a superposition of localized states on the m th amide unit with coefficients Ψ_{em} is defined as $W_e = (\sum_m |\Psi_{em}|^4)^{-1}$. It interpolates between $W_e = 1$ for a localized state and $W_e = n$ when e is completely delocalized, and Ψ_{em} has equal $1/\sqrt{n}$ contributions from all of the n local modes. The uppermost in Fig. 2 shows the PR distribution (PRD) of eigenstates of M42 in a given frequency range (a 10- cm^{-1} binning was used). The eigenstates at the two band edges are highly localized, as evident from their low PR. These modes are dominated by the coil and the turn segment, as demonstrated by the normal mode decomposition (NMD), whereas the mid-band modes have more β -strand character. Amide I vibrations in the

Author contributions: W.Z. and S.M. designed research; W.Z. and D.V.V. performed research; W.Z. contributed new reagents/analytic tools; W.Z. and D.A. analyzed data; and W.Z. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: AD, Alzheimer's disease; 2DCS, two-dimensional correlation spectroscopy; SAS, sensitivity analysis signal.

*To whom correspondence should be addressed. E-mail: smukamel@uci.edu.

© 2007 by The National Academy of Sciences of the USA

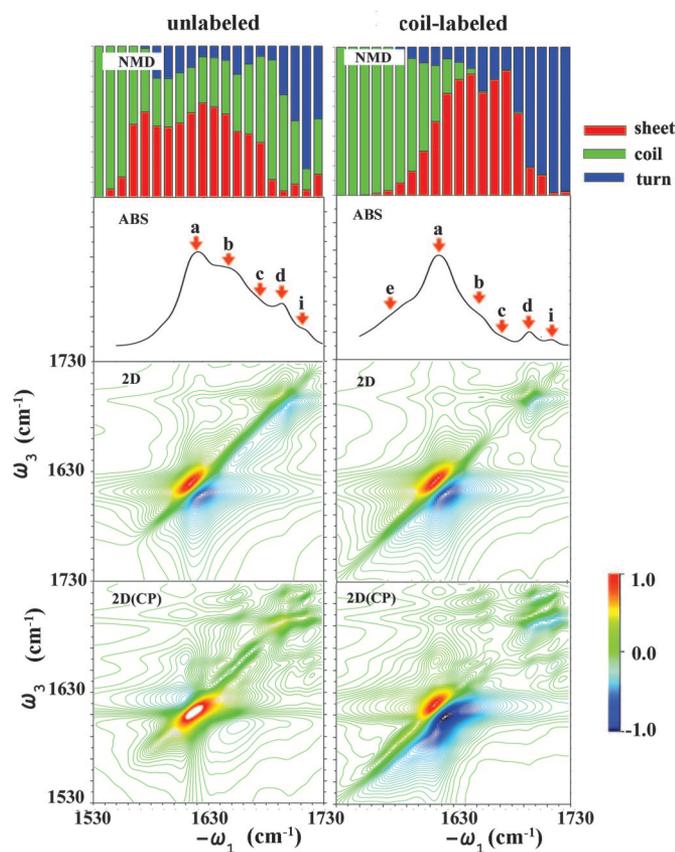


Fig. 1. Starting at the top, shown are the NMD diagram, the absorption signal (Abs), the $xyxy$ polarization 2DCS (2D), and the coherent-control-optimized-polarization [2D(CP)] 2DCS of unlabeled (*Left*) and coil-labeled (*Right*) M42. In NMD, the β -strand, coil, and turn content are shown in red, green, and blue, respectively.

coil segments are thus much more localized, which is to be expected based on their highly irregular structure.

