Supporting Information: Direct Observation of Insulin Association Dynamics with Time Resolved X-Ray Scattering

Dolev Rimmerman[†], *Denis Leshchev*[†], *Darren J. Hsu*[†], *Jiyun Hong*[†], *Irina Kosheleva*[‡], *Lin X. Chen*[*,†,§]

[†]Department of Chemistry, Northwestern University, Evanston, IL 60208

[‡]Center for Advanced Radiation Sources, The University of Chicago, Illinois 60637, USA

[§]Chemical Sciences and Engineering Division, Argonne National Laboratory, Argonne, IL 60439

*To whom correspondences should be addressed. E-mail: l-chen@northwestern.edu

Supporting information

SI.1 Experimental Methods

Sample preparation

Bovine insulin was purchased from Sigma Aldrich (I5550) and used without any further purification. The Insulin was dissolved in an aqueous mixture of ethanol (20% v/v), 0.1 M NaCl and 0.27 M HCl as used in previous studies of the insulin dimer-monomer system[1,2]. In order to study the insulin monomeric state, samples were dissolved in an aqueous mixture of 20% acetic acid (v/v). The protein concentration for all experiments was 5 mg/mL and samples were filtered using a 0.2 μm filter prior to use. For all experiments the samples were used immediately after preparation. In the case of TRXSS experiments, the sample was loaded into a reservoir and replaced with fresh sample after 8 hours of use.

Temperature dependent UV-CD

UV-CD measurements were taken on a temperature controlled Jasco J-815 CD instrument at Northwestern University's Keck Biophysics Facility. The sample was equilibrated at each temperature (15-50 °C) for several minutes prior to collection of the UV-CD spectrum between 240-340 nm. The intensity of the peak at 275 nm was monitored to determine the association state of the protein at each temperature.

Temperature dependent static SAXS for Guinier and Kratky plot analysis

Static SAXS measurements were taken at Sector 5 of the Advanced Photon Sources (APS) at Argonne National Labs[3]. Samples were flowed at a rate of 30 μL/s through a 0.7 mm capillary. The capillary flow cell assembly and sample input tubing were temperature controlled, allowing for static temperature measurements (15-50 °C). Scattering curves were obtained from three area detectors simultaneously, allowing for collection of SAXS/WAXS in a single exposure. These SAXS data were used for Guinier and Kratky plot analysis.

T-Jump X-Ray Scattering

TRXSS experiments were carried out at BioCARS 14-ID-B beamline at the Advanced Photon Source (APS). Details of the x-ray scattering setup and generic TRXSS pump-probe data acquisition methodology at BioCARS have been previously published.[4,5] For conducting T-jump experiments, the sample was excited by 7 ns laser pulses with a 1.443 μm wavelength which corresponds to an overtone of O-H stretch.[6] The

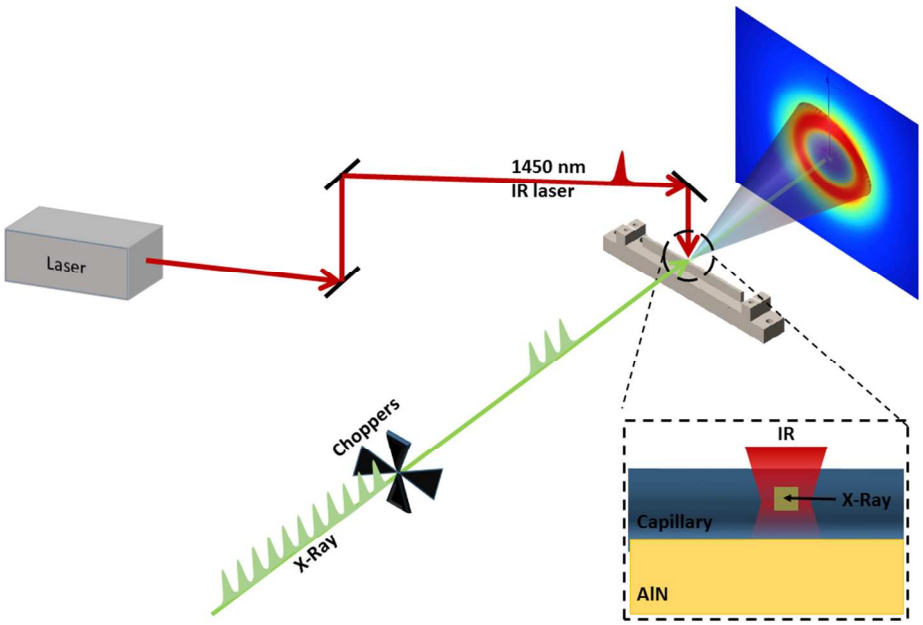
laser beam was focused on a capillary with the sample solution to an elliptical spot 65 μm x 350 μm delivering energy of ~1 mJ in each pulse, leading to a power density of ~56 mJ/mm^2 at the focal spot. Electronically delayed x-ray pulses with a pink spectrum centered at 11.65 keV were used to probe the laser induced structural changes in the sample. A pair of K-B mirrors focused the x-ray beam on the sample to a beamsize of 35 μm by 35 μm. The x-ray and laser beams were overlapped on the capillary with sample solution at 90 degrees angle between each other and the capillary. Due to the Gaussian shape of the laser beam and the perpendicular geometry of the pump and probe, some inhomogeneity in the T-jump is unavoidable and exists along the x-ray probe propagation direction, in particular at the sides of the capillary. It is estimated that at the very edges of the capillary, where the x-ray enters and leaves the sample, the laser excitation is ~40% of the power density at the center of the capillary, where the excitation is more uniform. The excitation is also more uniform across the vertical axis of the x-ray beam, where the homogeneity is ~90%. Therefore, the final magnitude of the T-jump (~8 °C) represents an ensemble average of the excited state population.

The scattered x-rays were collected with the Rayonix340 camera placed at 365 mm distance from the sample. In order to reduce the background contribution, a cone-shaped chamber filled with helium was installed between the sample and the detector. The chopper-shutter system described in literature was employed to reduce the repetition rate of the synchrotron source operating in 24 bunch mode.[5] Variation of the opening time of the high speed triangular chopper allowed us to control the time resolution of the experiment at the cost of incoming flux. For the late time delays (beyond 5 μs) the chopper opening time was set to 3.7 μs therefore selecting x-rays pulses from 24 electron micro-bunches in the ring. Similarly, for short time delays (below 5 μs) the chopper opening time was chosen to provide x-rays from a single x-ray bunch with pulse duration of ~100 ps. The repetition rate of the experiment was adjusted from 20 Hz to 0.8 Hz depending on the investigated time delay range. Data collection was set up to collect a time series of specific positive time delays spanning from nanoseconds to milliseconds, interwoven with negative time delays to allow for calculation of difference curves. The positive time delays in the data collection series were rotated randomly to verify that there were no erroneous signals. The collected images were polarization and geometry corrected and detector nonuniformities were masked out. The images were then azimuthally averaged to produce 1D scattering curves. The curves were scaled by normalization around the isosbestic point of the solvent at q~2.0Å, and then difference curves were produced from the subtraction of negative time delay data from positive time delay data. Data taken at different repetition rates and bunch modes included a shared

time delay point (5 μs) to verify the reproducibility of the data under different scan conditions (rep-rate and time resolution) and was found to be consistent under all the conditions used in the experiment. For the short time delays series (t<5 μs), typically ~40 images were collected for each Laser-ON time delay; overall 22 time delays were collected for this dataset at 35°C. For late time delays (t>5 μs), typically 10 images were collected for each time point; overall 16 time delays were collected for this dataset at 35°C. Similar data collection parameters were used for collection of 25°C and 30°C datasets.

During data collection, the sample solution was loaded into the custom built capillary flow cell system using a syringe pump (see SI figure 1). The 0.70 mm diameter open ended capillary was held on a temperature controlled capillary holder constructed from Aluminum Nitride (AlN) for the purpose of the experiment. The capillary was placed inside a channel with depth of 150 μm and radius of 350 μm in order to provide thermal contact between the AlN holder and the bottom of the capillary while allowing the x-ray to propagate through the side of the capillary. During data collection, the capillary was translated horizontally in steps of 250 μm after each pair of pump-probe pulses such that each time a fresh volume of sample was illuminated. Each x-ray image was recorded by using ~280 x-ray pulses, which correspond to a total translation of the capillary (and the sample holder) by ~7 cm by the end of the exposure. After each image recording the capillary was translated back to the initial position and a new portion of the sample was withdrawn into capillary. Before starting a new exposure, the sample holder was allowed to thermalize for 5 seconds, so the new portion of sample solution can reach the preset temperature. The temperature stability prior to exposure was verified by measuring the temperature with a thermistor mounted on the AlN holder as well as a thermistor mounted at the top of the capillary away from the exposed region.

Temperature dependent static scattering curves for protein and pure buffer were similarly collected at the BioCARS beamline (without laser excitation). The static data from the TRXSS beamline was used for obtaining the static difference curves for WAXS analysis and estimating the absolute magnitudes of difference signals.



SI figure 1 - T-jump Transient X-ray Scattering instrument setup.

SI.2 UV-CD and SAXS results

As mentioned in the main text, both SAXS and UV-CD measurements were carried out in order to verify the association state of the insulin oligomerization state at the temperature range of 15-50°C. Insulin solutions were prepared identically to the procedure for TRXS measurements. SAXS results were analyzed with Guinier analysis to extract radius of gyration, according to standard procedures.

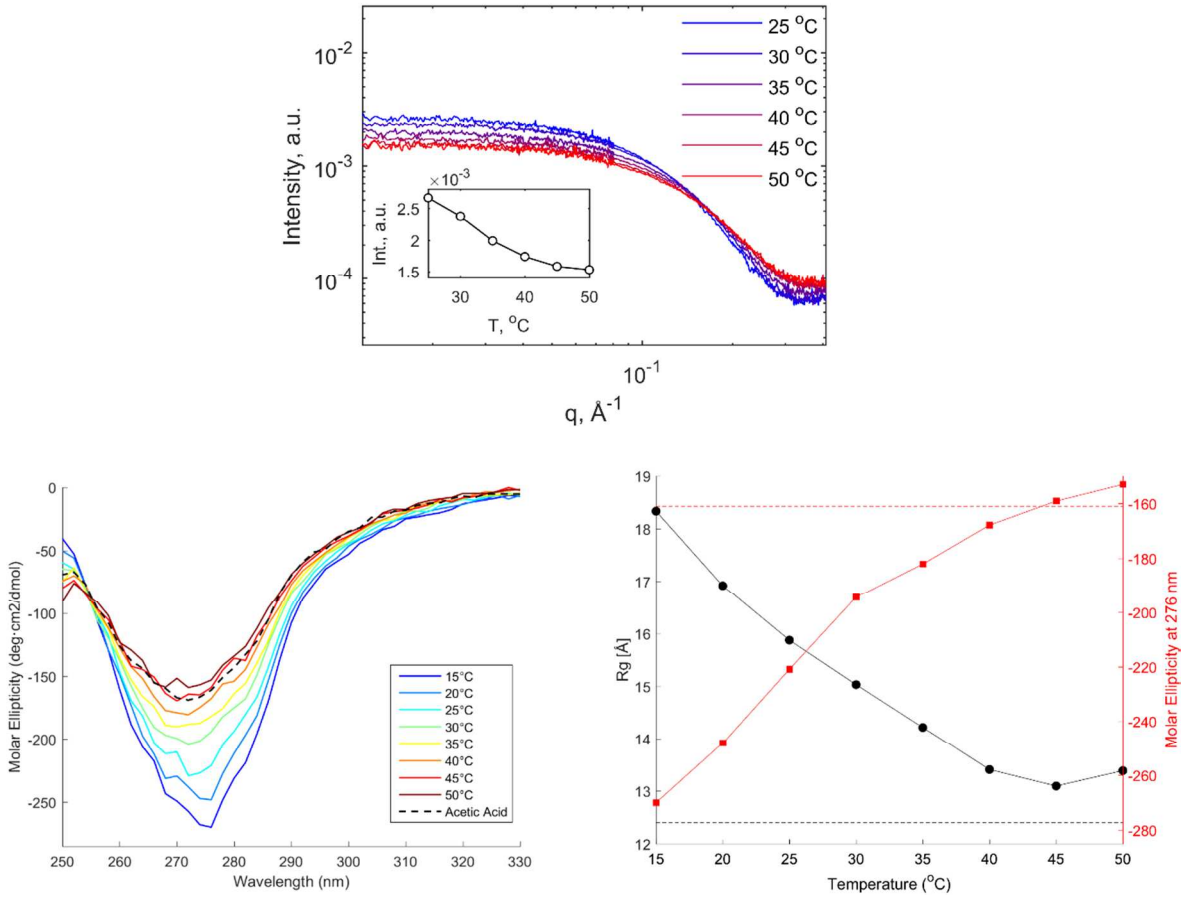
UV-CD spectra were measured and molar ellipticity at 276 nm was extracted from the spectra at different temperatures. Previous studies have shown that the tyrosyl signal at 276 nm is strongly affected by insulin oligomerization, in particular by the process of forming dimers from monomers.[7,8] This band is strongly negative in the dimer state, however it becomes less negative in the monomeric state.

