Supporting Information:

A Free Energy Barrier Caused by the Refolding of an Oligomeric Intermediate Controls the Lag Time of Amyloid Formation by hIAPP

Arnaldo L. Serrano,[‖,†] Justin P. Lomont,[‖,†] Ling-Hsien Tu,[‡] Daniel P. Raleigh,[‡] Martin T. Zanni*[,†]

[†]Department of Chemistry, University of Wisconsin-Madison, Madison WI 53706.

[‡]Department of Chemistry, Stony Brook University, Stony Brook, NY 11790.

SI Materials and Methods

Synthesis and Purification of Human Islet Amyloid Polypeptide

Peptide synthesis was performed on a CEM Liberty microwave peptide synthesizer on a 0.1 mmol scale using 9-fluornylmethoxycarbonyl (Fmoc) chemistry. Solvents used were ACS-grade. 5-(4'-fmoc-aminomethyl-3', 5-dimethoxyphenol) valeric acid (PAL-PEG-PS) resin was used to form an amidated C-terminus. Standard Fmoc reaction cycles were used. Fmoc protected pseudoproline dipeptide derivatives were incorporated at positions 9–10 and 19–20 to facilitate the synthesis.[1,2] The β-branched residues, arginine, and all pseudoproline dipeptide derivatives were double coupled. C^{13}=O^{18} isotope labeled Phe (F), Gly (G), Ala (A), Ile (I) and Leu (L) residues were prepared from the C1 C^{13} labeled amino acids (Cambridge Isotope Laboratories) by incorporating O^{18} via oxygen exchange from H_2O^{18} as reported previously.[3] A maximum temperature of 50°C was used for the coupling of His and Cys to reduce racemization. Peptides were cleaved from the resin by standard trifluoroacetic acid (TFA) methods. Crude peptides were partially dissolved in 20% acetic acid (v/v), frozen in liquid nitrogen, and lyophilized to increase their solubility. The dry peptide was redissolved in dimethyl sulfoxide at room temperature to promote the formation of the disulfide bond between residues 2 and 7. Purification was carried out by reverse-phase high-performance liquid chromatography using a Vydac or Proto 300 C18 preparative column (10 mm 250 mm). A two buffer gradient was used: buffer A consisted of 100% H_2O and 0.045% HCl (v/v) and buffer B included 80% acetonitrile, 20% H_2O, and 0.045% HCl. HCl was used as the counterion instead of TFA because residual TFA can influence amyloid formation. Matrix-assisted laser desorption/ionization time of flight mass spectrometry confirmed the correct molecular mass.

Sample Preparation

Lyophilized hIAPP was dissolved in deuterated hexafluoroisopropanol (d-HFIP) to deuterate exchangeable protons and subsequently aliquoted into 1 μmol fractions and lyophilized. Aggregation was initiated by dissolving the dried samples in 2, 4 or 6.5 μL of 20 mM phosophate buffer (pH* ~ 7), for a total peptide concentrations were either 150, 250, or 500 μM as described in the text. Isotope dilution experiments were carried out by mixing d-HFIP stock solutions of unlabeled and FGAIL labeled hIAPP prior to drying under vacuum in a ratio of 4:1 unlabeled:labeled peptide for a total peptide concentration of 1 mM. Freshly dissolved samples were sealed between two 2x25 mm CaF_2 windows separated by a 56 μm Teflon spacer and experiments were carried out under an atmosphere of dry air at room temperature (~22 °C).

Two-Dimensional Infrared Spectroscopy

Details of how two-dimensional infrared (2D IR) spectra are collected and processed have been described previously.[4] Briefly, a home-built Ti:sapphire oscillator is used to generate 800 nm pulses, which are amplified in a neodymium-doped yttrium lithium fluoride pumped regenerative amplifier (Spectra Physics Spitfire). Difference frequency generation of the outputs from a home-built optical parametric amplifier produces ~100 fs mid-IR pulses centered at 6 μm, which are split (90/10) into pump and probe pulses. The pump pulses are passed through a Ge acousto-optic modulator pulse shaper which generates the two pump pulses used in the 2D IR experiment, and these are overlapped and focused at the sample. A liquid nitrogen cooled mercury cadmium telluride array (Infrared Systems) is used to detect the signal. The approximate delay between initiating hIAPP aggregation and the collection of the first 2D IR spectrum is ~2 min. Spectra are then collected every ~40 seconds until aggregation is complete. A running average of 10 spectra (~7 minutes) is used to produce the kinetic traces for plotting.