For a more direct comparison of the coherence lengths in different motifs, we have applied the sensitivity analysis technique for the PRD in Fig. 2. Details of the protocol are given in ref. 19. The basic idea is to introduce a small perturbation to the energies of a specific segment, e.g., β -strand, in the Hamiltonian. The sensitivity analysis signal (SAS) given by the difference between the PRD of the perturbed and unperturbed system, provides a qualitative measure of the degree of localization of states belonging to that segment. Fig. 2 gives the SAS of M42 for the β -sheet, the coil and the turn segments. The arrows mark the strongest peaks, which have the PR value of ≈ 4 for the β -strand, ≈ 2 for the coil, and ≈ 8 for the turn, indicating that the β -strand has a longer coherence length.

The lower resolution and normal mode delocalization complicate the interpretation of the cross peak compared with NMR. However, isotope labeling and a judicious design of polarization configurations can be used to manipulate the 2DCS signals by enhancing desired spectral features. $^{13}\text{C}^{18}\text{O}$ isotope labeling of a given peptide residue can induce a 65-cm^{-1} red shift of the amide-I vibrational frequency, creating peaks well separated from the unlabeled band and providing structural information on desired segments (17). Two-dimensional signals depend on interferences among many contributions (Liouville space pathways) (20). This interference may be controlled by varying the relative polarizations of the various beams, thereby eliminating diagonal peaks and amplifying the cross-peaks.

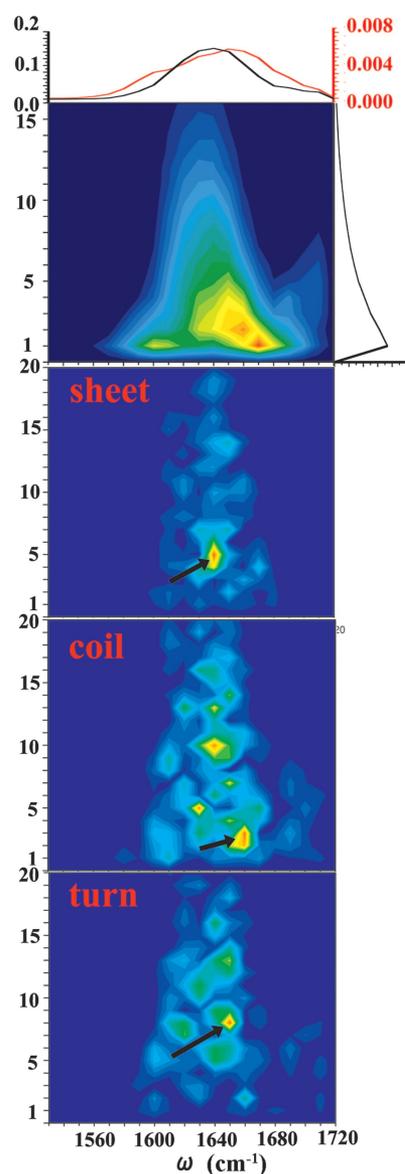


Fig. 2. The PRD binned over frequency and the SAS for sheet, coil, and turn for unlabeled M42. The top marginals are frequency-dependent average PR (black) and the variance of PR (red), and the right marginals are PRD integrated over the frequencies. The y axis gives the magnitude of PR, and the black arrows point to the strongest peak in SAS of PR.

We have constructed linear combinations of 2DCS signals corresponding to different pulse polarization configurations. A genetic algorithm (21) was applied to search for the best combination that suppresses the diagonal $1,655\text{-cm}^{-1}$ peak. This results in a much richer cross peak pattern in the signal [Fig. 1 *Left*, 2D(CP)] compared with the noncontrolled xxx signal (Fig. 1 *Left*, 2D). We have projected the normal modes contributing to the diagonal peaks onto the local amide modes along the backbone to assign the cross-peaks to positions along the structure. The *i* modes (Fig. 3 *Left*) are dominantly localized within the turn segment, with residue 28 making the largest weight, whereas the *d* modes (Fig. 3 *Right*) are almost evenly spread among the coil and the 28–30 residues of the turn. Considering the large distance between the coil and the turn (see Fig. 3), their interaction should be negligible. We therefore conclude that these two cross-peaks reflect turn–turn interactions within the 28–30 residues.