For pure monomer reference data, both UV-CD and SAXS data were collected on insulin in 20% Acetic Acid at 25°C. Acetic acid was used as co-solvent with water as it is known to completely dissociate insulin, even in the presence of zinc[7,9]. Our analysis of the SAXS data reveals that the R_g of monomer insulin at 25°C is 12.2 Å, similar to previous results published by Uversky et al.[7] UV-CD results indicated that in Acetic acid the molar ellipticity is -161 $\frac{deg \cdot cm^2}{dmol}$, in the range previously published for monomeric insulin. The results for 20% ethanol solutions at different temperatures are shown in SI table 1 and SI figure 2. Our results indicate that at 45 °C the insulin in EtOH, pH=0 solution approximately reaches the

molar ellipticity at 276 nm as the acetic acid sample, indicating the presence of mostly monomers in solution. Furthermore, the radius of gyration reaches its minimum at the same temperature, indicating that monomer radius of gyration may be slightly larger at higher temperatures in 20% ethanol.

Temperature (°C)	15	20	25	30	35	40	45	50
R_g (Å)	18.3	16.9	15.8	15.0	14.2	13.4	13.1	13.4
[θ] ($\frac{deg \cdot cm^2}{dmol}$) at 276 nm	-270	-248	-221	-194	-182	-168	-159	-153

SI Table 1- R_g and mean residue ellipticity of tyrosyl band as a function of temperature.

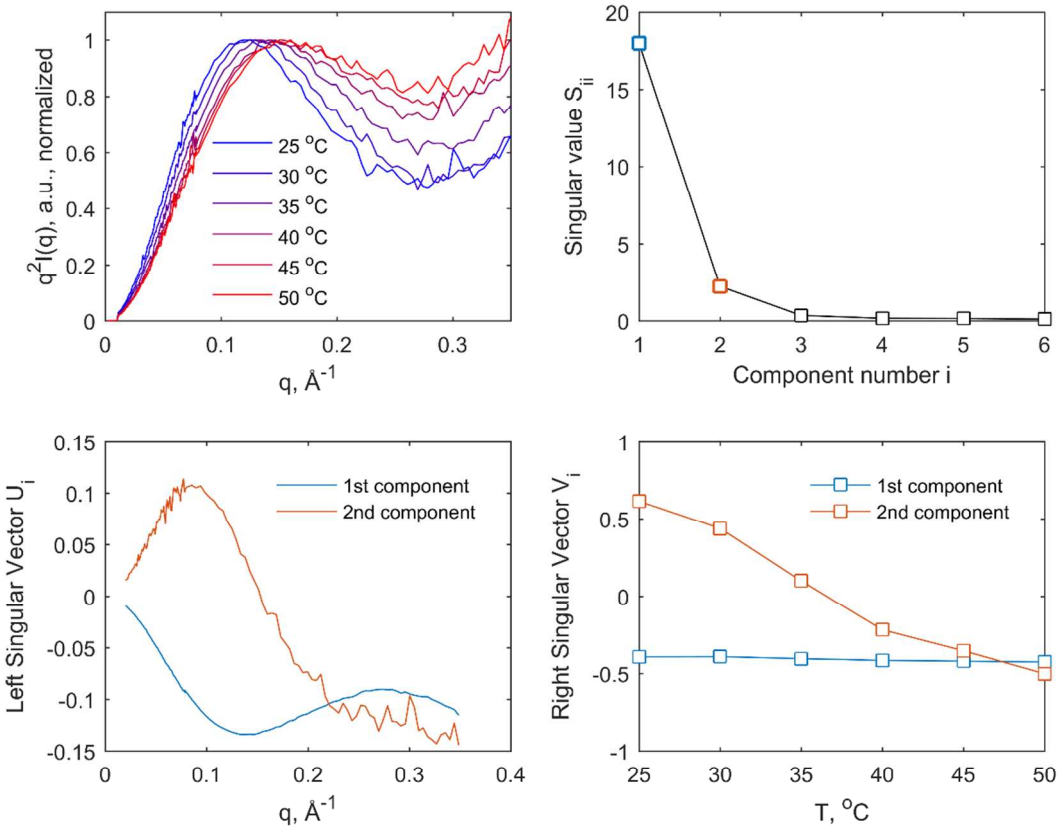


SI figure 2 – (top) SAXS curves of the insulin monomer-dimer mixture in water- (20%) ethanol buffer. The inset shows the intensity at forward scattering at different temperatures. (bottom-left) UV-CD spectra of insulin in 20% EtOH solution at pH=0 at different temperatures. The dashed line is the spectra from the insulin monomer state in Acetic Acid solution at room temperature. (bottom-right) R_g and molar ellipticity of at 278nm as a function of temperature. the dashed lines are the reference values for insulin monomers from acetic acid solution.

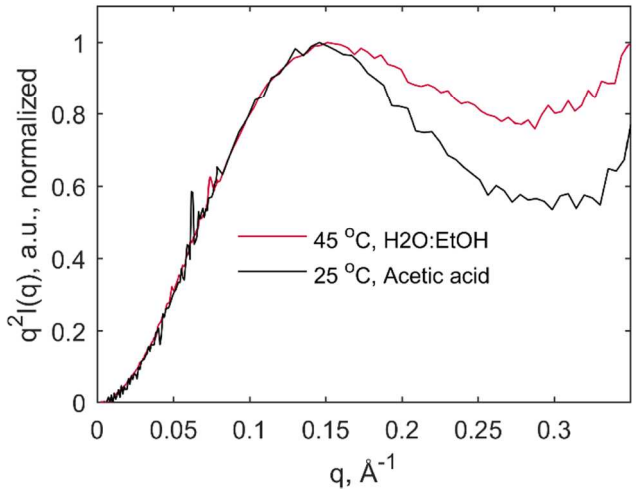
SI 3. Kratky curve analysis for static data

The temperature dependence of Kratky curves derived from static SAXS data was analyzed in order to evaluate the conformation of the insulin in its dimeric and monomeric forms (SI-figure 3). All the

collected curves exhibit a Gaussian curve shape indicating the folded confirmation of the protein. The temperature increase results in the shift of the maximum intensity towards higher q values in agreement with observed decrease in Rg. At high temperatures, the shape of the curve becomes similar to the one collected for isolated monomers in a buffer containing acetic acid (SI-figure 4). The monomer curve exhibits a strong peak in the intensity at 0.18 $Å^{-1}$, which has been shown to indicate its folded form.[10] The higher plateau for insulin in EtOH at 45°C in comparison with acetic acid indicates that under the former conditions insulin adopts a more disordered form than the compact monomer which exists under the latter conditions. To further investigate whether any additional species can be present in the investigated monomer-dimer mixture, we have used singular value decomposition (SVD) of Kratky curves. The analysis shows that the data can be well reproduced by a low-rank approximation consisting of only two components, i.e. supporting the premise of presence of only two species in the solution (dimers and monomers). We note the possible existence of the third SVD component which may have marginal contribution to changes in monomer structure due to temperature.



SI Figure 3. (top -left) Kratky curves for insulin in EtOH-water mixture at temperatures between 25-50 ℃. (top-right) S values from SVD analysis of the Kratky curves. (bottom-left) U vectors from SVD analysis of the Kratky curves. (bottom-right) V vectors from SVD analysis of the Kratky curves.

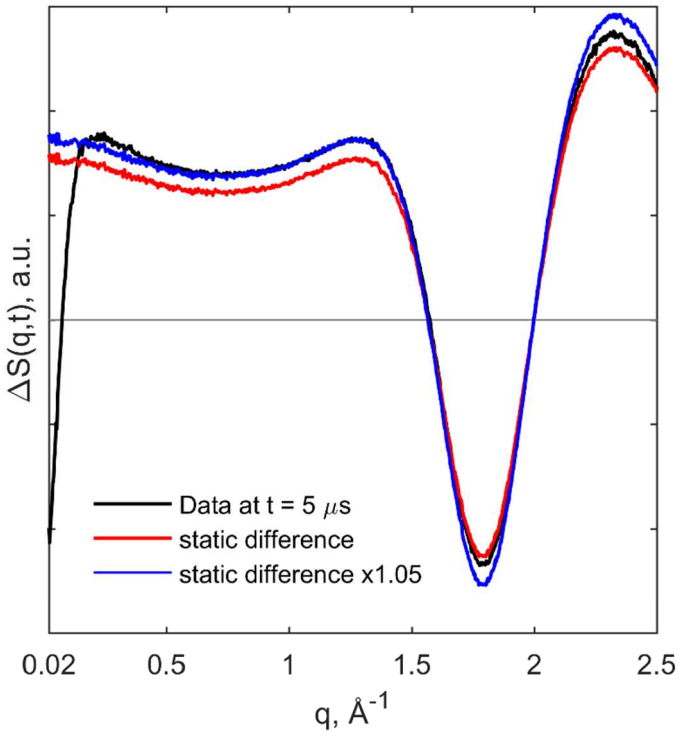


SI Figure 4. Comparison of Kratky curves for insulin in EtOH-water mixture at temperatures between 45 ℃ with the Kratky curve from insulin monomers in 20% acetic acid at room temperature.

SI.4 Comparison of T-jump in buffer to subtraction of static buffer scattering at different temperatures

In order to determine the extent of the T-jump in the sample after the IR pump, we compared the T-jump TRXSS curves recorded at 5μs time delay with the scattering difference calculated from the subtraction of static curves. In order to find the best agreement, the static data was taken at 2°C intervals from 25-39 °C and the differential was derived from the static data at 25°C. The best agreement between the T-jump and static temperature differential was at 8°C difference (i.e the differential between 33°C and 25°C), as can be seen in SI figure 5. The slight discrepancy between static and T-jump derived differences is attributed to (1) slightly higher magnitude of the T-jump compared to the static difference and (2) contribution from temperature dependent capillary scattering. During the laser induced T-jump measurements the increase of the capillary temperature is negligible, resulting in a small contribution to the scattering differences. By contrast, in the static measurements, the temperature is achieved by increasing the capillary wall temperature, in this case, to either 25 or 33 °C, and hence the capillary scattering has a small, though non-negligable, contribution to the static difference. Since capillary scattering contribution peaks at approximately 1.5 Å, the static curve was scaled such that it matches the T-jump data at q<1.5 Å at 5 us time delay. From the comparison it is evident that the capillary contribution in the scattering pattern is about 5% of the observed static

difference and hence puts a 5% uncertainty of our estimation of the achieved T-jump (0.4 degrees Celsius).