Lag Phase and Fibril State TEM Images

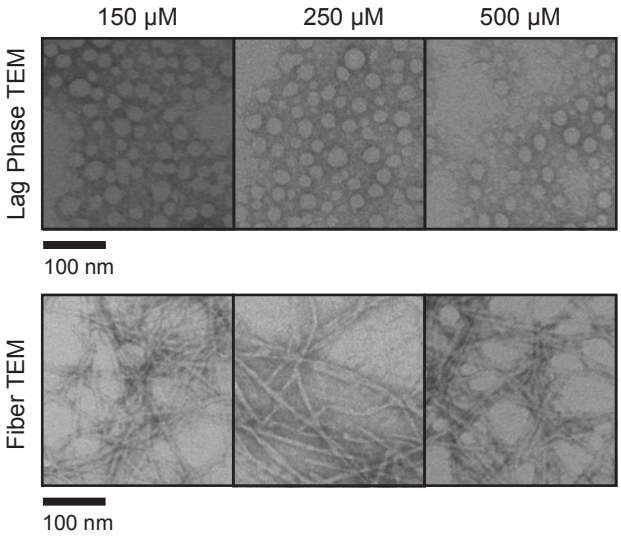

Figure S1. TEM images collected during the lag phase and after equilibration at 150, 250, and 500 μM concentrations. No fibrils are observed during the lag phase. Lag phase TEM images were collected ca. 15 minutes after initiating aggregation, and fibril TEM images were collected ca. 6 hours after initiating aggregation.

Additional Experimental Results

In Table S1 can be found the diagonal cut frequencies for the three states identified in the labeled region of the 2D IR spectra at the experimental concentrations, for a number of repetitions, along with averages and standard deviations. Representative diagonal cut data for experiments listed in Table S1 are reproduced in Fig. S2.

Frequencies (cm^{-1})	500 μM			250 μM		150 μM		Average (SD)
monomer	NA	1581	1586	1591	1585	1590	1580	1586 (4)
intermediate	1573	1570	1571	1570	1570	NA	NA	1571 (1)
fibril	1568	1559	1561	1563	1564	1560	1564	1563 (3)

Table S1. The frequencies for the peaks identified as monomer, intermediate, and fibril for different repetitions of experiments at 500, 250, and 150 μM, with the average frequencies and standard deviations.

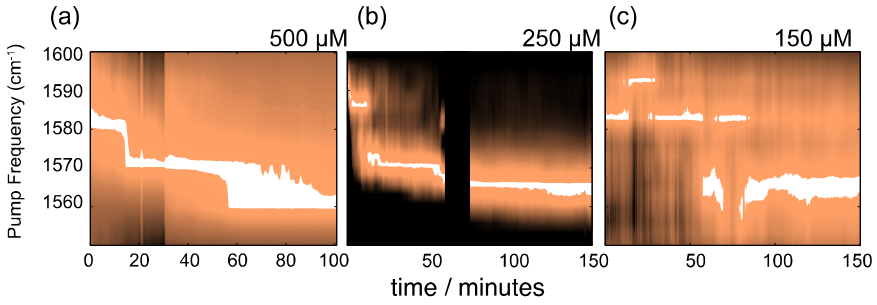

Figure S2. Diagonal cuts in the isotope labeled region for representative reproductions of the data in Table S1. The black cut-out in the middle panel at ca. 60-70 minutes occurred because the liquid nitrogen cooled detector needed to be refilled during the course of the experiment.

