Supplementary Material (ESI) for Chemical Communications

Direct Observation of Myoglobin Structural Dynamics from 100 picoseconds to 1 microsecond with Picosecond X-ray Solution Scattering

Kyung Hwan Kim, Key Young Oang, Jeongho Kim, Jae Hyuk Lee, Youngmin Kim, Hyotcherl Ihee*

[a] *Center for Time-Resolved Diffraction, Department of Chemistry, Graduate School of Nanoscience & Technology (WCU), KAIST, Daejeon 305-701, Korea.*

Sample preparation

Carbomonoxy myoglobin (MbCO) solution was prepared as follows. Horse heart Mb purchased from Sigma-Aldrich was dissolved in a 100 mM Na-phosphate buffer at pH 7.0 and centrifuged to remove the aggregates. The Mb solution was reduced by adding sodium dithionite (10 mM), bubbled by CO gas, passed through desalting column to remove sodium dithionite, and concentrated to 8 mM. The resulting solution was put into an air-tight vial, bubbled by nitrogen gas, and treated again by adding sodium dithionite (the final concentration of sodium dithionite was set to 10 mM) under a nitrogen atmosphere to make sure the complete reduction. The reduced Mb solution was bubbled again by CO gas for over 30 min to form MbCO. An aliquot of the resulting 8 mM MbCO solution was transferred into a 1 mm X-ray capillary (Hampton Research) and immediately sealed with epoxy to minimize gas exchange while CO gas was purged continuously into the capillary.

Data collection

Time-resolved X-ray solution scattering data were collected at the ID14B BioCARS beamline at the Advanced Photon Source while the storage ring was operated in the standard operating top-up mode at 7 GeV. The electron bunch fill pattern at this mode has 102 mA in 24 single bunches with a nominal current of 4.25 mA and a spacing of 153 nanoseconds between single bunches.

X-ray pulses were generated by passing the electron bunches through two undulators with 23 and 27 mm periods. The X-ray spectrum is peaked at 12 keV with a long wavelength tail and the bandwidth is ~3%. To reduce the data acquisition time, the full spectrum was used without being monochromatized. The blurring effect to the scattering curve due to the polychromaticity with the ~3% bandwidth is not substantial. The 100-ps-long (FWHM) X-ray pulse containing ~10^9 photons was focused by Kirkpatrick-Baez focusing mirrors, trimmed by slits and delivered to the sample position at a spot size of 0.14 × 0.07 mm^2 (horizontal × vertical, FWHM). Single X-ray pulses were isolated from the high-frequency pulse train by a heat-load chopper, a Jülich high-speed chopper and a millisecond shutter. The heat-load chopper positioned upstream of the rest reduces the heat-load to the high-speed chopper. Single X-ray pulses are isolated by the high-speed chopper where a triangular-shaped rotor with a tunnel is spinning at ~1 kHz, a subharmonic of the synchrotron. The repetition rate was futher reduced from 1 kHz down to 10 Hz using a millisecond shutter to allow long time delays and enough time to move the probing spot to a fresh position at every pair of laser and X-ray pulses.

The protein sample was excited by ~30 ps laser pulses at 532 nm generated by the picosecond laser system at the 14IDB beamline. Femtosecond pulse train at 780 nm were generated from a Tsunami Ti:sapphire laser oscillator and was used to seed the picosecond amplifer system (Spectra Physics Spitfire Pro). The output from the amplifer was sent to a TOPAS optical parametric amplifer to convert the wavelength to 532 nm and stretched to ~30 ps by an echelon. The laser pulse of ~40 μJ energy is focused to a spot of 0.65 mm by 0.16 mm size at the sample position, yielding a power density of 0.4 mJ/mm^2. The sample is contained in a sealed capillary with 1-mm diameter and excited by a laser pulse from the top. The sample was maintained at 25 °C with a cold nitrogen stream (Oxford Cryostream). The X-ray pulse probes 0.2 mm deep from the top in perpendicular geometry. The X-ray pulses scattered by the sample were collected with a two-dimensional MarCCD. To reduce the scattering from the air, a cone filled with helium gas was used between the sample and the MarCCD. At the repetition rate of 10 Hz, each image was exposed to 1000 X-ray shots, amounting to the exposure time of 2 miniutes per image. To avoid radiation damage and provide fresh sample for each pair of X-ray and laser pulses, the capillary containing the sample was translated back and forth along its long axis over a 20-mm