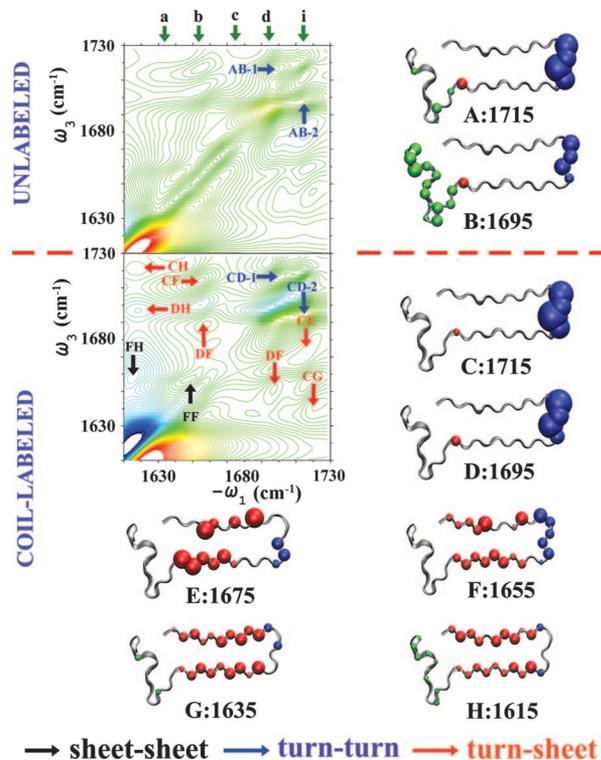


Fig. 3. The 2DCS signal of unlabeled (above the dashed line) and coil-labeled (below the dashed line) M42 with optimized polarization configuration (Figs. 1 and 3) on an expanded scale (1,630–1,730 cm^{-1}) and the projection of the normal modes contributing to the specified cross-peaks onto the local amide modes along the backbone. The volume of the balls represent the amplitude of the local mode population. The population is an average over a given residue on each layer. In the 2DCS plot, blue arrows denote the cross-peaks attributed to the turn–turn interaction, red arrows denote the turn–sheet interactions, and black arrows denote the sheet–sheet interactions. In the normal mode projection plots, the turn, sheet, and coil contributions are shown in blue, red, and green, respectively. Green arrows above the 2DCS denote absorption peak positions.

Most peaks in the M42 spectra have significant contributions from more than one structural motif and therefore may not be unambiguously assigned. Upon isotope-labeling of the coil segment (residues 1–16), the peaks will be dominated by 1 structural motif, as is evident from the NMD shown in Fig. 1. The new shoulder *e* in the linear absorption (Fig. 1, *Right Abs*) is dominated by the coil segment. Peaks *a*, *b*, and *c* are dominated by the sheet and peaks *d* and *i* belong to the turn. The resulting 2DCS (Fig. 1 *Right*, 2D) has a better cross-peak resolution than in the unlabeled sample, but the main cross-peak pattern is still unresolved.

The resolution is greatly improved by applying a genetic algorithm aimed at eliminating the 1,655- cm^{-1} diagonal peak of the coil-labeled fibril. Most cross-peaks may be clearly assigned in the resulting signal displayed in Fig. 1 *Right* [2D(CP)]. In Fig. 3, C:1715 and D:1695 demonstrate that, for the coil-labeled sample, peaks *d* and *i* are dominated by the turn, the cross-peaks CD1 (1,695, 1,715) and CD2 (1,715, 1,695) are thus related to turn–turn interactions. In Fig. 3, H:1615 and G:1635 show that the 1,615- and 1,635- cm^{-1} frequency windows are dominated by the β -strand motif. The CH, DH, and CG cross-peaks therefore originate from interactions between the turn and the sheet motifs close to the turn segment (mainly residues 24 and 25 and residues 32 and 33). The normal modes in the 1,675- cm^{-1} window (Fig. 3, E:1675) also are dominated by the sheet motif, the local mode population is rather nonuniformly