Modeling of Two-Dimensional Infrared Spectra

Spectra modelling was performed using the molecular snap-shot approach.[5,6] For each of the three simulated spectra in Fig. 3, homogenously broadened 2DIR spectra are generated for each of a large number of sample structures, which are then added together to produce an ensemble averaged spectrum. Specifically, for a given snapshot, local Amide-I vibrational mode frequencies are selected from a normal distribution centered around 1585 cm^{-1} and with a standard deviation of 12 cm^{-1}.[5,7] For each snapshot, the one-quantum floating oscillator Hamiltonian matrix for the Amide-I subspace is generated using coupling constants from the dihedral map generated by Jansen and coworkers[8] for the nearest neighbor terms, while all other coupling terms are generated using transition dipole coupling with values parameterized by Krimm.[9] The two-quantum Hamiltonian and dipole matrices are then generated from the one-quantum Hamiltonian and dipole matrices using the harmonic approximation,[10] from which two-dimensional stick spectra for rephasing and nonrephasing responses are generated. Two-dimensional histograms of 500 such stick spectra are then convoluted with two-dimensional homogenous lineshapes, for rephasing and nonrephasing spectra respectively, with 5 cm^{-1} HWHM, and added together to produce the final simulated absorptive 2DIR spectra. The models only include the isotope labeled residues, FGAIL, as interactions with unlabeled residues should be negligible. In order to account for the effect of incomplete isotope substitution, 10% of the residues are randomly assigned a +40 cm^{-1} shift. The backbone configurations for the initial monomeric peptide, T=0 min, were generated from a distribution of coiled polypeptide structures produced using a self-avoiding chain Monte Carlo simulation where the C_α-carbon self-avoiding distance was set to 2 Å and where the phi-psi angle pairs where selected using a Metropolis algorithm to reproduce the distribution of non-secondary structure dihedrals as determined by the Top 500 distribution.[11] The intermediate spectrum, T=20 min, is modeled using an ideal β-sheet structure with (phi,psi)=(120°,120°) The final fibril spectrum is modeled using a longer ideal β-sheet with 2 residues, (F and G) adopting coil configurations, in the same manner as the monomer model, in order to represent their role in the disordered turn region of hIAPP as determined by ssNMR.[12]

Isotope Dilution Experiments

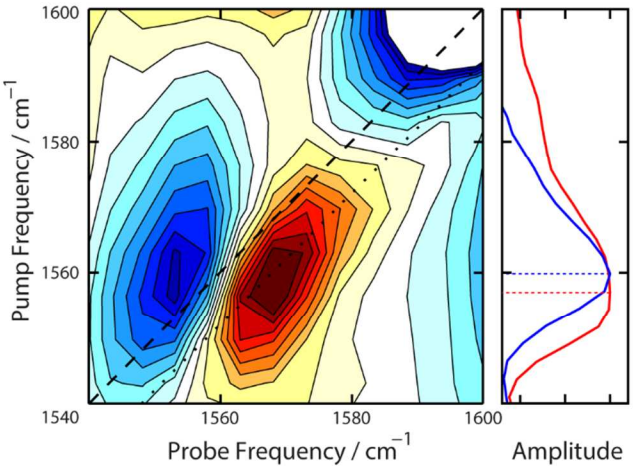

Figure S3. 2D IR spectrum (left) of 1:4 isotope labeled to unlabeled sample of fibrillar hIAPP (total concentration 500 μM), with diagonal cuts (right) of the diluted (blue) and fully labeled (red) spectra overlain for comparison.

Isotope dilution experiments involving diluting the isotope labeled peptide with unlabeled peptide can be used to eliminate effects caused by interstrand coupling and thus can be used to validate spectral assignments and isolate the origin of frequency shifts.[4,13–16] This experimental approach has been used previously to study amyloid structures and kinetics and to elucidate the structure of the FGAIL intermediate,[17] and we use the same approach here.