range. To ensure that successive pulses in the 10-Hz pulse train excite a fresh portion of the protein solution, the sample was translated by 0.2 mm after each probe pulse. The laser-off images were acquired with laser pulses arriving 5 μs earlier than the X-ray pulse in order to probe the ground state while assuring the same average temperature of the solution. These laser-off images were used to compute the time-resolved X-ray scattering differences. Usually a laser-off image was collected after every 3 or 4 laser-on images to compensate for slow drifts in the beamline. To attain enough signal-to-noise ratio, ~60 images were acquired at each time delay. The measured time delays are basically spread evenly in logarithmic time scale and are as follows: –5 μs, 0 ps, 100 ps, 300 ps, 500 ps, 1 ns, 3 ns, 10 ns, 50 ns, 70 ns, 100 ns, 300 ns, 1 us, 3.16 ms, 10 ms, 31.6 ms.

Data processing

Two-dimensional scattering images were azimuthally averaged to give one-dimensional scattering curves. To convert the scattering angle to q, the center of mass position of the undulator spectrum (1.033 Å) was used as the reference wavelength. The laser-induced scattering intensity changes are less than a few percent of the static scattering intensity and thus a careful normalization is necessary to extract accurate laser-induced scatteing differences. As a normalization reference, we used the q positon of 2.07 $Å^{-1}$, which is the isosbestic point of the water scattering curves with respect to temperature increase, so that the difference scattering intensity at this q value is zero. Figure S1 show time-resolved data at all time-delays. In Figure S2A, the data at 3.16 ms, 10 ms and 31.6 ms are superimposed. At these time delays, the signal at low q values is negligible, indicating that the contribution from transiently generated species disappears. Moreover, the difference curve at 10 ms can be well represented by a sum of bulk water scattering intensity changes due to the temperature change at constant density ($(\partial S(q)/\partial T)_\rho$) and the density change at constant temperature ($(\partial S(q)/\partial \rho)_T$) as shown in Figure S2B, confirming that the differences at late time delays are due to solvent heating. To remove the solvent heating effect, this solvent heating signal was subtracted from the difference curves (see Figure S3).

Data analysis

(1) Singular value decomposition

In order to determine the kinetic model, we need to examine how many distinct transient species are involved in the dynamic process of interest and how fast the population of each species changes after photolysis. For this purpose, we applied the singular value decomposition (SVD) analysis and principal component analysis (PCA), which will be detailed in the next section, to our experimental data in the *q*-range of 0.08 – 2.5 Å $^{-1}$. From the experimental scattering curves measured at various time delays, we can build a $n_q \times n_t$ matrix *A*, where n_q is the number of *q* points at a given time-delay point and n_t is the number of time-delay points. For the data presented in this paper, n_q and n_t are 231 and 11, respectively. Then, the matrix *A* can be decomposed while satisfying the relationship of $A = USV^T$, where *U* is an $n_q \times n_t$ matrix whose columns are called left singular vectors (i.e. time-independent *q* spectra) of *A*, *V* is an $n_t \times n_t$ matrix whose columns are called right singular vectors (i.e. amplitude changes of *U* as time evolves) of *A*, and *S* is a diagonal $n_t \times n_t$ matrix whose diagonal elements are called singular values of *A* and can possess only non-negative values. The matrices *U* and *V* have the properties of $U^T U = I_{n_t}$ and $V^T V = I_{n_t}$, respectively, where I_{n_t} is the identity matrix. Since the diagonal elements (i.e. singular values) of *S*, which represent the weight of left singular vectors in *U*, are ordered so that $s_1 \geq s_2 \geq ... \geq s_n \geq 0$, (both left and right) singular vectors on more left are supposed to have larger contributions to the constructed experimental data. In this manner, we can extract the time-independent scattering intensity components (from left singular vectors), which, when combined together, contains the information on the scattering curves of distinct transient species, and the time evolution of their amplitudes (from right singular vectors), that is, the information on the population dynamics of the transient species. Thus, the SVD analysis provides a model-independent estimation of the number of structurally distinguishable species and the dynamics of each species.

