# Supplementary information:

# Insulin Hexamer Dissociation Dynamics Revealed by Photoinduced T-jumps and Time-Resolved X-Ray Solution Scattering

Dolev Rimmerman,<sup>a</sup> Denis Leshchev, <sup>a</sup> Darren J. Hsu, <sup>a</sup> Jiyun Hong, <sup>a</sup> Baxter Abraham, <sup>b</sup> Irina Kosheleva, <sup>c</sup> Robert Henning, <sup>c</sup> Lin X. Chen\*

E-mail: lchen@anl.gov

#### SI.1 Experimental Methods

The sample preparation method and details of the TRXSS T-jump instrument have been previously published and are available in SI.1 of SI Reference [1].<sup>1</sup> In brief, TRXSS experiments were carried out at BioCARS 14-ID-B beamline at the Advanced Photon Source (APS), the full details of the beamline, including chopper setup, is detailed in SI Reference [2].<sup>2</sup> The sample was excited by 7 ns laser pulses with a 1.443 nm wavelength to generate a T-jump in the buffer. The laser beam was focused on a capillary with the sample solution to an elliptical spot 65  $\mu$ m x 350  $\mu$ m delivering energy of ~1 mJ in each pulse, leading to a power density of ~56 mJ/mm2 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  $\mu$ m by 35  $\mu$ m. The chopper-shutter system described in literature<sup>2</sup> was employed to reduce the repetition rate of the synchrotron source operating in 24 bunch mode. The chopper opening time was used to modify the time resolution of the experiment at the cost of x-ray flux. For the late time delay data (delays longer than 5  $\mu$ s) the opening time was calibrated to 3.7  $\mu$ s therefore generating x-rays pulses comprised of 24 electron micro-bunches in the ring. Similarly, for short time delays (delays shorter than 5  $\mu$ s) the opening time was chosen to generate x-ray pulses 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.

During data collection, the sample solution was loaded into the custom built capillary flow cell system using a syringe pump (a full description, including an image of the setup is available in SI Reference [1]).<sup>1</sup> The 0.70 mm diameter open ended capillary was held on a temperature controlled capillary holder constructed from Aluminum Nitride (AIN) 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 AIN holder and the bottom of the capillary was translated horizontally in steps of 250 µm after each pair of pump-probe pulses such that a fresh spot on the capillary was illuminated each time. 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. 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 AIN holder as well as a thermistor mounted at the top of the capillary away from the

<sup>&</sup>lt;sup>a.</sup> Department of Chemistry, Northwestern University, Evanston, Illinois 60208, USA

E-mail: I-chen@northwestern.edu

<sup>&</sup>lt;sup>b.</sup> Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, USA

<sup>&</sup>lt;sup>c.</sup> Center for Advanced Radiation Sources, The University of Chicago, Illinois 60637, USA

<sup>&</sup>lt;sup>d.</sup> Chemical Sciences and Engineering Division, Argonne National Laboratory, Argonne, Illinois 60439, USA

exposed region. The estimation of the degree of temperature jump obtained in the measurement (8 °c) was published previously<sup>1</sup>.

# SI.2 Solvent subtraction.

Subtraction of the solvent contribution to the TRXSS signals was performed according to the standard procedures published before<sup>1</sup>. The separately measured buffer data recorded at 15 ns and 5  $\mu$ s were used to calculate the characteristic liquid response to isochoric changes in temperature and isothermal changes in density, respectively. This contribution, along with capillary contribution, were fitted to the high-q portion of the data (q $\geq$ 0.5 Å<sup>-1</sup>) as shown in Figure S1. This contribution was then subtracted to obtain the protein-only TRXSS signals analyzed and discussed in the main text.



Figure S1. Solvent contribution fitting to the high-q portion of the short time delay (<10  $\mu$ s, left) and long time delay (>10  $\mu$ s, right) data.

# SI.3 SVD analysis of the TRXSS data.

Singular value decomposition (SVD) analysis was carried out by previously reported procedures.<sup>1</sup> As previously described, the data collected for short time delays (t < 10  $\mu$ s) and for long time delays (t ≥ 10  $\mu$ s, LTD) was analyzed separately due to difference in the signal-to-noise ratio (see note S4 for further details). Qualitative inspection of the SVD of the data recorded for t < 10  $\mu$ s shows two significant components, whereas the SVD of data at t ≥ 10  $\mu$ s shows three significant components. Quantitatively these observations are supported by inspection of the autocorrelation values of the left singular vectors.<sup>3</sup> For the short time scales (t < 10  $\mu$ s) autocorrelation values of the first two components are ~0.95 and ~0.6 which is significantly larger than those for other components having values ≤0.3. For the long time scales the autocorrelation values for the first three components are ~0.95, ~0.9 and 0.8, which is much higher than the values for the higher order components with autocorrelations <0.4. From the SVD analysis, and the protein associated TRXSS signals, it is apparent that the last time delay in the short time-series (3  $\mu$ s) and the first time delay of the long time series (10  $\mu$ s) are nearly identical, and therefore one state is shared between the two datasets. Therefore, we conclude that the entire data recorded for all time delays reflects a presence of four species. To estimate the time scales of transitions