Si Figure 5 (black) Scattering difference patterns of T-jump from 25 ℃. (red) difference scattering pattern of solvent scattering at 33 ℃ and 25 ℃. (blue) scaled difference scattering pattern of solvent scattering at 33 ℃ and 25 ℃ to reproduce signal height at q<1.5Å^{-1}.

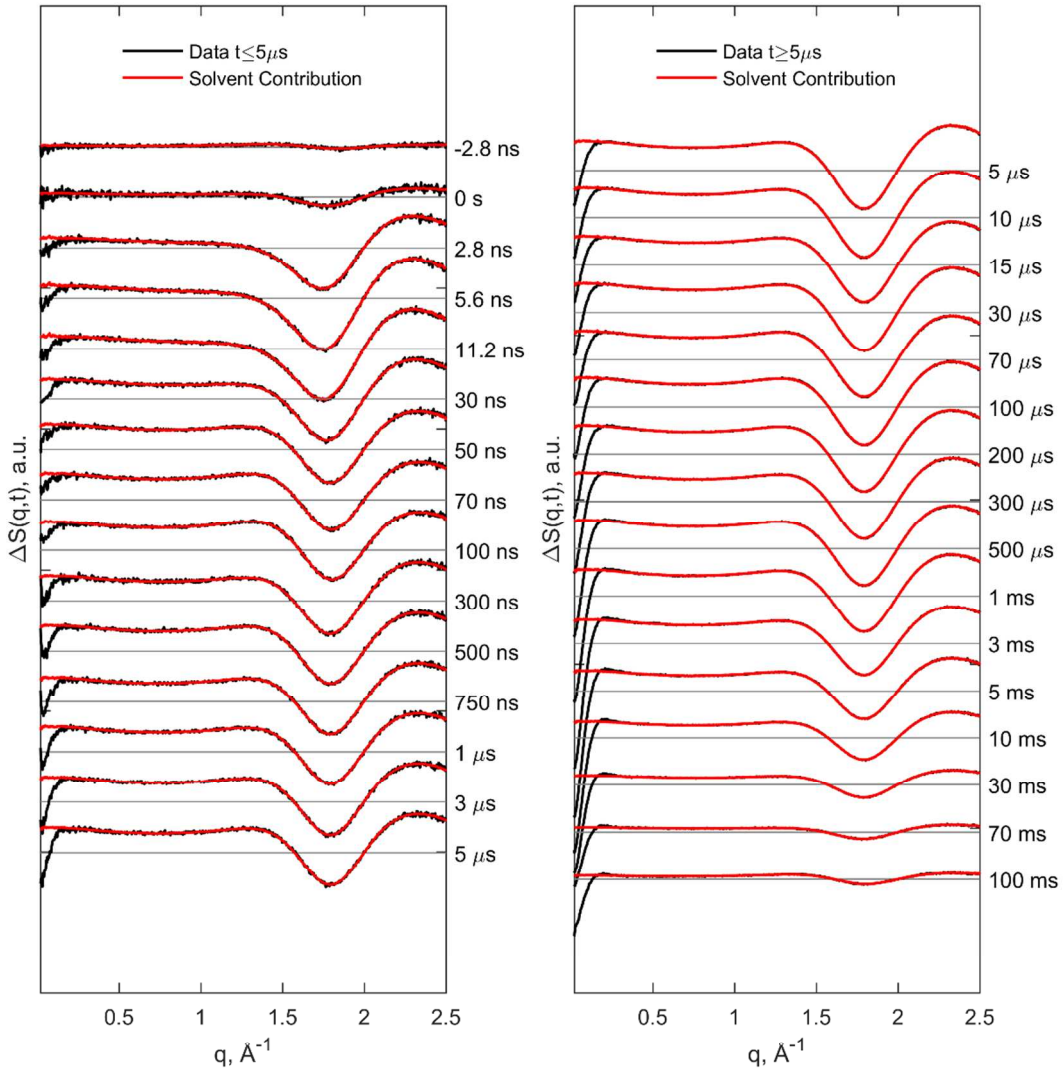
SI.5 Solvent contribution subtraction and hydrodynamics

The solvent hydrodynamic response in TRXS experiments can be described as a sum of two components:

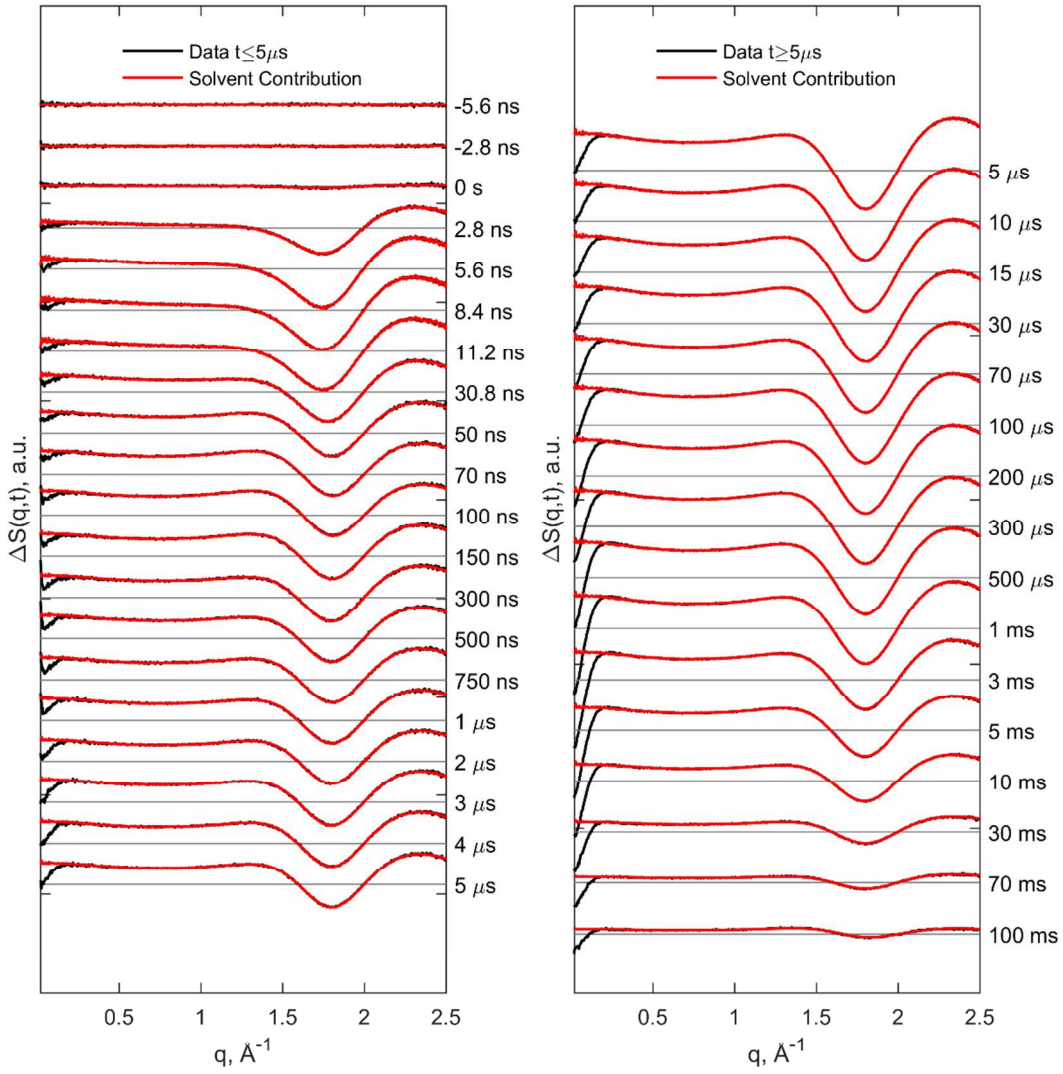
$$\Delta S_{solvent}(q,t) = \Delta T(t)\frac{\partial S(q)}{\partial T}|_{\rho} + \Delta\rho(t)\frac{\partial S(q)}{\partial \rho}|_{T}$$

Where $\Delta T(t)$ and $\Delta\rho(t)$ are time dependent changes in the temperature and density, respectively; $\frac{\partial S(q)}{\partial T}|_{\rho}$ is the change in the scattering intensity due to variations and temperature and constant density; $\frac{\partial S(q)}{\partial \rho}|_{T}$ is the change in the scattering intensity due to variations in the density at constant temperature. In the current experiments, we have determined the corresponding differentials by measuring the solvent contribution before (at 10 ns) and after (5 μs) the solvent expansion and following the standard procedure described in the literature[11]. While overall agreement between the fitted solvent response

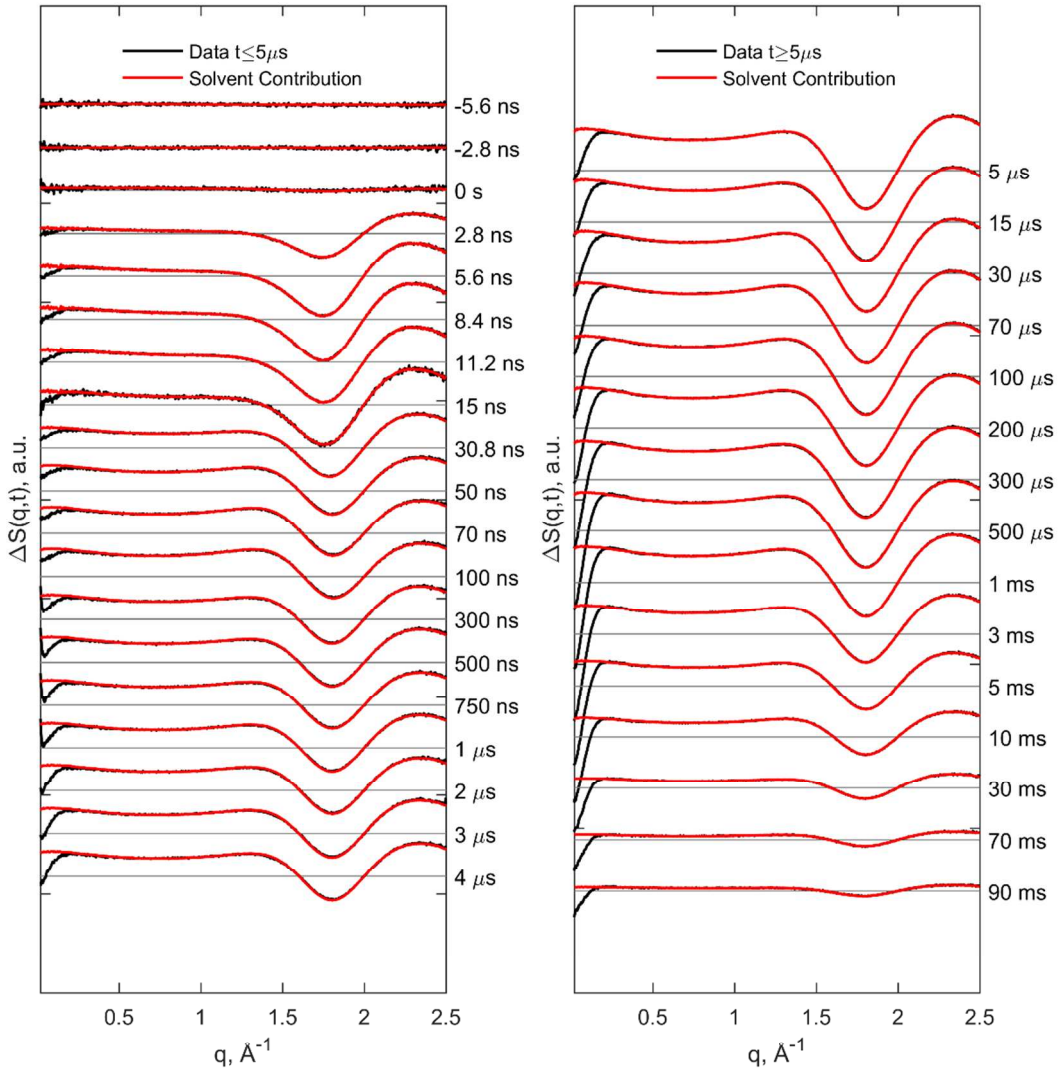
and insulin T-jump data in WAXS region ($0.5 < q < 2.5$ Å^{-1}) have been found satisfactory, for some of the curves the subtraction of solvent contribution resulted in slowly varying residuals across the WAXS region with magnitude of <3% from the differential heating signal. These residuals are suspected to arise due to fluctuations in the scattering from the capillary and the air enclosure around the sample due to slight instabilities in the sample holder temperature and air currents in the experimental chamber. These contributions have been retrieved using singular value decomposition analysis of the solvent heat response measurements. After taking these contributions into account, the fitting of the solvent contribution in the WAXS region have significantly improved making the extraction of the protein signal more reliable. We note, that the added terms did not alter the protein signal itself as these background contributions are comprised of slowly varying functions remaining essentially constant within the region where the protein contribution is significant. SI figures 6-8 show the solvent subtraction for each of the dataset collected at initial temperatures of 25, 30 and 35 ^{o}C. Finally, the hydrodynamic response of the solvent, i.e. the time dependent changes in temperature and density of the solvent, were obtained by scaling the extracted magnitudes of the thermodynamic differentials to the known values of the temperature and density before and after T-jump. The hydrodynamic response of the solvent for initial temperatures of 25 and 30 ^{o}C are shown in SI figures 9-10, whereas the 35 ^{o}C data is presented in the main text.



