

## Supporting Information

# Conformational Substates of Myoglobin Intermediate Resolved by Picosecond X-ray Solution Scattering

Key Young Oang,<sup>†, ‡, #</sup> Jong Goo Kim,<sup>†, ‡, #</sup> Cheolhee Yang,<sup>†, ‡</sup> Tae Wu Kim,<sup>†, ‡</sup> Youngmin Kim,<sup>†, ‡</sup> Kyung Hwan Kim,<sup>†, ‡</sup> Jeongho Kim,<sup>§</sup> and Hyotcherl Ihee<sup>\*, †, ‡</sup>

<sup>†</sup> Center for Nanomaterials and Chemical Reactions, Institute for Basic Science (IBS), Daejeon 305-701, Korea

<sup>‡</sup> Department of Chemistry, Graduate School of Nanoscience & Technology (WCU), Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea

<sup>§</sup> Department of Chemistry, Inha University, Incheon 402-751, Korea

\* Corresponding author: [hyotcherl.ihee@kaist.ac.kr](mailto:hyotcherl.ihee@kaist.ac.kr)

# These authors contributed equally to this work.

## 1. Sample preparation

Solution samples of carbonmonoxy myoglobin (MbCO) were prepared as follows. Sperm whale Mb (swMb) gene was synthesized following a well-established protocol<sup>1</sup> with some modifications. First, amino acid sequences of swMb were reversely translated by using optimal genetic codon in *E. coli* and twenty-two oligonucleotides containing 31 – 46 nucleotides in length were synthesized. Subsequently, the restriction enzyme sites with NdeI and BamHI were inserted at the first and the last nucleotides of each oligonucleotide. For each oligonucleotide, four hundred picomoles of oligonucleotides were phosphorylated by using T4 polynucleotide kinase and then ligated with Taq ligase. A full-length swMb gene was constructed from the ligation product by polymerase cycling assembly (PCA) using Pfu polymerase. The full-length gene produced by PCA was amplified by polymerase chain reaction (PCR). The PCR mixtures contained Pfu polymerase, buffer, dNTPs, the PCA product, and PF1 or PR11 oligonucleotides. The PCR-amplified DNA and pET1b vector were cleaved with NdeI and BamHI, and the product was inserted into the vector by ligation. The sequence mutation was corrected by using site-directed mutagenesis. The pET15-swMb plasmid was inserted to BL21 (DE3). The *E. coli* was grown at 37 °C in LB broth in the presence of ampicillin. When the optical density reached 0.4, IPTG was added to the culture broth. The harvested cells in late stationary phase were stored at –70 °C. When used, the cells were thawed and sonicated at 4 °C in 20-mM NaPO<sub>4</sub> buffer (pH 7.0) containing 150-mM NaCl and 1-mM PMSF. The cell debris was removed by centrifugation at 17,000 rpm. The supernatant was loaded into nickel affinity column equilibrated in 20-mM NaPO<sub>4</sub> buffer (pH 7.0) containing 150-mM NaCl. The eluted swMb was bubbled by CO gas and dialyzed using thrombin cleavage buffer (20-mM Tris-HCl at pH 8.4, 150-mM NaCl, and 10-mM CaCl<sub>2</sub>) and treated with thrombin at 4 °C for 3 days in order to remove His-tag. The resulting swMb contains three extra residues (Gly-Ser-Met) in the N-terminus compared with the naturally occurring wild-type swMb. The sample was then applied to the second nickel affinity column to remove uncut His-tag swMb. The sample solution was further purified by gel permeation chromatography using a gel filtration column equilibrated with 20-mM Tris-HCl buffer (pH 8.0). Subsequently, the buffer was changed to 100-mM NaPO<sub>4</sub> (pH 7.0) by dialysis. The Mb sample solution was concentrated to 8 mM, reduced by adding sodium dithionite (10 mM), and bubbled by CO gas for 30 minutes so that MbCO can be formed. The sample solution was prepared just before the X-ray solution scattering measurement. An aliquot of the

resulting MbCO solution was transferred into a 1-mm diameter X-ray capillary (Hampton Research) and immediately sealed with epoxy to minimize gas exchange while CO gas was purged continuously into the capillary.