between species, we have performed a standard model-free kinetic fitting of the both data sets.<sup>4–6</sup> First, we fitted the long time series data with three exponential terms. Second, we fitted short time delay data with two exponentials; time scale of one of the exponentials was fixed to the shortest time scale retrieved from long time delay ( $t \ge 10 \mu$ s) dataset. The SVD decomposition, autocorrelation values for SVD components and kinetic fitting for both data sets are shown in Figures S2 and S3. The best-fit time scales are as follows:  $0.9 \pm 0.1 \mu$ s,  $730 \pm 20 \mu$ s,  $20.5 \pm 0.7 \mu$ s and  $300 \pm 15 \mu$ s.



Figure S2. Left panel: protein associated TRXSS data recorded for time delays less than 10 µs and corresponding fit with kinetic model. Right panels (from top to bottom): singular value resulting from the SVD; autocorrelation values for the components; left singular vector resulting from the SVD; right singular vector resulting from the SVD (dots) and results of kinetic fitting (lines).



Figure S3. Left panel: protein associated TRXSS data recorded for time delays more than 10 µs and corresponding fit with kinetic model. Right panels (from top to bottom): singular value resulting from the SVD; autocorrelation values for the components; left singular vector resulting from the SVD; right singular vector resulting from the SVD (dots) and results of kinetic fitting (lines).

#### SI.4 SVD of merged dataset

As mentioned above, the SVD analysis of the short and long-time delay data was performed separately due to abrupt differences in the signal-to-noise ratio between the two methods of measurements. As expected, attempts to perform SVD analysis on the merged dataset resulted in the appearance of components which are mainly attributed to differences in noise level (Figure S4). The qualitative inspection of the left singular vectors Ui(q), resulting from the SVD of the merged dataset, show that only the first three components carry signal whereas the following components appear to correspond to the noise. While the signal components in the merged dataset seem to correlate with the shapes of the first three Ui(q) vectors obtained from the SVD of late time delay data, the signal to noise ratio for these components is substantially lower. The inspection of the time-dependent amplitudes of the second and third components, the right singular vector Vi(t), shows the presence of abrupt change in the signal to

noise ratio at t=10  $\mu$ s – the time delay of separation of the two datasets. Additionally, at short time delays t<10  $\mu$ s, the time dependent amplitude of the third component V3(t) appears to represent only random noise and does not carry robust kinetic information. Finally, the forth and the fifth components, which can be regarded as noise based on the analysis of the left singular vectors U4,5(q), demonstrate kinetics that are strongly correlated to the first three components at time delays > 10  $\mu$ s. These observations suggest that the results of the SVD represent a mixture of the noise and the signals, which makes it difficult to extract kinetic information from the data. Similar observations have been made in the literature, and it has been suggested that the standard SVD procedure works most robustly on the data with evenly distributed noise.<sup>3</sup> For these reasons, the main findings of this work are based on the SVD results obtained from the analysis of the short and long time delay data separately.



Figure S4. Results of SVD performed on the merged dataset containing both data for short and long time delays. (left) The time series of the data. (right-top) Singular values. (right-middle) 1-5<sup>th</sup> left singular vectors. (right-bottom) 1-5<sup>th</sup> right singular vectors. The singular values and vectors in right panels are color-coded according to the legend depicted on the right from right-middle panel.

#### SI.5 Choice of the time delays for LCA

As discussed in the main text, it is possible to choose the key time delays which will reflect the presence of majority of excited population in the intermediate states based on the kinetics retrieved from the SVD analysis, as well as based on the overall progression of the integrated TRXSS signals. In the main text we discuss the population dynamics based on the time delays 10 ns, 20  $\mu$ s, 2 ms and 100 ms. This is justified as, for example, 2ms is << 20.5 ms and >> 0.9  $\mu$ s time scales retrieved from the kinetic analysis, and therefore this choice reflects the time delay when the majority of the population is in TD state (see main text for the state assignment). Similar arguments can be done for other chosen time delays. We note that the choice of the time delays for LCA is arbitrary as long as the chosen time delays satisfy the condition of being sufficiently different from the time scales retrieved from the kinetic analysis. To verify that the results of LCA are independent of the specific choice of time delays we performed LCA of the data with different sets of time delays. The resulting population dynamics from different LCA runs are shown in Figure S3. We observe that the population dynamics do not change substantially and are very similar to those presented in the main text, which indicates the robustness of the analysis.