Fig. S3 shows isotope dilution for 1:4 labeled:unlabeled peptide in the fibrillar state. We observe a blue-shift from 1558 cm^{-1} to 1560 cm^{-1} in the β-sheet transition frequency for the diluted spectrum, indicative of stronger interstrand coupling in the undiluted sample, consistent with a β-sheet structure in the FGAIL region. For hIAPP with all five FGAIL residues C^{13}=O^{18} labeled, we observe a larger absolute red-shift between the monomer and fibril states in the isotope labeled region, relative to that observed for single isotope labels on individual residues this region in a previous study.[17] This is indicative of a transition that is delocalized over multiple residues in each individual strand, expected when all five FGAIL residues are close to a β-sheet configuration, as expected from ssNMR data on hIAPP.[12]

Derivation of the 3-State Mass Action Kinetic Model

Here we detail the derivation of the 3-state model of the kinetics expressed in equation 1, using the approximations stated in the main text. The fast equilibration approximation, used in classical nucleation models,[18] assumes that small, fast diffusing, species achieve detailed balance prior to the rate limiting step of the reaction. This would imply from equations 1 and 2 that, for a small oligomer of size l:

$$k_{-olig} c_l = k_+ c_1 c_{l-1}$$

$$c_l = \frac{c_1 c_{l-1}}{c_{sup}} \quad \text{.(S1)}$$

Assuming that this expression holds for all species smaller than n leads to:

$$c_{n-1} = \frac{c_1^{n-1}}{c_{sup}^{n-2}} \text{ (S2)}$$

And to the following concerted rate equation for the concentration of the nucleus species:

$$\frac{dc_n}{dt} = k_+ \frac{c_1^n}{c_{sup}^{n-2}} - k_{-olig} c_n + k_{-a} c_n^{\ddagger} - k_{+a} c_n \text{ .(S3)}$$

Where the first term is just the bimolecular rate for formation of c_n from c_1 and c_{n-1}, where c_{n-1} was substituted using equation S2. k_{+a} and k_{-a} are the forward and reverse rate constants for formation of the activated (double-dagger) species. As the species represented by $c_n^{\ddagger}$ is an activated complex, further simplification is possible using the standard assumptions of activated complex/transition state theory. From the mechanism in equation 1 and assuming a quasi-steady state population of the activated complex:

$$c_n^{\ddagger} = \frac{k_{+a} c_n + k_- f_{n+1}}{k_{-a} + k_+ c_1} \approx \frac{k_{+a}}{k_{-a}} c_n = c_n e^{-\Delta G^{\ddagger}/RT} \text{ (S4)}$$

where it has been assumed that the unimolecular rates are much faster than the bimolecular rates ($k_{+a} \gg k_- f; k_{-a} \gg k_+ c_1$). This allows the rate of passage from the nucleus, n, to the first fibril state, $n+1$, to be written as:

$$k_+ c_1 c_n e^{-\Delta G^{\ddagger}/RT} \text{ .(S5)}$$

Recognizing that at equilibrium ($c_1 = c_{crit}$), the most unstable species (excluding the activated state, $c_n^{\ddagger}$,which we assume is only transiently populated and not present significantly at equilibrium) is the nucleus which thus represents an extremum in concertation such that, $f_{n+1} \approx c_n$, we may find an expression for the reverse rate constant for formation of the first fibril state from the following detailed balance condition:

$$k_+ c_{crit} c_n e^{-\Delta G^{\ddagger}/RT} = k_{rev} f_{n+1} \approx k_{rev} c_n$$

$$k_{rev} = k_+ \frac{k_-}{k_+} e^{-\Delta G^{\ddagger}/RT} = k_- e^{-\Delta G^{\ddagger}/RT} \text{ .(S6)}$$

If we assume that $f_{n+1} \approx c_n$, which holds at equilibrium as described above, is also a good approximation at all times, we get for the rate equation for the smallest fibril species:

$$\frac{df_{n+1}}{dt} = \left(k_+ c_1 - k_-\right) c_n e^{-\Delta G^{\ddagger}/RT} - k_+ c_1 f_{n+1} + k_- f_{n+2} \text{ .(S7)}$$