As shown in Figure S4, the singular values and autocorrelation values of the corresponding singular vectors suggest that four singular vectors are enough to represent our experimental data. In other words, the contribution becomes negligible from the fifth singular vector and beyond. By simultaneously fitting the right singular vectors of the four singular vector components with multiple exponentials, the best fitting result was obtained when we use four exponentials (with

time constants of 360 ps, 9 ns, 70 ns and 150 ns) in the time range of 100 ps – 1 μs. The presence of four time-independent components indicates that there exist at least four structurally distinct intermediate species.

(2) Principal component analysis

Using the four time constants obtained by the SVD analysis, we then applied principal component analysis (PCA) using a simple sequential kinetic model. As a result, our experimental data were decomposed to species-associated difference scattering curves corresponding to five intermediate components (C_1 to C_5):

$$\Delta S_{theory}(q_i, t_j) = \sum_{k=1}^{5} [C_k(t_j)] \Delta S_{C_k}(q_i) + \Delta S_{water}(q_i) \quad (1)$$

where $\Delta S_{theory}(q_i, t_j)$ is the theoretical difference scattering curve at a given q and t value, $\Delta S_{C_k}(q_i)$ is the species-associated difference scattering curve corresponding to k th intermediate component at a given q value, $[C_k(t_j)]$ is the instantaneous population of k th intermediate component that obeys the simple sequential kinetic model at a given t value. The difference scattering curve of bulk water ($\Delta S_{water}(q_i)$) was also included to account for the contribution of solvent heating to the total difference scattering curves. Then, we minimized the discrepancy between theory and experiment using Minuit package:

$$\chi_i^2 = \sum_{j=1}^{n_t} \left(\frac{\Delta S_{\exp}(q_i, t_j) - \Delta S_{theory}(q_i, t_j)}{\sigma_{ij}} \right)^2 \quad (2)$$

where χ_i^2 is the discrepancy between theory and experiment at a given q value, $\Delta S_{\exp}(q_i, t_j)$ is the experimental difference scattering curve at a given q and t value and σ_{ij} is the experimental standard deviation at a given q and t value.