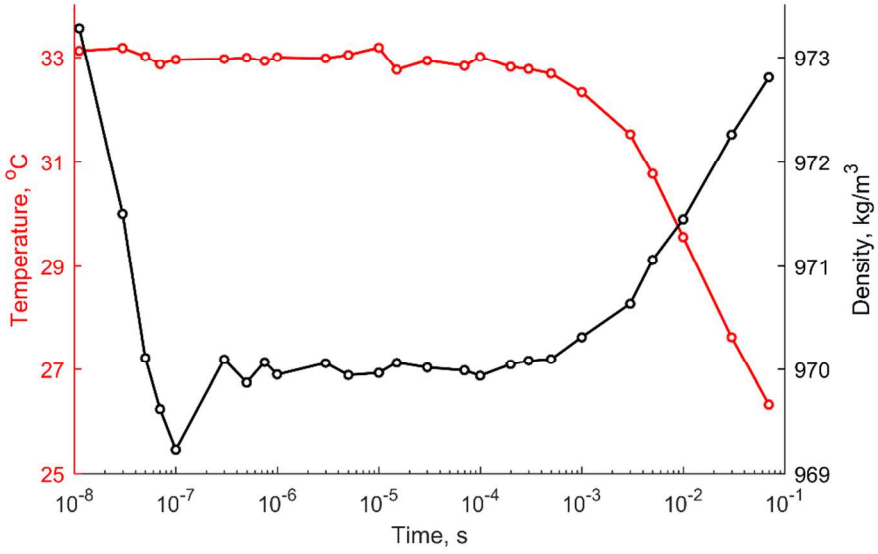
SI Figure 6. Solvent contribution subtraction from the insulin T-jump data collected at initial temperature of 25 ℃. The left panel represents the data collected with 7 ns time resolution up to 5 μs; The right panel represents the data collected with 3.7 μs time resolution from 5 μs up to 100 ms.



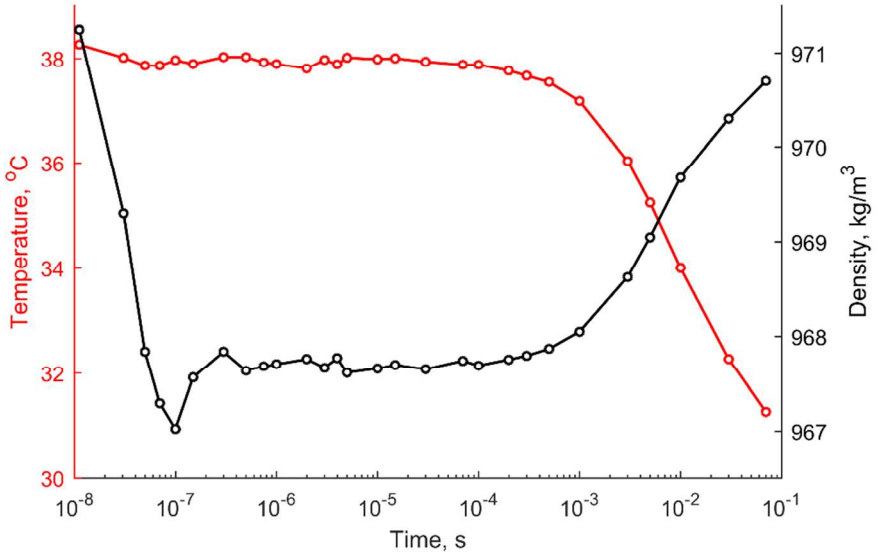
SI Figure 7. Solvent contribution subtraction from the insulin T-jump data collected at initial temperature of 30 ℃. The left panel represents the data collected with 7 ns time resolution up to 5 μs; The right panel represents the data collected with 3.7 μs time resolution from 5 μs up to 100 ms.



SI Figure 8. Solvent contribution subtraction from the insulin T-jump data collected at initial temperature of 35 ℃. The left panel represents the data collected with 7 ns time resolution up to 5 μs; The right panel represents the data collected with 3.7 μs time resolution from 5 μs up to 100 ms.



SI Figure 9. Hydrodynamic response of the solvent extracted from the T-jump measurements with initial temperature of 25 ℃.



SI Figure 10. Hydrodynamic response of the solvent extracted from the T-jump measurements with initial temperature of 30 ℃.

SI.6 Global analysis of the protein signal

The global analysis procedure for TRXS data performed here is described in the literature.[12–17] Briefly, the procedure can be described as follows. First, singular value decomposition (SVD) of the data is performed in order to determine the number of independent components (vectors) required to describe the entire data set. Due to the difference in noise level between data collected using 100 ps x-ray pulse duration (time delays shorter than 5 μs) and data collected with 3.7 μs x-ray pulse duration (time delays longer than 5 μs), each time series was treated independently. Using the standard SVD formula decomposition for short and long time series data will yield

$$\Delta S_{protein}(q, t < 5\mu s) = U_S S_S V_S^T$$

$$\Delta S_{protein}(q, t > 5\mu s) = U_L S_L V_L^T$$

Where indexes 'S' and 'L' stand for short and long time series; S_S and S_L are the matrices which diagonal elements corresponding to singular values; $U_S(q)$ and $U_L(q)$ represent the scattering signal associated with the components; $V_S(t)$ and $V_L(t)$ reflect kinetics of the components. The datasets and corresponding decompositions for 25, 30 and 35 ^{o}C initial temperatures are shown in SI figures 11-16. Additionally, in order to avoid bias in retrieved kinetic constants due to solvent hydrodynamics, SVD was performed on the data collected after 50 ns, i.e. after sample expansion has taken place. Overall, the comparison of low rank approximation with three components and the data demonstrates that in case of all initial temperatures, early time delay data shows presence of three components. Similarly, the late time delay data also have shown the presence of three components. Since short and long time series are overlapping at 5 us, one of the underlying transient states is shared between the two datasets. Based on that we conclude that each dataset for each initial temperature can be described with five transient species. For convenience, in the following we will consider only three components from each decomposition.

The next step of the analysis is to transform the SVD components, which do not have physical meaning, to species associated difference scattering signals and corresponding kinetics. This can be accomplished through the fitting of SVD component kinetics V_S and V_L with theoretical models

$$V_{S,theory} = C_S R_S$$

$$V_{L,theory} = C_L R_L$$

where C_S and C_L are the time dependent concentrations of the species; R_S and R_L are rotation matrices defined from the fitting of the component kinetics V_S and V_L by C_S and C_L, respectively. After that the species associated difference scattering patterns can be calculated as

$$U_{S,theory} = U_S S_S R_S^T$$

$$U_{L,theory} = U_L S_L R_L^T$$

It was found that the data can be successfully described by a set of consecutive processes which, following the notation from the main text, can be written as follows:

$$D \xrightarrow{T-jump} D^* \rightarrow D_1 \rightarrow D_2 \xrightarrow{dissociation} 2M \xrightarrow{cooling} 2M' \xrightarrow{reassociation} D$$

Early time delay data was analyzed with the kinetic model $D \xrightarrow{T-jump} D^* \rightarrow D_1 \rightarrow D_2$, while the later time delays were analyzed using second part of the model $D_2 \xrightarrow{dissociation} 2M \xrightarrow{cooling} 2M' \xrightarrow{reassociation} D$. We note that the transition $D \xrightarrow{T-jump} D^*$ was considered quasi instantaneous as the signal from D^* arises together with the temperature increase due to the laser heating. All the processes, except for the re-association, have found to agree with first order kinetic equations. The re-association process was found to be well described with the stretched exponential decay law. Overall, the following system of equations was employed to describe the data:

$$\frac{\partial[D^*]}{\partial t} = -\frac{[D^*]}{\tau_1}$$

$$\frac{\partial[D_1]}{\partial t} = \frac{[D^*]}{\tau_1} - \frac{[D_1]}{\tau_2}$$

$$\frac{\partial[D_2]}{\partial t} = \frac{[D_1]}{\tau_2} - \frac{[D_2]}{\tau_3}$$

$$\frac{\partial[2M]}{\partial t} = \frac{[D_2]}{\tau_3} - \frac{[2M]}{\tau_4}$$

$$\frac{\partial[2M']}{\partial t} = \frac{[2M]}{\tau_4} - \frac{[2M']}{\tau_5}\beta\left(\frac{t}{\tau_5}\right)^{\beta-1}$$

The following initial conditions were applied: $[D^*](0) = 1$; $[D_1](0) = 0$; $[D_2](0) = 0$; $[2M](0) = 0$; $[2M'](0) = 0$. The solutions of the set of equations can be ascribed as $C_S = \{D^*, D_1, D_2\}$ and