## 2. Data collection

Time-resolved X-ray solution scattering data were collected at the BioCARS 14-ID-B beamline at the Advanced Photon Source while the storage ring was operated in the standard top-up mode.<sup>2</sup> The electron bunch fill pattern in this mode has 102 mA in 24 single bunches with a nominal current of 4.25 mA per bunch and a spacing of 153.4 ns 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 has the bandwidth of  $\sim 3\%$ . To reduce the data acquisition time, the full spectrum was used without being monochromatized. The blurring effect of the polychromaticity with  $\sim 3\%$  bandwidth to the scattering curve is not substantial. The 100-ps-long (FWHM) X-ray pulse containing  $\sim 10^{10}$  photons was focused by Kirkpatrick-Baez focusing mirrors, trimmed by slits and delivered to the sample position with a spot size of  $0.14 \times 0.07 \text{ mm}^2$  (horizontal  $\times$  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. The high-speed chopper has a triangular-shaped rotor with a tunnel spinning at  $\sim 1 \text{ kHz}$ , a subharmonic of the synchrotron, and isolates single X-ray pulses. The repetition rate was further reduced from 1 kHz down to typically 40 Hz using a millisecond shutter to allow long time delays and enough time to move the capillary to a fresh position for every pair of laser and X-ray pulses.

The protein sample was excited by  $\sim 35$ -ps-long laser pulses at 532 nm generated from a picosecond laser system at the 14-ID-B beamline. A picosecond Ti: Sapphire amplifier (Spectra Physics Spitfire Pro) was seeded by femtosecond pulse train at 780 nm generated from a Ti: Sapphire oscillator (Tsunami). The output from the amplifier was converted to 532 nm by a TOPAS optical parametric amplifier and stretched to  $\sim 35$  ps by an echelon pair. To completely eliminate the polarization-dependent anisotropy in the one-dimensional curve obtained by azimuthal integration,<sup>3</sup> the circularly polarized laser pulse of  $\sim 37 \mu\text{J}$  energy was

focused to a spot of 0.6 mm by 0.13 mm size at the sample position, yielding an energy density of  $0.6 \text{ mJ/mm}^2$ .<sup>4,5</sup> The sample was contained in a sealed capillary of 1-mm diameter and excited by a laser pulse from the top. The sample was maintained at 22 °C with a cold nitrogen stream (Oxford Cryostream). The X-ray pulse transmitted 0.2 mm deep from the top of the capillary in perpendicular geometry. The X-ray pulses scattered by the sample were collected with a two-dimensional Mar165 CCD detector. To reduce the scattering from the air, a cone filled with helium gas was used between the sample and the CCD detector. At the repetition rate of 40 Hz, the exposure time for each image was 20 seconds, corresponding to 800 laser/X-ray pulse pairs 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 with the movement synchronized with the pulse trains of laser and X-rays. To ensure that successive pulses in the 40-Hz pulse train excite a fresh portion of the protein solution, the sample was translated by 0.2 mm after each pump/probe pair.

The laser-off images were acquired with the laser pulse arriving 5  $\mu\text{s}$  later than the X-ray pulse in order to probe the (unexcited) ground state while assuring the same average temperature of the solution. These laser-off images were used as a reference for calculating the time-resolved difference X-ray scattering patterns. Usually, a laser-off image was collected after every 3 or 4 laser-on images to compensate for slow drifts of the X-ray intensity in the beamline. To attain enough signal-to-noise ratio, about 80 images were acquired at each time delay. The measured time delays are basically spread evenly in logarithmic time scale as follows:  $-5 \mu\text{s}$ , 100 ps, 178 ps, 316 ps, 562 ps, 1 ns, 1.78 ns, 3.16 ns, 5.62 ns, 10 ns, 17.8 ns, 31.6 ns, 56.2 ns, 100 ns, 178 ns, 316 ns, 562 ns, 1  $\mu\text{s}$ , 1.78  $\mu\text{s}$ , 3.16  $\mu\text{s}$ , 5.62  $\mu\text{s}$ , 10  $\mu\text{s}$ , 17.8  $\mu\text{s}$ , 31.6  $\mu\text{s}$ , 56.2  $\mu\text{s}$ , 100  $\mu\text{s}$ , 178  $\mu\text{s}$ , 316  $\mu\text{s}$ , 562  $\mu\text{s}$ , 1 ms, 1.78 ms, 3.16 ms, 5.62 ms, 10 ms.

### **3. Data processing**

Two-dimensional scattering images were azimuthally averaged to give one-dimensional scattering curves. To convert the scattering angle ( $\theta$ ) to  $q$ , the center-of-mass position of the undulator spectrum 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 scattering differences. As a normalization reference, we used the  $q$  position of  $2.07 \text{ \AA}^{-1}$ , 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 shows time-resolved X-ray solution scattering data of MbCO containing solvent heating at all the time delays. In Figure S2, the data at 10 ms is presented as an example. At this time delay, the signal at low  $q$  values is negligible, indicating the absence of the contribution from transiently generated species. Moreover, the difference curve at 10 ms can be well represented by a sum of scattering intensity changes from bulk water arising from 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 S2, confirming that the difference scattering at late time delays are mainly attributed to solvent heating. To remove the solvent heating effect, this signal from solvent heating was subtracted from the difference curves at all the time delays (see Figure S3).