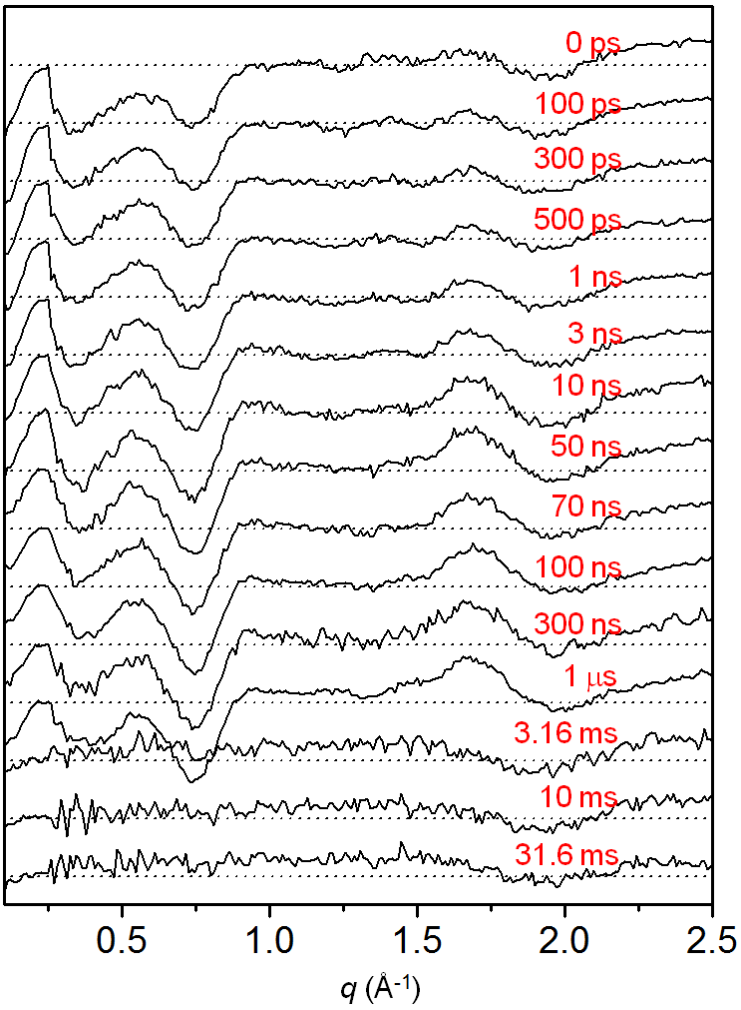

Figure S1. Time-resolved X-ray scattering data at all time-delays.

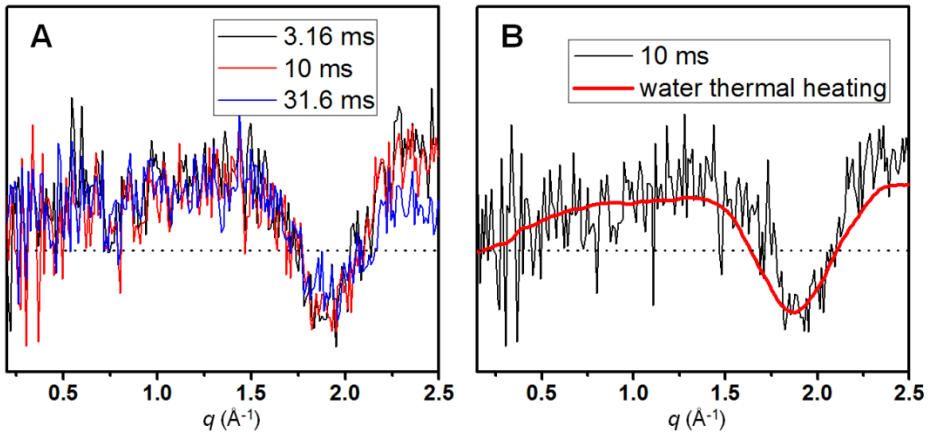

Figure S2. (A) The scattering data at 3.16 ms, 10 ms and 31.6 ms are superimposed. At these time delays, the low q signal is absent, indicating that the contribution from transiently generated species disappear. (B) Difference curve at 10 ms can be well represented by a sum of bulk water scattering intensity changes due to the temperature change at constant density ($(\partial S(q)/\partial T)_\rho$) and the density change at constant temperature ($(\partial S(q)/\partial \rho)_T$), confirming that the differences at late time delays are due to solvent heating.