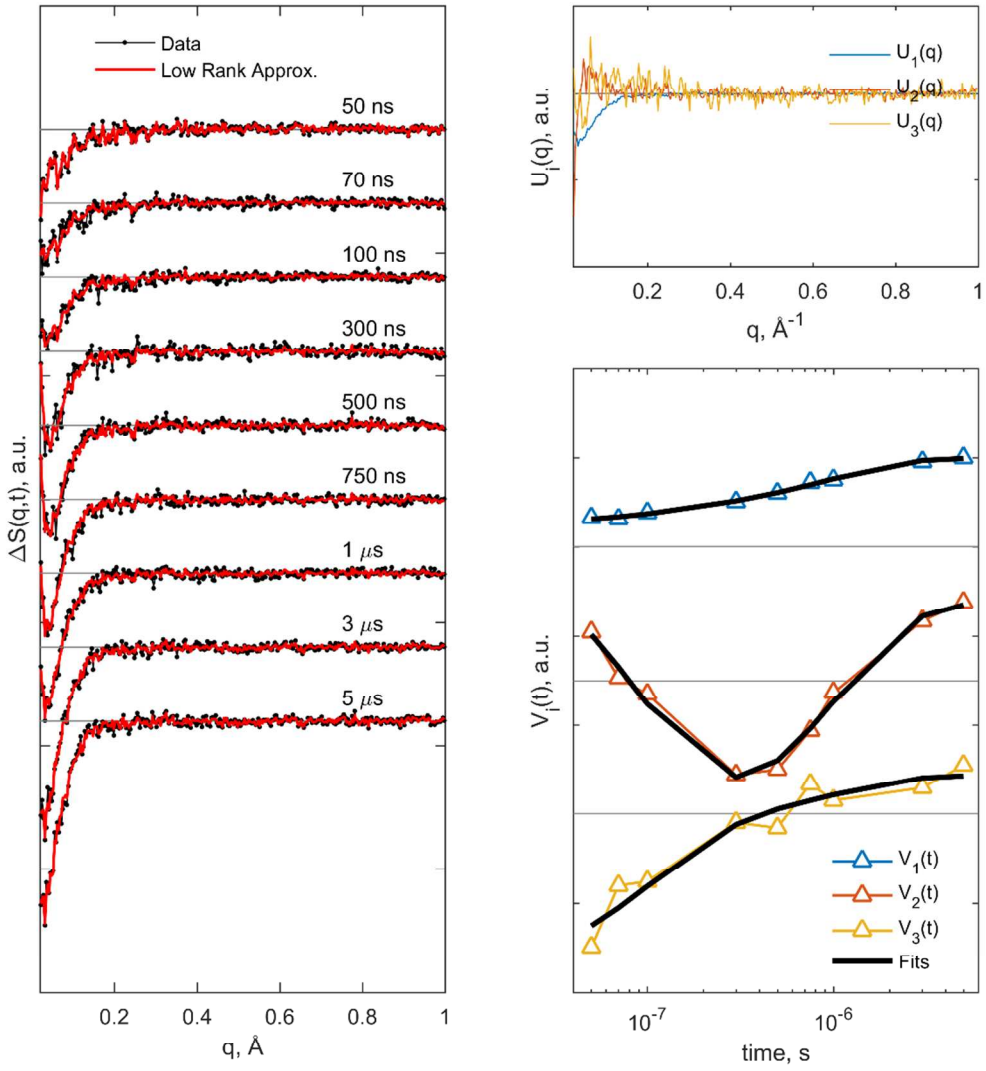
$C_L = \{D_2, 2M, 2M'\}$. In order to find the best fit set of kinetic parameters $\{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \beta\}$, the solutions of the kinetic equations have to fitted to $V_S(t)$ and $V_L(t)$, by using following estimators:

$$\chi_S^2 = \sum_t (C_S R_S - V_S)^T S_S (C_\alpha R_S - V_S),$$

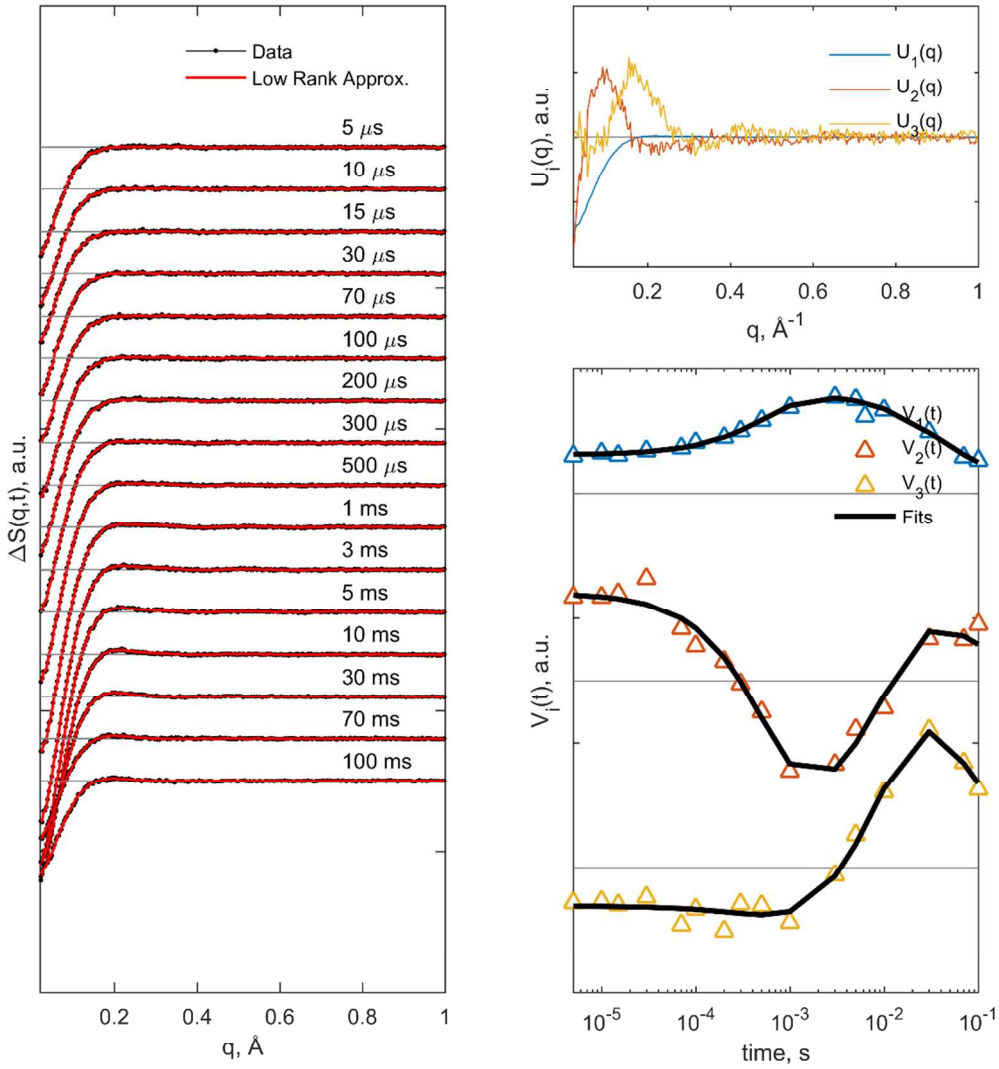
$$\chi_L^2 = \sum_t (C_L R_L - V_L)^T S_L (C_L R_L - V_L),$$

Where multiplication by singular values S_S and S_L is introduced in order to take into account different noise levels of different components.

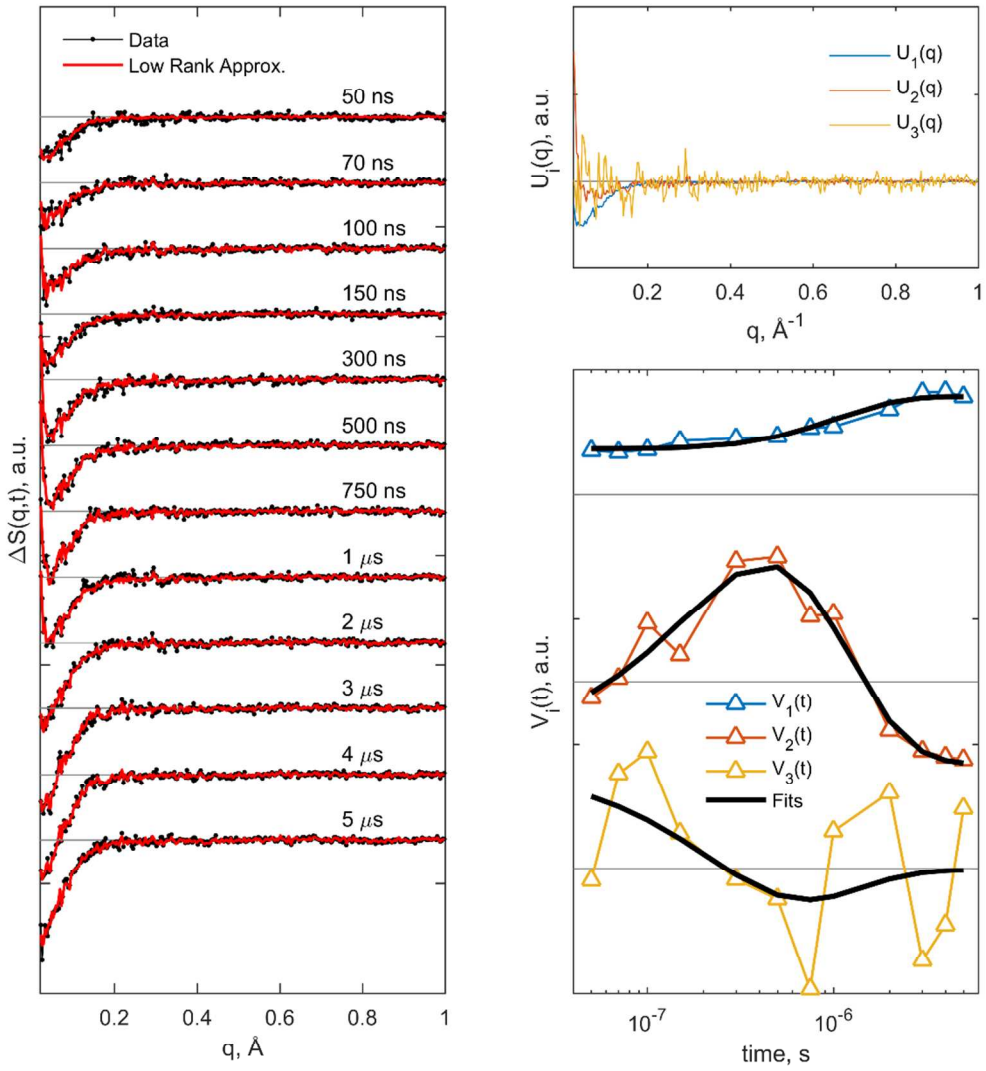
SI Figures 11-16 show the results of the fitting of V_S and V_L in each dataset collected at initial temperature of 25, 30 and 35 ^{o}C. The resulting species associated kinetics for 25 ^{o}C and 30 ^{o}C initial temperatures are shown in SI figure 17-18 (35 ^{o}C can be found in the main text); the species associated signals for all temperatures are shown in SI figure 19. Overall, the species associated signals obtained from different temperatures are in good agreement; slight differences can indicate the uncertainty associated with the global fitting procedure, as well as possible temperature dependence of the transient species structures. The kinetic parameters for each temperature are summarized in SI table 2. We note that the time constants for $D^* \rightarrow D_1$ and $D_1 \rightarrow D_2$ processes are not strongly affected by the temperature. We suggest that this transitions are at their 'speed limit' of corresponding protein structural rearrangements. The time constant of $D^* \rightarrow D_1$ for 25 ^{o}C is significantly smaller than the corresponding values from 30 ^{o}C and 35 ^{o}C. This is suspected to be due to possible interferences with hydrodynamic processes. On the other hand, the dissociation process $D_2 \rightarrow 2M$ lifetime systematically decrease with the temperature, following the expected temperature dependence of this process. The constants we extract for the final transient state which appears during the system cool down do not represent physically meaningful values as the process is convoluted with the cooling process itself. Therefore lifetimes for the final transient state are not comparable.