## 4. Data analysis

### 4.1. 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 kinetic analysis, which will be detailed in the next section, to our experimental data in the  $q$  range of  $0.15 - 1.0 \text{ \AA}^{-1}$ . From the experimental scattering curves measured at various time delays, we can build an  $n_q \times n_t$  matrix  $\mathbf{A}$ , where  $n_q$  is the number of  $q$  points in the scattering curve 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 414 and 33, respectively. Then, the matrix  $\mathbf{A}$  can be decomposed while satisfying the relationship of  $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ , where  $\mathbf{U}$  is an  $n_q \times n_t$  matrix whose columns are called left singular vectors (LSVs) (i.e. time-independent  $q$  spectra) of  $\mathbf{A}$ ,  $\mathbf{V}$  is an  $n_t \times n_t$  matrix whose columns are called right singular vectors (rSVs) (i.e. amplitude changes of  $\mathbf{U}$  as time evolves) of  $\mathbf{A}$ , and  $\mathbf{S}$  is an  $n_t \times n_t$  diagonal matrix whose diagonal elements are called singular values of  $\mathbf{A}$  and can possess only non-negative values. The matrices  $\mathbf{U}$  and  $\mathbf{V}$  have the properties of  $\mathbf{U}^T\mathbf{U} = \mathbf{I}_{n_t}$  and  $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{n_t}$ , respectively, where  $\mathbf{I}_{n_t}$  is the identity matrix. Since the diagonal elements (i.e. singular values)

of  $\mathbf{S}$ , which represent the weight of left singular vectors in  $\mathbf{U}$ , are ordered so that  $s_1 \geq s_2 \geq \dots \geq s_n \geq 0$ , (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 the left singular vectors and the time evolution of their amplitudes from the right singular vectors. The former, when combined together, can give the information on the scattering curves of distinct transient species, while the latter contains 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 autocorrelations of the corresponding singular vectors suggest that four singular vectors are enough to represent our experimental data of Mb because the contribution of each singular vector (lSV or rSV) to the data is proportional to its corresponding singular value and the autocorrelation of  $\mathbf{U}$  or  $\mathbf{V}$  matrix can serve as a good measure of signal-to-noise ratio of the singular vectors.<sup>6</sup> In other words, the contribution from the fifth singular vector and beyond becomes negligible. This judgment is consistent with previous studies. In this regard, the first four right singular vectors multiplied by singular values were fit by six exponentials sharing common relaxation times, yielding the relaxation times of  $460 \pm 160$  ps,  $3.6 \pm 0.7$  ns,  $92 \pm 25$  ns,  $1.4 \pm 0.2$   $\mu$ s,  $90 \pm 20$   $\mu$ s, and  $1.2 \pm 0.2$  ms. To verify the validity of using six exponentials, we also tried fitting the data using exponentials fewer or more than six as shown in Figure S5. The fit clearly deviates from the data when we use five exponentials, and the data are overfitted when we use seven exponentials. Therefore, it is valid to use six exponentials for fitting the data.

## 4.2. Kinetic analysis

Using the first four singular vectors of significant singular values obtained from the SVD analysis of the experimental data, we performed kinetic analysis. New matrices,  $\mathbf{U}'$ ,  $\mathbf{V}'$ , and  $\mathbf{S}'$ , can be defined by removing non-significant components from  $\mathbf{U}$ ,  $\mathbf{V}$ , and  $\mathbf{S}$ , respectively. In other words,  $\mathbf{U}'$  is an  $n_q \times 4$  matrix containing only the first four left singular vectors of  $\mathbf{U}$ ,  $\mathbf{S}'$  is a  $4 \times 4$  diagonal matrix containing the first four singular values of  $\mathbf{S}$ , and  $\mathbf{V}'$  is an  $n_t \times 4$  matrix containing the first four right singular vectors of  $\mathbf{V}$ . Here we define a matrix  $\mathbf{C}$ , of which the columns represent time-dependent concentrations of transiently formed intermediate species and can be described by a candidate kinetic model (for example, one of