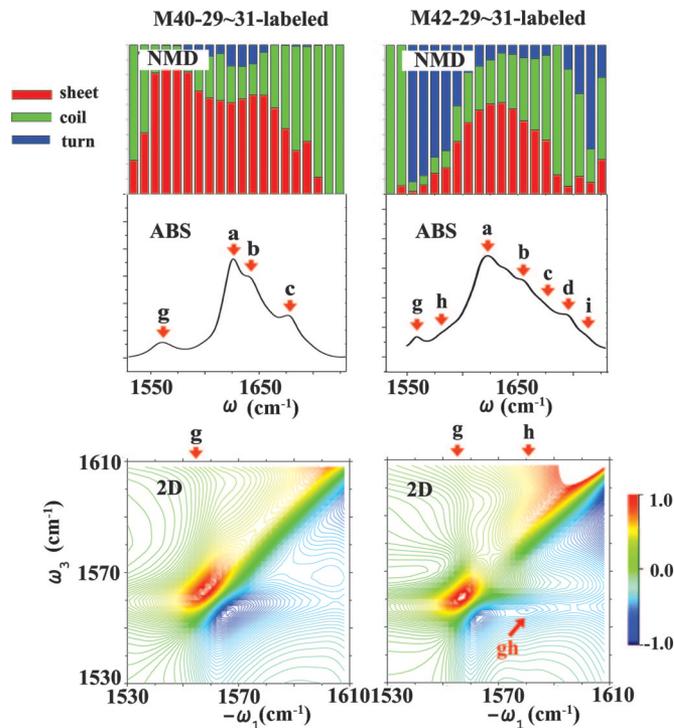


Fig. 4. Starting at the top, shown are the NMD, the absorption signal (ABS), and the 2DCS signal for M40 (*Left*), with xxxx polarization configuration in an expanded scale (1530–1610), and M42 (*Right*).

distributed, and there is no contribution from mode 25. The CE cross-peak therefore originates primarily from the interaction between the turn and residue 32. The normal modes at $\approx 1,655 \text{ cm}^{-1}$ (Fig. 3, F:1655), however, have a significant contribution from both the sheet and the turn; therefore, the CF, DF1, and DF2 peaks should carry information about turn–turn and turn–sheet interactions. The additional cross-peaks, FF and FH (marked by black arrows), are related to sheet–sheet interactions.

2DCS may also help differentiate between local structures in different fibril models. Residues 29–31, for instance, belong to the turn structure in M42 but to the β -strand in a structure model of A β 40 proposed by Tycko (M40). The simulated absorption spectra with these residues isotopically labeled are shown in Fig. 4, ABS. The signals have a same new intense peak, *g*, at $\approx 1,560 \text{ cm}^{-1}$. The corresponding labeled 2DCS signals depicted in Fig. 4, 2D and M42, show a cross-peak *gh* at (–1,585, 1,560), which is absent in M40. The NMD shown in Fig. 4 indicates that $\approx 60\%$ of the content of M40 in this frequency window belongs to the 29–31 segment (primarily mode 31); the remaining 40% is distributed among many other modes. This distribution may explain the absence of distinct cross-peaks. As in the absorption, the cross-peak features of various contents overlap and form a broad shoulder covering all significant peak features. For M42, this peak has a dominant ($\approx 90\%$) turn component, particularly residue 31 ($\approx 73\%$). Similar analysis shows that $\approx 85\%$ of the 29–31 segment is on the eigenmodes between (1,575, 1,585) for M42. The (1,585, 1,560) cross-peak is, thus, directly related to interactions within residues 29–31 of the turn. In addition to providing distinct signatures of local structure variations, the cross peaks also reveal differences in the coupling pattern within this local structure and provide a distinct spectral window for investigating this coupling pattern, which is not available from the absorption spectrum.