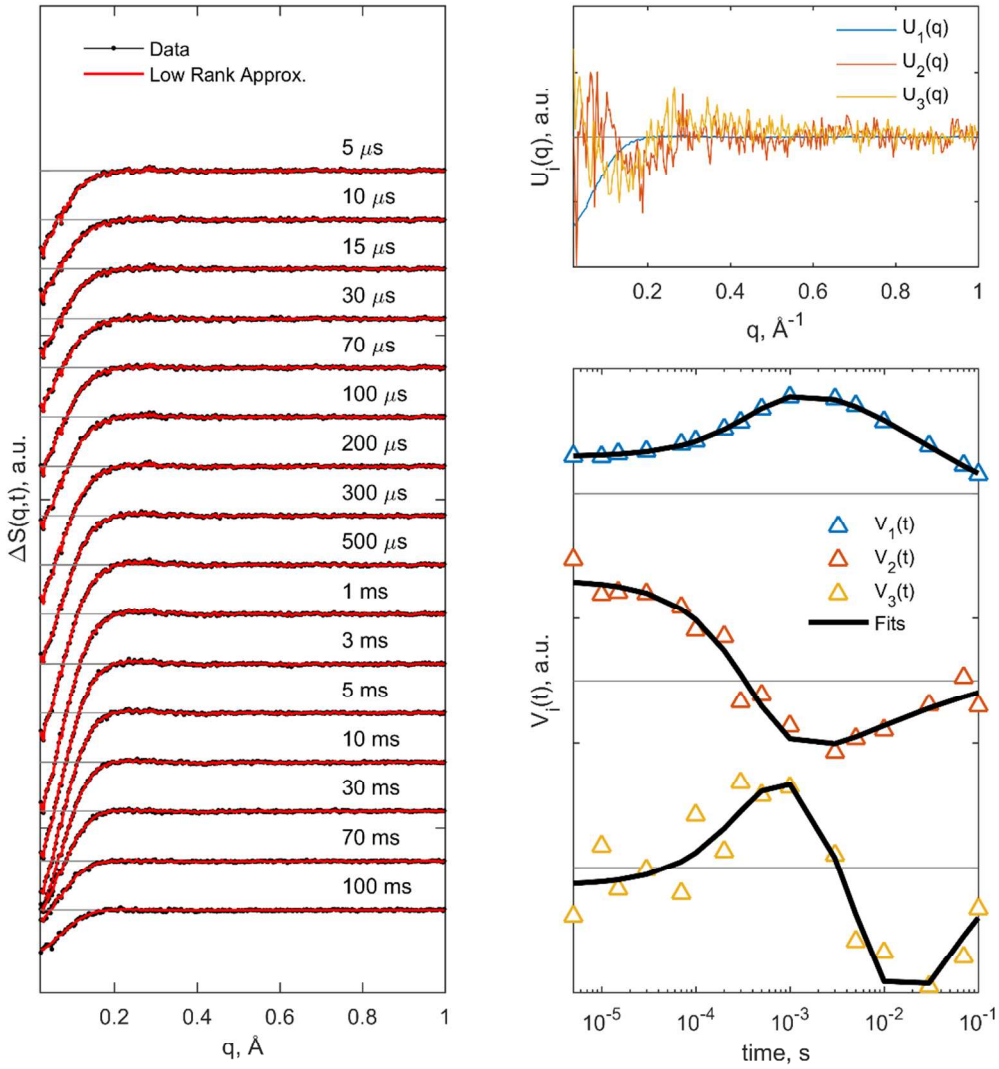
SI figure 11. (Left panel) comparison of the short time series data (t<5 μs) collected at 25 ℃ and low rank approximation using three SVD components; (top-right panel) left singular vectors U(q); (bottom-right panel) component kinetics V(t) and corresponding fits.



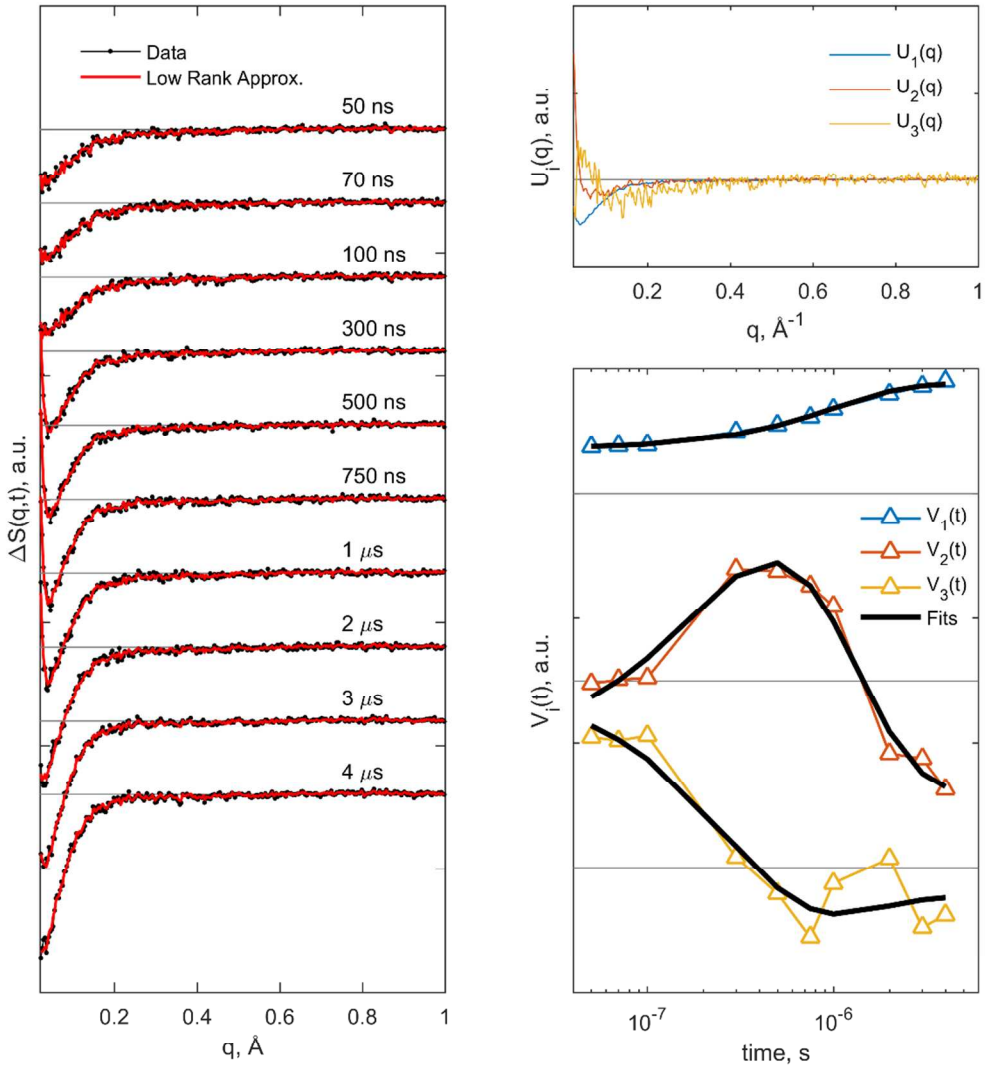
SI figure 12. (Left panel) comparison of the long time series data (t>5 μs) collected at 25 ℃ and low rank approximation using three SVD components; (top-right panel) left singular vectors U(q); (bottom-right panel) component kinetics V(t) and corresponding fits.



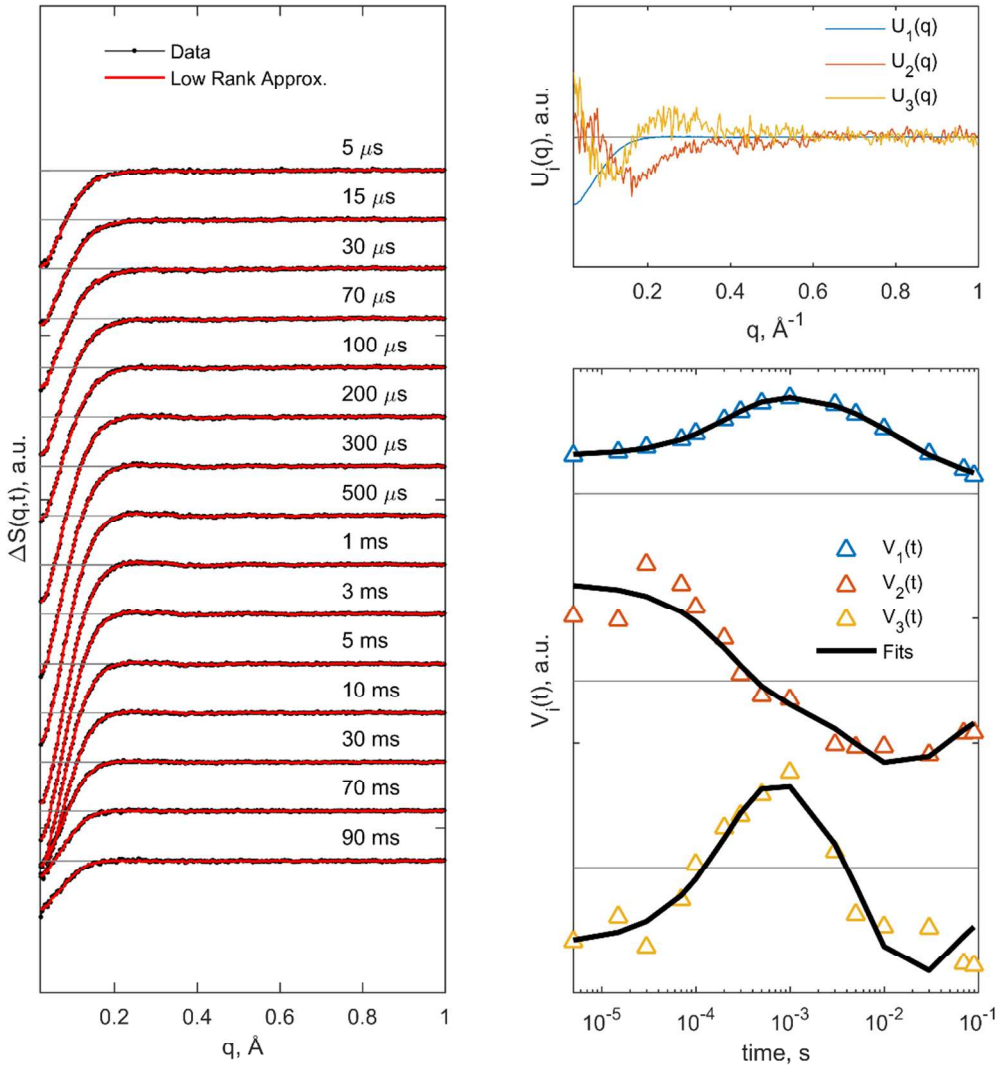
SI figure 13. (Left panel) comparison of the short time series data (t<5 μs) collected at 30 ℃ and low rank approximation using three SVD components; (top-right panel) left singular vectors U(q); (bottom-right panel) component kinetics V(t) and corresponding fits.