Figure S5. Population dynamics retrieved from LCA by using different sets of key time delays. The data points are the results of the LCA analysis, whereas lines represent guides for the eye obtained by using Savitzky-Golay linear three point filter. The color code of the transient populations is the same to the one in Figure 2 in main text. The LCA results obtained with different sets are shown by using different symbols/lines as shown in the legend (not color coded).

## SI.6 Comparison to PDB structures

Comparison of time-resolved and static differences to crystal structure data is often helpful in understanding and corroborating the origin of differences between proteins under different conditions. For this reason, we have compared the difference signals from our static data and TRXSS experiments with difference signals derived from PDB structures of insulin hexamers  $(2AIY)^7$  and dimers  $(2A3G)^8$  using CRYSOL<sup>9</sup>. However, as in the case of our previous study<sup>1</sup>, while the curves are in general agreement with each other in the SAXS region, some differences exist around the isosbestic point  $(0.12-0.16 \text{ Å}^{-1})$ , therefore we can only suggest a qualitative agreement between experimental data and theoretical curve. The discrepancies are ascribed to the fact that the theoretical insulin structure, and particularly its degree of flexibility, is significantly different from the insulin structure in H2O:EtOH solution at pH=2, as also evidenced by the imperfect bell shape in the Kratky plots<sup>1</sup>. Given the known shortcomings of the calculation thoroughly discussed in our previous work, the lack of total agreement between PDB derived differences and experimental data is not unexpected. Future work utilizing more robust calculation techniques, such as MD simulations, may provide better disordered structures for comparison to experimental data.



Figure S6. Comparison of PDB derived difference data with TRXSS and static difference curves.

## **References**

- D. Rimmerman, D. Leshchev, D. J. Hsu, J. Hong, I. Kosheleva and L. X. Chen, Direct Observation of Insulin Association Dynamics with Time-Resolved X-ray Scattering, J. Phys. Chem. Lett., 2017, 8, 4413–4418.
- T. Graber, S. Anderson, H. Brewer, Y.-S. Chen, H. S. Cho, N. Dashdorj, R. W. Henning, I. Kosheleva, G. Macha, M. Meron, R. Pahl, Z. Ren, S. Ruan, F. Schotte, V. Šrajer, P. J. Viccaro, F. Westferro, P. Anfinrud and K. Moffat, BioCARS: a synchrotron resource for time-resolved X-ray science, *J. Synchrotron Radiat.*, 2011, **18**, 658–670.
- E. R. Henry and J. Hofrichter, in *Methods in Enzymology*, Academic Press, 1992, vol. 210, pp. 129–192.
- 4 K. Y. Oang, C. Yang, S. Muniyappan, J. Kim and H. Ihee, 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**, 44013.
- 5 M. Schmidt, R. Pahl, V. Srajer, S. Anderson, Z. Ren, H. Ihee, S. Rajagopal and K. Moffat, Protein kinetics: structures of intermediates and reaction mechanism from time-resolved x-ray data, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 4799.
- M. Cammarata, M. Levantino, F. Schotte, P. a Anfinrud, F. Ewald, J. Choi, A. Cupane, M. Wulff and
  H. Ihee, Tracking the structural dynamics of proteins in solution using time-resolved wide-angle
  X-ray scattering., *Nat. Methods*, 2008, 5, 881–886.
- S. I. O'Donoghue, X. Chang, R. Abseher, M. Nilges and J. J. Led, Unraveling the symmetry ambiguity in a hexamer: calculation of the R6 human insulin structure., *J. Biomol. NMR*, 2000, 16, 93–108.

- 8 G. D. Smith, W. A. Pangborn and R. H. Blessing, The structure of T 6 bovine insulin, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2005, **61**, 1476–1482.
- D. Svergun, C. Barberato and M. H. J. Koch, CRYSOL a Program to Evaluate X-ray Solution
  Scattering of Biological Macromolecules from Atomic Coordinates, J. Appl. Crystallogr., 1995, 28, 768–773.