If mass balance is enforced between f_{n+1} and c_n, given that $c_n^{\ddagger}$ is in a quasi-steady state, the last two terms in equation S3 may be rewritten to give:

$$\frac{dc_n}{dt} = k_+ \frac{c_1^n}{c_{sup}^{n-2}} - k_{-olig} c_n - \left(k_+ c_1 - k_-\right) c_n e^{-\Delta G^{\ddagger}/RT} \text{ (S8)}$$

Which is the final form of the concerted rate equation for the oligomer, as written in the main text. To get the final expression for the fibril rate, first consider the rate equation for a fibril species larger than $n+1$ as implied by the mechanism in equation 1:

$$\frac{df_i}{dt} = k_+ c_1 f_{i-1} - \left(k_+ c_1 + k_-\right) f_i + k_- f_{i+1} \text{(S9)}$$

If we sum this expression from $m = n+2$ to $m = \infty$ and add it to equation (S7) we get:

$$\frac{d}{dt} f = \left(k_+ c_1 - k_-\right) c_n e^{-\Delta G^{\ddagger}/RT} \text{ (S10)}$$

Where $f = \sum_{i=n+1}^{\infty} f_i$. This is nearly the final form used in the main text. One additional term is introduced that adds an autocatalytic mechanism for formation of fibrils. One common mechanism of this type is fragmentation, the possibility of a fibril of length i splitting into two fibrils of length l and $i-l$. As this can occur at $i-1 \approx i$ sites for a linear fibril, the total rate of fragmentation should be linearly proportional to the total number of peptides currently in a fibrillar state, $c_{tot} - c_1 - nc_n$, and it is assumed the rate constant for this process is the same as for detachment of a single peptide from a fibril, k_-, giving the final form for df/dt:

$$\frac{df}{dt} = \left(k_+ c_1 - k_-\right) c_n e^{-\Delta G^{\ddagger}/RT} + k_-(c_{tot} - c_1 - nc_n) \text{ .(S11)}$$

Finally, if mass balance is enforced between the expressions in equations S8, S9 and the concentration of monomers, the rate equation for monomers becomes:

$$\frac{dc_1}{dt} = n\left(k_{-olig} c_n - k_+ \frac{c_1^n}{c_{sup}^{n-2}} \right) + k_- \sum_{i=n+2}^{\infty} f_i - k_+ c_1 \sum_{i=n+1}^{\infty} f_i \text{ (S12)}$$

where we've treated the monomer lost in formation of the smallest fibril species in equation S7 as negligible. Using our final approximation, $f = \sum_{i=n+1}^{\infty} f_i \approx \sum_{i=i+2}^{\infty} f_i$, equation S12 leads to the final expression for the rate equation for monomers which completes the set of coupled equations used to model the data in the main text.

In Fig. S4 we show the predicted t_{50} dependence with concentration of our model when fragmentation and the barrier are excluded. The purpose for this is that an analytical solution to this scenario exists, developed by Powers and Powers,[19] which appears as a dotted line in Fig. S4. This comparison is done to show that in spite of our approximations (in particular pre-equilibration) we can still obtain the correct scaling behavior and abrupt change at the supercritical concentration that is observed in the analytical model, suggesting that the need to include fragmentation and a free energy barrier is not due to our approximations but from a failure of classical models to reproduce the 2DIR data.

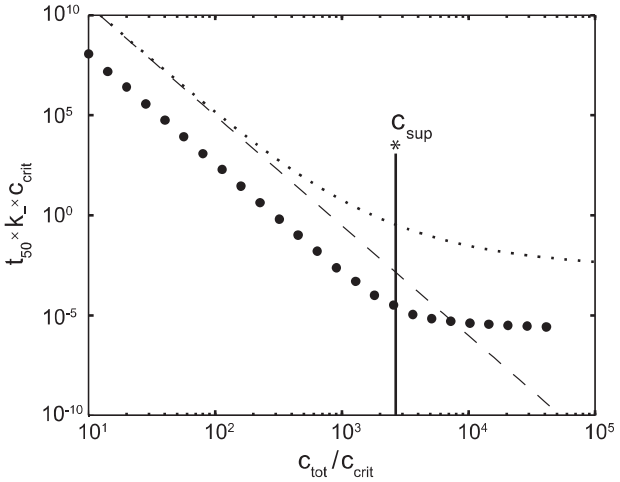

Figure S4. Comparison between the concentration dependence of t_{50} for our model when fragmentation and the barrier are both excluded (circles), the analytic classical nucleation model (dashed lines) and the modified analytic model by Powers that accounts for saturation of the supercritical concentration (dotted lines).