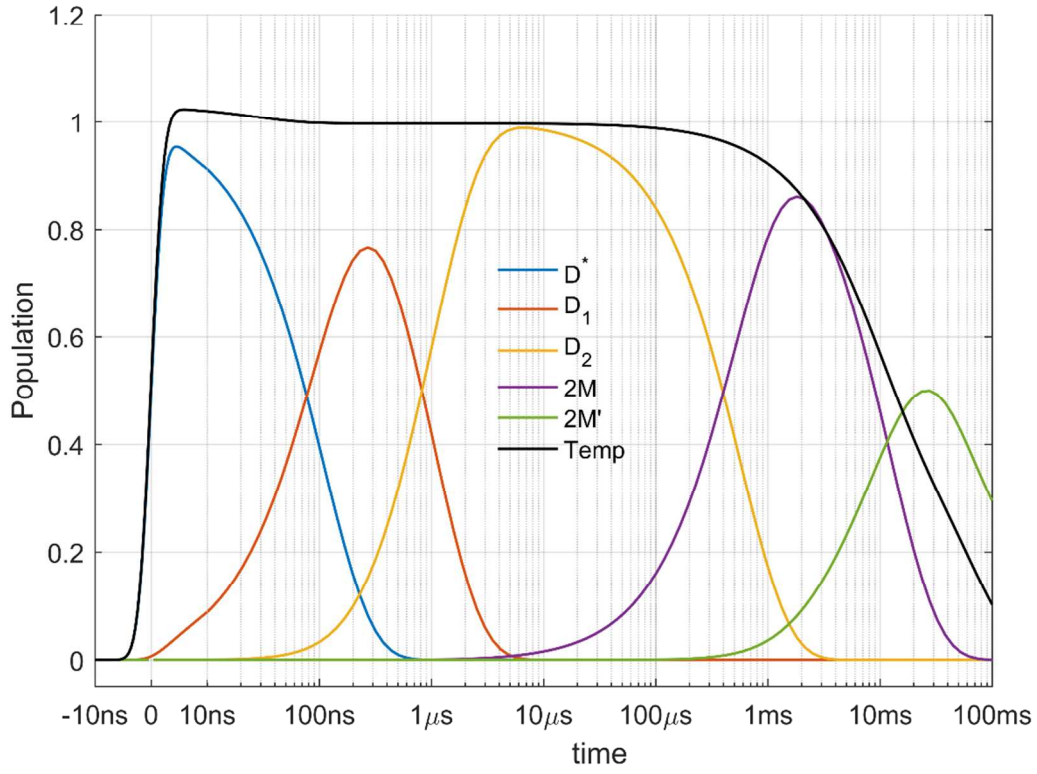
SI figure 14. (Left panel) comparison of the long time series data (t>5 μs) collected at 30 ℃ and low rank approximation using three SVD components; (top-right panel) left singular vectors U(q); (bottom-right panel) component kinetics V(t) and corresponding fits.



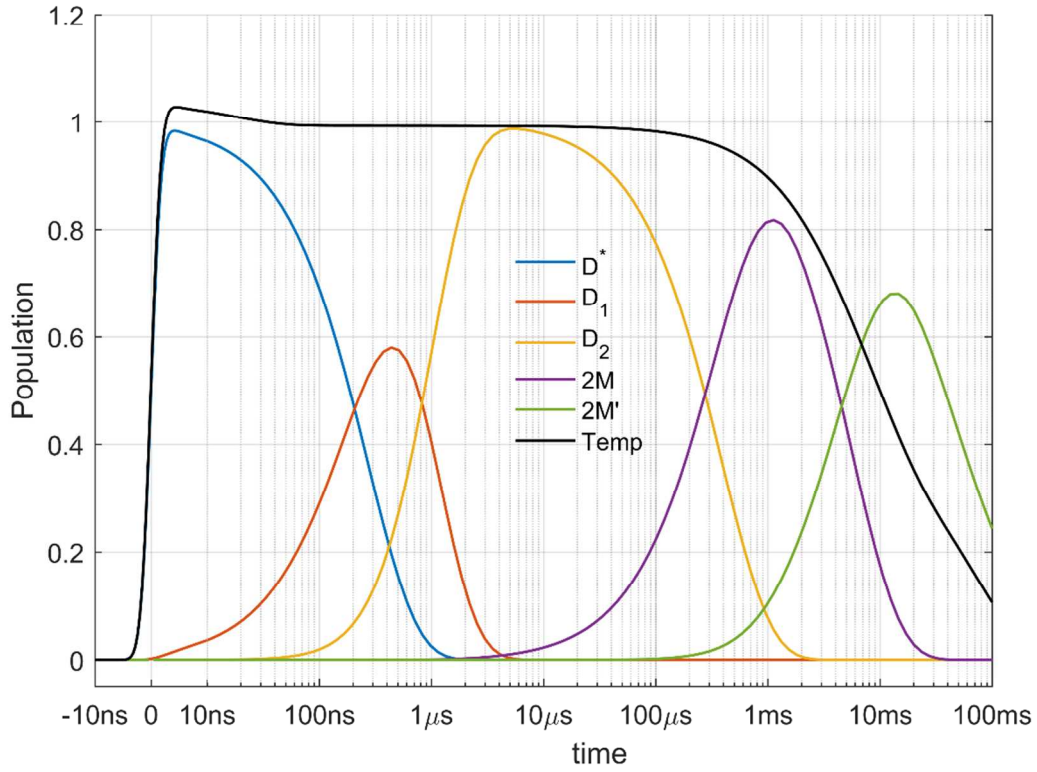
SI figure 15. (Left panel) comparison of the short time series data (t<5 μs) collected at 35 ℃ and low rank approximation using three SVD components; (top-right panel) left singular vectors U(q); (bottom-right panel) component kinetics V(t) and corresponding fits.



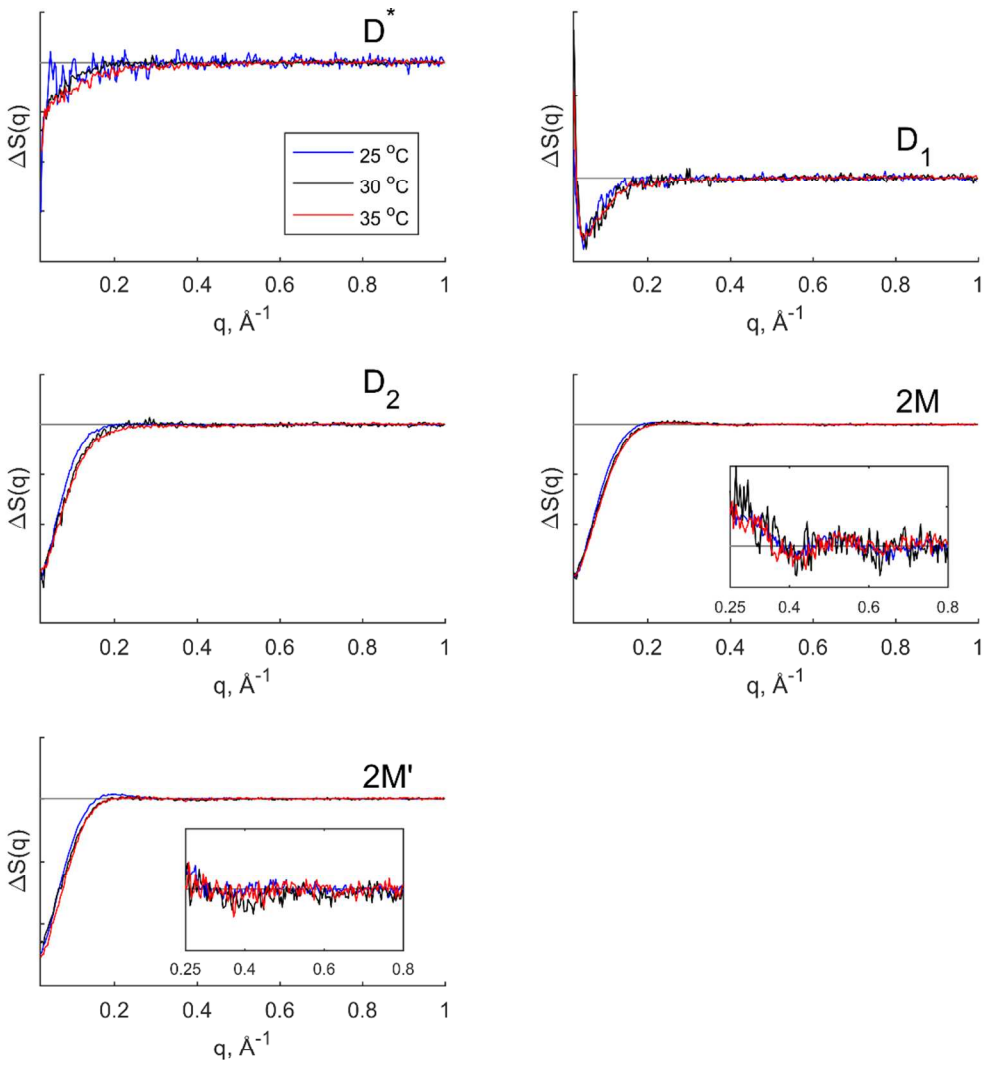
SI figure 16. (Left panel) comparison of the long time series data (t>5 μs) collected at 35 ℃ and low rank approximation using three SVD components; (top-right panel) left singular vectors U(q); (bottom-right panel) component kinetics V(t) and corresponding fits.



SI Figure 17. Kinetics extracted from the global analysis for the data collected at initial temperature of 25 ℃.



SI Figure 18. Kinetics extracted from the global analysis for the data collected at initial temperature of 30 ℃.



SI Figure 19. Species associated difference scattering patterns extracted from the global analysis for each temperature. Each figure is marked with the corresponding species and all the curves have the same color code following the legend in top-left subpanel.

Initial Temperature	$D^* \rightarrow D_1$	$D_1 \rightarrow D_2$	$D_2 \rightarrow 2M$	$2M \rightarrow 2M'$	$2M' \rightarrow D$	
	τ_1	τ_2	τ_3	τ_4	τ_5	β
25 °C	110 ns	1 µs	570 µs	12 ms	180 µs	0.16
30 °C	270 ns	820 ns	390 µs	5 ms	33 ms	0.53
35 °C	310 ns	900 ns	240 µs	7 ms	920 µs	0.15

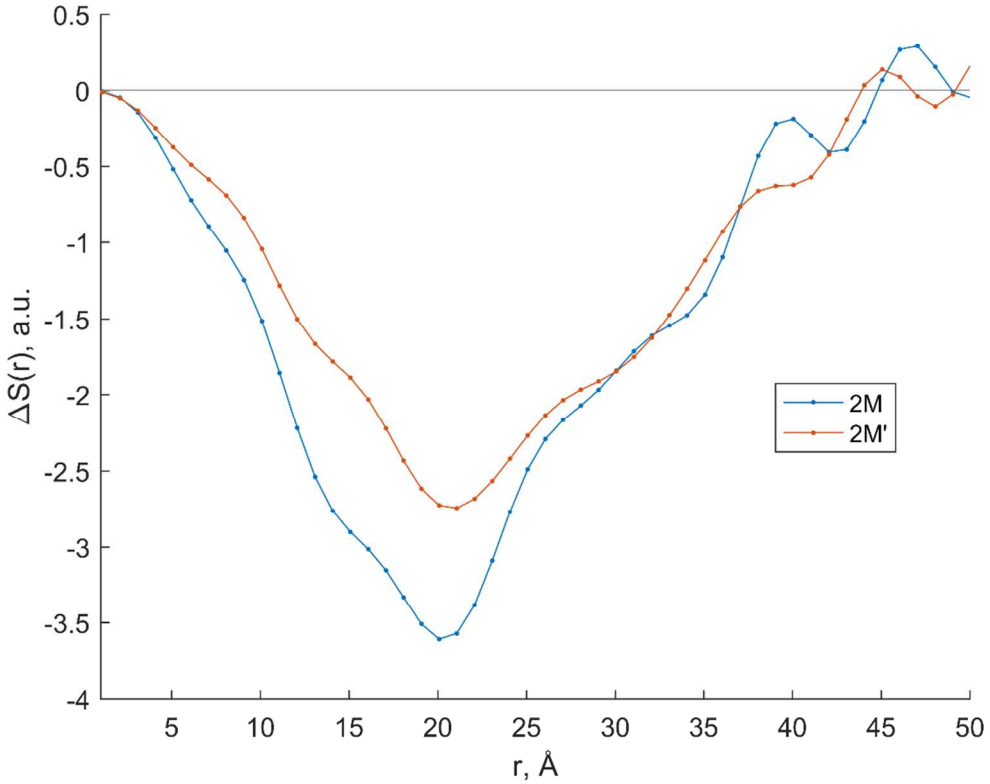
SI Table 2. Kinetic parameters obtained from the global analysis.