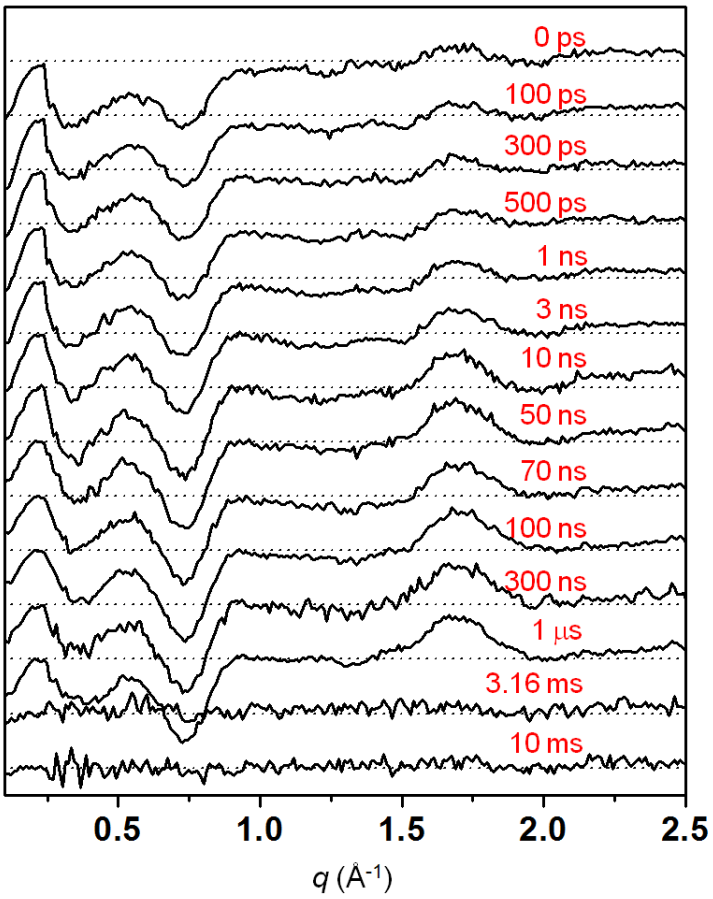

Figure S3. Heat-free, time-resolved X-ray scattering data. Thermal heating signal was subtracted from the difference curves presented in Figure S1.

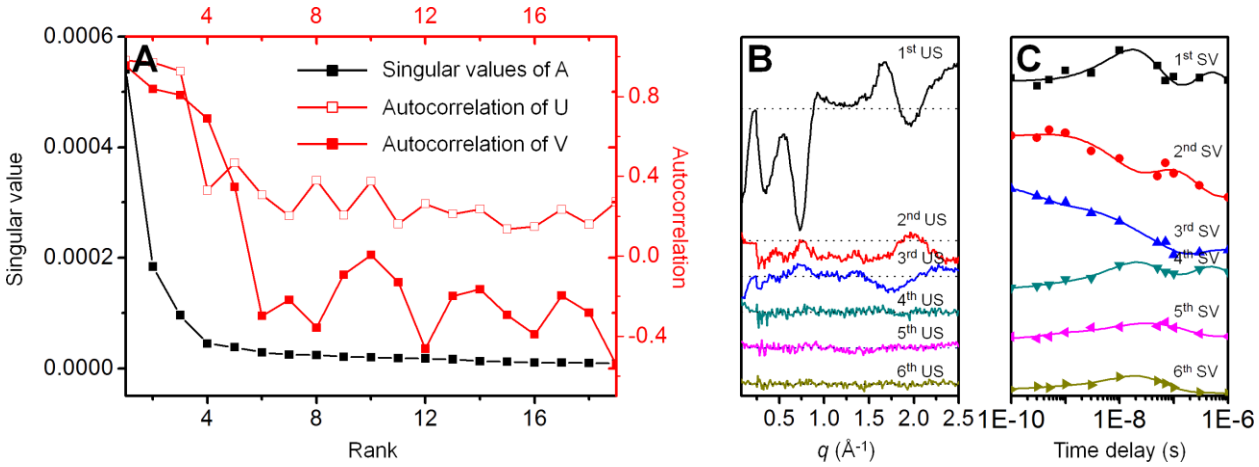

Figure S4. (A) Singular values of *A* (black solid square) and autocorrelations of left singular vectors (*U*, red open square) and right singular vectors (*V*, red solid square) are shown. Judging from the amplitude of the singular values and the autocorrelations, we employed the first four singular vectors for our kinetic analysis. (B) The first six left singular vectors scaled by singular values of *A* are shown. (C) The first six right singular vectors scaled by singular values of *A* (points) and their fits (lines) obtained by using four exponentials (with time constants of 360 ps, 9 ns, 70 ns and 150 ns that were obtained by fitting only the first four right singular vectors) in the time range of 100 ps – 1 μs. This figure indicates that the four time constants are enough to describe the kinetics within the signal-to-noise ratio of our data.