This study had focused on the peptide backbone structure. However, there are significant differences among the side chains for the different structures (7). These differences can be investigated by measuring other vibrational bands and their cross-peaks with the amide I band. One notable advantage of 2DCS is that it should allow us to monitor the primary steps in the fibril aggregation process (22) of selectively labeled Ab peptides with high (50-fs) temporal resolution. Light scattering, when the aggregate size is comparable to the optical wavelength, may complicate these measurements. This problem could be overcome by looking at oligomers rather than fibrils.

Finally, there are several other pulse polarization configurations that are chirality-specific, i.e., they vanish for nonchiral systems (23). Such signals are the nonlinear analogues of circular dichroism. They are one to three orders of magnitude weaker but could potentially provide additional fine detail about the structure. Exploring these signals will be an interesting direction for the future.

Methods

The Protein Data Bank file of M40 was kindly provided by Robert Tycko (National Institutes of Health, Bethesda, MD), M42 was obtained from Protein Data Bank, and the initial tail structures were added manually. Five layers of strand were stacked along the z axis for the periodic boundary box. One of the differences between M40 and M42 is that M42 comprises only one monomer per unit length instead of two to four in M40 (6). Although each fibril in M42 consists of four protofilaments so that a plane perpendicular to the fibril axis will have four monomers, the distance between the two neighboring monomer on the plane is 44 Å, and the electrostatic interaction between them can be negligible so that the IR signal will be identical to the single protofilament. Therefore, we used one monomer for each layer in M42, and we use two monomers for M40.

The MD simulation was carried out by using the NAMD package (24). The simulated annealing technique was used to generate initial structures with heterogeneous tail configurations. We fixed the nontail structure and raised the temperature of the system from 0 K to 3,000 K with 5-K increments. A 5,000-step equilibration run was carried out with a 2-fs time step after every increase in temperature. A 5-ns equilibration was then carried out at 3,000 K, followed by a 5-ns dynamics run, during which 20 structures were sampled for the cooling. The cooling procedure was then performed for all of the 20 configurations, with the temperature of the system from 3,000 K to 273 K with 5-K decrements and 5,000 steps of equilibration run after

every decrease in temperature to harvest the 20 initial configurations for the later simulation. A 10-ns equilibration was then carried out for all of the 20 initial structures with fixed C α atoms along the peptide backbone of the nontail part. Finally a 10-ns dynamics run with fixed C α atoms along the nontail part was performed, and 3,000 snapshots were used for the simulation of the two-dimensional infrared signal by the SPECTRON package (25). The truncated NEE technique (25) with a Lorentzian line-shape with FWHM at 5.5 cm $^{-1}$ for all excitons. $h = 0.3$ was used to simulate the signal. $^{13}\text{C}^{18}\text{O}$ isotope labeling was simulated by -65 cm^{-1} red-shifting of the local model frequency.

The NMD was calculated by summing the squares of its expansion coefficients $|Y_{em}|^2$ for all of the local modes belonging to a given structural motif. The components were then binned for all eigenmodes in the frequency window ($\pm 5\text{ cm}^{-1}$) to obtain the NMD figures. The projection of normal mode on backbone (Fig. 4, peaks a–h) was similarly obtained by calculating the squares of its expansion coefficients for each local mode in the local basis, averaged over all of the five layers, then binned by the mode frequencies.

For the two-dimensional coherent-control polarization-optimized signals, we constructed the following superposition of the three linearly independent tensor components $T_j = xxxy, xyxy, xyxx$:

$$W(\omega_1, \omega_3) = \sum_j c_j T_j.$$

The complex coefficients c_j were optimized by using a genetic algorithm aimed at minimizing the control target taken to be the ratio of the integrated diagonal line in the absolute magnitude of the two-dimensional spectrum to the integrated diagonal peak at 1,655 cm $^{-1}$, with $\delta = 10\text{ cm}^{-1}$. Fast exponential convergence was achieved by using 10 members in a population within 100–200 generations. The noise reduction parameter (21) was $\varepsilon = 0.6$ –0.7, and the parameters of the selection were $L_i = 5$ and $L_f = 6$. The optimized coefficients for M42 in Fig. 1 *Left*, were $c_1 = -0.34 - i \times 0.56$, $c_2 = -0.044 - i \times 0.54$, and $c_3 = 0.34 - i \times 0.40$. For M42 in Fig. 1 *Right*, the optimized coefficients were $c_1 = 0.57 + i \times 0.28$, $c_2 = 0.04 + i \times 0.22$, and $c_3 = -0.59 - i \times 0.44$.