SI 7. Pair distribution function of 2M and 2M' states

In order to get insight into the structural difference between 2M and 2M' species, difference radial distribution function (dRDF) was calculated based on the species associated difference signals obtained from the global analysis of the data recorded at initial temperature of 35°C. The standard formula have been used:

$$\Delta S(r) = \int_{q_{min}}^{q_{max}} \mathrm{q}\Delta S(q)\,\sin(qr)dq,$$

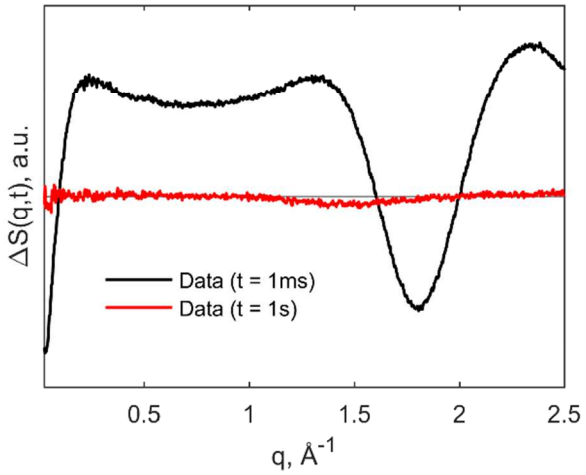
where r is the interatomic distance; q_{min} and q_{max} were taken as 0.02 and 1 Å^{-1}, i.e the span of the measured protein signal. The resulting dRDF is shown in SI figure 20. Overall shape is represented by a negative feature reflecting the loss of electron density due to dissociation process. The main difference between the two species, 2M and 2M', is observed in the region of $r < 25\,\text{Å}$, which reflects the difference in the structure of the insulin monomer structure. Specifically, the smaller amplitude of 2M' in this region indicates that the structure of the 2M' is closer to the one in the (ground state) dimer confirmation (see main text).



SI Figure 20. dRDF of 2M and 2M' species.

SI 8. Comparison of scattering differential signal at different time delays

SI figure 21 shows the comparison of the raw differential signal at 1ms in comparison to 1s. The comparison clearly shows that by 1s, both heating and protein signals have largely disappeared as the system has returned to the ground state.



SI Figure 21. Comparison of protein signal at 1ms and 1s time delays.

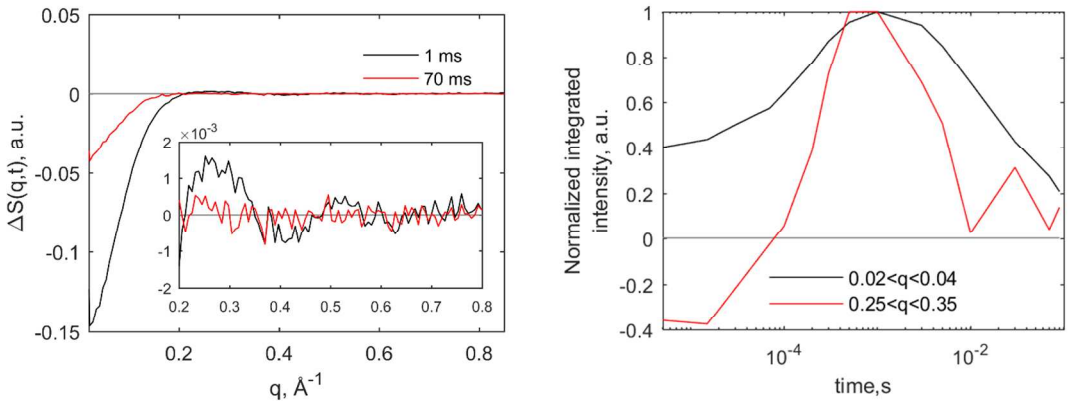
SI 9. Comparison of the WAXS signal from static SAXS differences with theoretical curves

Theoretical curves were calculated from PDB entries 2A3G (dimer)[18] and 2JV1 (monomer)[19] and the difference between the theoretical curves was compared to the 2M species associated signal, as well as differential signal from static data at different temperatures. As the static data used for Guinier and Kratky plot analysis was deemed to be too noisy for WAXS region analysis, temperature dependent static data was collected at the BioCARS beamline and used to calculate the WAXS differences.

SI 10. Comparison of the WAXS signal at different time delays and integrated signal analysis

SI figure 22 shows comparison of the signal at 1ms and 70ms, representing the maximum populations of the 2M and 2M' states. The inset shows the WAXS region magnified to better show the features disappearing at long time delays, indicating the formation of the 2M' state. The figure also shows the

comparison of SAXS and WAXS integrated region to show that the WAXS signal (0.25<q<0.35 $Å^{-1}$), proportional to the 2M population after 100 μs, decays faster compared to the SAXS signal (0.02<q<0.04 $Å^{-1}$) which, at longer time delays, is proportional to the 2M' population. These dynamics support the assignment of the 2M' to the induced fit step required to form the dimers as discussed in the main text.



SI Figure 22. (left) Comparison of protein signal at 1ms and 70 ms, showing the disappearance of the WAXS feature for 35 ℃ T-jump. (right) Integration of WAXS and SAXS region for comparison of dynamics.

SI References

(1) Ganim, Z.; Jones, K. C.; Tokmakoff, A. Insulin Dimer Dissociation and Unfolding Revealed by Amide I Two-Dimensional Infrared Spectroscopy. *Phys. Chem. Chem. Phys.* **2010**, *12* (14), 3579–3588.

(2) Zhang, X.-X.; Jones, K. C.; Fitzpatrick, A.; Peng, C. S.; Feng, C.-J.; Baiz, C. R.; Tokmakoff, A. Studying Protein–Protein Binding through T-Jump Induced Dissociation: Transient 2D IR Spectroscopy of Insulin Dimer. *J. Phys. Chem. B* **2016**, *120* (23), 5134–5145.

(3) Weigand, S. J.; Keane, D. T. DND-CAT's New Triple Area Detector System for Simultaneous Data Collection at Multiple Length Scales. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* **2011**, *649* (1), 61–63.

(4) Cammarata, M.; Levantino, M.; Schotte, F.; Anfinrud, P. a; Ewald, F.; Choi, J.; Cupane, A.; Wulff, M.; Ihee, H. Tracking the Structural Dynamics of Proteins in Solution Using Time-Resolved Wide-Angle X-Ray Scattering. *Nat. Methods* **2008**, *5* (10), 881–886.

(5) Graber, T.; Anderson, S.; Brewer, H.; Chen, Y. S.; Cho, H. S.; Dashdorj, N.; Henning, R. W.; Kosheleva, I.; MacHa, G.; Meron, M.; et al. BioCARS: A Synchrotron Resource for Time-Resolved X-Ray Science. *J. Synchrotron Radiat.* **2011**, *18* (4), 658–670.

(6) Ma, H.; Wan, C.; Zewail, A. H. Ultrafast T-Jump in Water: Studies of Conformation and Reaction Dynamics at the Thermal Limit. *J. Am. Chem. Soc.* **2006**, *128* (19), 6338–6340.

(7) Uversky, V. N.; Garriques, L. N.; Millett, I. S.; Frokjaer, S.; Brange, J.; Doniach, S.; Fink, A. L. Prediction of the Association State of Insulin Using Spectral Parameters. *J. Pharm. Sci.* **2003**, *92*

(4), 847–858.

(8) Strickland, E. H.; Mercola, D. Near-Ultraviolet Tyrosyl Circular Dichroism of Pig Insulin Monomers, Dimers, and Hexamers. Dipole-Dipole Coupling Calculations in the Monopole Approximation. *Biochemistry* **1976**, *15* (17), 3875–3884.

(9) Nielsen, L.; Khurana, R.; Coats, A.; Frokjaer, S.; Brange, J.; Vyas, S.; Uversky, V. N.; Fink, A. L. Effect of Environmental Factors on the Kinetics of Insulin Fibril Formation: Elucidation of the Molecular Mechanism †. *Biochemistry* **2001**, *40* (20), 6036–6046.

(10) Ahmad, A.; Millett, I. S.; Doniach, S.; Uversky, V. N.; Fink, A. L. Partially Folded Intermediates in Insulin Fibrillation. *Biochemistry* **2003**, *42* (39), 11404–11416.

(11) Cammarata, M.; Lorenc, M.; Kim, T. K.; Lee, J. H.; Kong, Q. Y.; Pontecorvo, E.; Lo Russo, M.; Schiró, G.; Cupane, A.; Wulff, M.; et al. Impulsive Solvent Heating Probed by Picosecond X-Ray Diffraction. *J. Chem. Phys.* **2006**, *124* (12), 124504.

(12) Oang, K. Y.; Yang, C.; Muniyappan, S.; Kim, J.; Ihee, H. SVD-Aided Pseudo Principal-Component Analysis: A New Method to Speed up and Improve Determination of the Optimum Kinetic Model from Time-Resolved Data. *Struct. Dyn.* **2017**, *4* (4), 44013.

(13) Cho, H. S.; Dashdorj, N.; Schotte, F.; Graber, T.; Henning, R.; Anfinrud, P. Protein Structural Dynamics in Solution Unveiled via 100-Ps Time-Resolved X-Ray Scattering. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (16), 7281–7286.

(14) Moffat, K. The Frontiers of Time-Resolved Macromolecular Crystallography: Movies and Chirped X-Ray Pulses. *Faraday Discuss.* **2003**, *122*, 65–77.

(15) Schmidt, M.; Pahl, R.; Srajer, V.; Anderson, S.; Ren, Z.; Ihee, H.; Rajagopal, S.; Moffat, K. Protein Kinetics: Structures of Intermediates and Reaction Mechanism from Time-Resolved X-Ray Data. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (14), 4799.

(16) Ihee, H.; Rajagopal, S.; Srajer, V.; Pahl, R.; Anderson, S.; Schmidt, M.; Schotte, F.; Anfinrud, P. A.; Wulff, M.; Moffat, K. Visualizing Reaction Pathways in Photoactive Yellow Protein from Nanoseconds to Seconds. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (20), 7145–7150.

(17) Cho, H. S.; Schotte, F.; Dashdorj, N.; Kyndt, J.; Henning, R.; Anfinrud, P. A. Picosecond Photobiology: Watching a Signaling Protein Function in Real Time via Time-Resolved Small- and Wide-Angle X-Ray Scattering. *J. Am. Chem. Soc.* **2016**, *138* (28), 8815–8823.

(18) Smith, G. D.; Pangborn, W. A.; Blessing, R. H. The Structure of T 6 Bovine Insulin. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2005**, *61* (11), 1476–1482.

(19) Bocian, W.; Sitkowski, J.; Bednarek, E.; Tarnowska, A.; Kawęcki, R.; Kozerski, L. Structure of Human Insulin Monomer in Water/acetonitrile Solution. *J. Biomol. NMR* **2008**, *40* (1), 55–64.