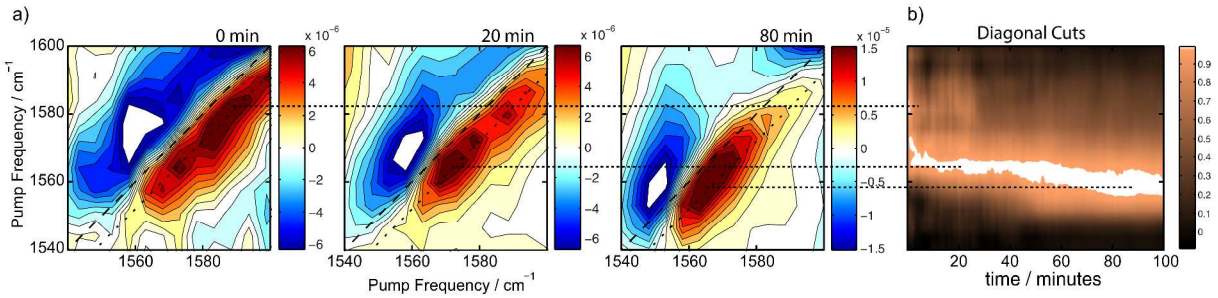

Figure S5. A reproduction of Figure 1 from the main text with color bar scales included. (a)-(c) 2D IR spectra of the isotopically labelled Amide-I vibrations of 0.5 mM FGAIL-labeled hIAPP in 20 mM phosphate buffer in the isotope labeled region at the indicted times after initiation of aggregation, and (d) as diagonal cuts through the fundamental transition in the 2D spectra as a function of time after initiation. Blue and red features correspond to negative and positive signals, respectively.

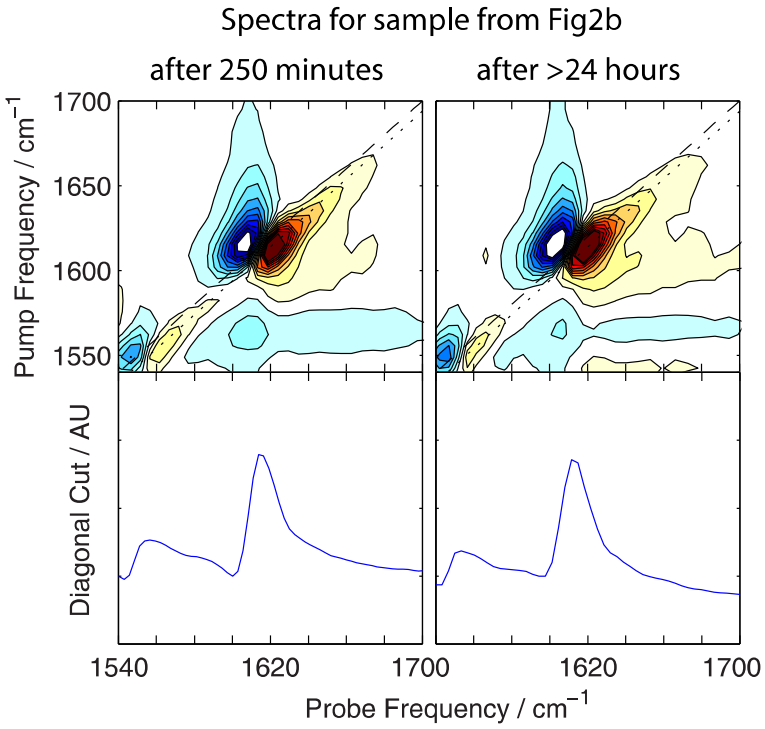

Figure S6. 2D IR spectra collected after a longer equilibration time to demonstrate that the spectrum does not evolve further after the final spectrum shown in the main text. The 2D IR

spectrum of the sample from Figure 2b in the main text are shown after 250 min and after greater than 24 hrs equilibration. The spectra show similar frequencies and intensities.