We thank Dr. R. Tycko for sharing structural data and Dr. R. Riek and Prof. Charles Glabe for helpful discussions. This work was supported by the National Institutes of Health Grant GM59230 and National Science Foundation Grant CHE-0446555.

1. Glenner GG (1980) *N Engl J Med* 302:1283–1292.
2. Selkoe DJ (2003) *Nature* 426:900–904.
3. Findeis MA (2000) *Biochim Biophys Acta* 1502:76–84.
4. Kang J, Lemaire HG, Unterbeck A, Salbaum JM, Masters CL, Grzeschik KH, Multhaup G, Beyreuther K, Mullerhill B (1987) *Nature* 325:733–736.
5. Riek R, Guntert P, Dobeli H, Wipf B, Wuthrich K (2001) *Eur J Biochem* 268:5930–5936, 8237–8245.
6. Knauer MF, Soreghan B, Burdick D, Kosmoski J, Glabe CG (1992) *Proc Natl Acad Sci USA* 89:7437–7441.
7. Luhrs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, Doeli H, Schubert D, Riek R (2005) *Proc Natl Acad Sci USA* 102:17342–17347.
8. Tycko R (2000) *Curr Opin Chem Biol* 4:500–506.
9. Petkova AT, Ishii Y, Balbach JJ, Antzutkin ON, Leapman RD, Delaglio F, Tycko R (2002) *Proc Natl Acad Sci USA* 99:16742–16747.
10. Petkova AT, Yau WM, Tycko R (2006) *Biochemistry* 45:498–512.
11. Tanimura Y, Mukamel S (1993) *J Chem Phys* 99:9496–9511.
12. Zheng JR, Kwak K, Asbury J, Chen X, Piletic IR, Fayer MD (2005) *Science* 309:1338–1343.
13. Asplund MC, Zanni MT, Hochstrasser RM (2000) *Proc Natl Acad Sci USA* 97:8219–8224, 15920.
14. Demirdoven N, Cheatum CM, Chung H, Khalil S, Knoester M, Tokmakoff J (2004) *J Am Chem Soc* 126:7981–7990.
15. Hamm P, Lim MH, Hochstrasser RM (1998) *J Phys Chem B* 102:6123–6138.
16. Mantsch HH, Chapman D (1996) *Infrared Spectroscopy of Biomolecules* (Wiley-Liss, New York).
17. Wang JP, Chen JX, Hochstrasser RM (2006) *J Phys Chem B* 110:7545–7555.
18. Thouless D (1974) *Phys Rep* 13:93–106.
19. Zhuang W, Abramavicius D, Mukamel S (2005) *Proc Natl Acad Sci USA* 102:7443–7448.
20. Mukamel S (1995) *Principle of Nonlinear Spectroscopy* (Oxford Univ Press, New York).
21. Voronine D, Abramavicius D, Mukamel S (2006) *J Chem Phys* 123:034104–034115.
22. Hwang W, Zhang SG, Kamm RD, Karplus M (2004) *Proc Natl Acad Sci USA* 101:12916–12921.
23. Zhuang W, Abramavicius D, Mukamel S (2006) *Proc Natl Acad Sci USA* 103:18934–18938.
24. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) *J Comp Phys* 151:283–312.
25. Zhuang W, Abramavicius D, Hayashi T, Mukamel S (2006) *J Phys Chem B* 110:3362–3374.