References

(1) Marek, P.; Woys, A.; Sutton, K.; Zanni, M.; Raleigh, D. *Org. Lett.* **2010**, *12*, 4848–4851 DOI: 10.1021/ol101981b.
(2) Abedini, A.; Raleigh, D. P. *Org. Lett.* **2005**, *7*, 693–696 DOI: 10.1021/ol047480+.
(3) Marecek, J.; Song, B.; Brewer, S.; Belyea, J.; Dyer, R. B.; Raleigh, D. P. *Org. Lett.* **2007**, *9*, 4935–4937 DOI: 10.1021/ol701913p.
(4) Shim, S.-H.; Gupta, R.; Ling, Y. L.; Strasfeld, D. B.; Raleigh, D. P.; Zanni, M. T. *Proc. Natl. Acad. Sci.* **2009**, *106*, 6614–6619 DOI: 10.1073/pnas.0805957106.
(5) Hamm, P.; Lim, M.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123–6138 DOI: 10.1021/jp9813286.
(6) Ganim, Z.; Tokmakoff, A. *Biophys. J.* **2006**, *91*, 2636–2646 DOI: 10.1529/biophysj.106.088070.
(7) Hamm, P.; Lim, M.; DeGrado, W. F.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 2036–2041.
(8) la Cour Jansen, T.; Dijkstra, A. G.; Watson, T. M.; Hirst, J. D.; Knoester, J. *J. Chem. Phys.* **2006**, *125*, 44312 DOI: 10.1063/1.2218516.
(9) Moore, W. H.; Krimm, S. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 4933–4935.
(10) Hamm, P.; Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy*; Cambridge University Press, 2011.
(11) Lovell, S. C.; Davis, I. W.; Arendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins* **2003**, *50*, 437–450 DOI: 10.1002/prot.10286.
(12) Luca, S.; Yau, W.-M.; Leapman, R.; Tycko, R. *Biochemistry (Mosc.)* **2007**, *46*, 13505–13522 DOI: 10.1021/bi701427q.
(13) Kim, Y. S.; Liu, L.; Axelsen, P. H.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 7720–7725 DOI: 10.1073/pnas.0802993105.
(14) Kim, Y. S.; Liu, L.; Axelsen, P. H.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci.* **2009**, *106*, 17751–17756 DOI: 10.1073/pnas.0909888106.
(15) Middleton, C. T.; Marek, P.; Cao, P.; Chiu, C.; Singh, S.; Woys, A. M.; de Pablo, J. J.; Raleigh, D. P.; Zanni, M. T. *Nat. Chem.* **2012**, *4*, 355–360 DOI: 10.1038/nchem.1293.
(16) Falvo, C.; Zhuang, W.; Kim, Y. S.; Axelsen, P. H.; Hochstrasser, R. M.; Mukamel, S. *J. Phys. Chem. B* **2012**, *116*, 3322–3330 DOI: 10.1021/jp2096423.
(17) Buchanan, L. E.; Dunkelberger, E. B.; Tran, H. Q.; Cheng, P.-N.; Chiu, C.-C.; Cao, P.; Raleigh, D. P.; Pablo, J. J. de; Nowick, J. S.; Zanni, M. T. *Proc. Natl. Acad. Sci.* **2013**, *110*, 19285–19290 DOI: 10.1073/pnas.1314481110.
(18) Hill, T. L. *Linear Aggregation Theory in Cell Biology*; Rich, A., Series Ed.; Springer Series in Molecular Biology; Springer New York: New York, NY, 1987.
(19) Powers, E. T.; Powers, D. L. *Biophys. J.* **2006**, *91*, 122–132 DOI: 10.1529/biophysj.105.073767.