the kinetic models shown in Figure S6). Then, the matrix  $C$  can be related to  $V'$  by using a parameter matrix  $P$  that satisfies  $V' = CP$ . In our analysis,  $C$  is an  $n_t \times 4$  matrix containing the time-dependent concentrations of four intermediates of Mb (**B**, **C**, **D**, and **S**), where **B** is presumably Mb with CO still in the primary docking site, **C** and **D** are presumably the intermediate states where the CO is located in Xe4 site and Xe1 site, respectively, and **S** means the Mb with free CO (in other words, CO has completely escaped to the solvent.), and  $P$  is a  $4 \times 4$  matrix containing coefficients for the time-dependent concentrations so that the linear combination of concentrations of the four intermediates can form the four right singular vectors in  $V'$ . Once  $C$  is expressed using a set of variable kinetic parameters based on a candidate kinetic model,  $P$  and  $C$  can be optimized by minimizing the discrepancy between  $V'$  (from the experiment) and  $CP$  (from the kinetic theory). We performed this optimization for each of the candidate models shown in Figure S6 and compared the minimized discrepancies of all the kinetic models (see  $\chi_{red}^2$  values listed in Table S1) to determine the optimum kinetic model that best fits the experimental data.

However, standard deviations for  $V'$  are not available from the experimental data and thus we instead used the following method to optimize  $P$  and  $C$ . Since  $V' = CP$ , the following relationships hold:

$$A' = U'S'V'^T = U'S'(CP)^T = U'S'P^TC^T = (U'S'P^T)C^T \quad (1)$$

where  $A'$  is an  $n_q \times n_t$  matrix that contains the theoretical difference scattering curve  $\Delta S_{theory}(q_i, t_j)$  at given  $q$  and  $t$  values. Theoretical difference scattering curves calculated by using Eqn. (1) were compared with the experimental difference scattering curves, and the matrix  $P$  and  $C$  were optimized by minimizing the discrepancy (quantified by “reduced” chi-square,  $\chi_{red}^2$ ) between the theoretical and experimental difference scattering curves using the Minuit package:<sup>7</sup>

$$\chi_{red}^2 = \left( \frac{1}{N - n - 1} \right) \sum_{i=1}^{n_q} \sum_{j=1}^{n_t} \left( \frac{\Delta S_{exp}(q_i, t_j) - \Delta S_{theory}(q_i, t_j)}{\sigma_{ij}} \right)^2 \quad (2)$$

where  $N$  is the number of data points and  $n$  is the number of fitting parameters.  $\Delta S_{exp}(q_i, t_j)$  and  $\Delta S_{theory}(q_i, t_j)$  are the experimental and theoretical difference scattering intensities at a given point of  $(q_i, t_j)$ , respectively, and  $\sigma_{ij}$  is the experimental standard deviation at  $(q_i, t_j)$ .

From Eqn. (1), we can define a matrix  $E$  as  $E = U'S'P^T$ , that is, a linear combination of the four left singular vectors in  $U'$  weighted by their singular values in  $S'$  with their ratios determined by  $P$ . Then, the matrix  $E$ , an  $n_t \times 4$  matrix, contains the four difference scattering curves directly associated with the four intermediate species of Mb. Therefore, by optimizing the matrices  $P$  and  $C$ , we obtain both the time-dependent concentrations (see the optimized  $C$  for the optimum kinetic model in Figure 3b in the main text.) and the (time-independent) scattering curves of the intermediate species (see the optimized  $E$  for the best-fit kinetic model in Figure S7.).

## 5. Comparison with previous studies using time-resolved X-ray solution scattering

We can compare the present work with previous studies on MbCO performed using time-resolved X-ray solution scattering (refs. 5 – 7 in the main text) as shown in Table S2. Ref. 5 (Ahn *et al.*) characterized the solution structure of a transient intermediate species (formed on nanosecond time scale after CO photolysis) of wild-type sperm whale MbCO by applying the rigid-body molecular dynamics (MD) simulation to nanosecond X-ray solution scattering data measured at a single time point (10 ns). Therefore, ref. 5 does not contain any information on the kinetics of structural transition of MbCO. Ref. 6 (Kim *et al.*) investigated structural dynamics of wild-type “horse heart” (instead of sperm whale) MbCO in the time range from 100 ps to 1  $\mu$ s (with two time points per decade) and only a simple sequential model was considered in the kinetic analysis. In contrast, our present work investigates the structural dynamics of wild-type sperm whale MbCO in a wider time range (100 ps – 10 ms) using the data measured at twice more time points (four time points per decade) than in ref. 6. Furthermore, in our present work, we applied the kinetic analysis considering all the plausible candidate models based on the SVD analyses in two different time ranges so that we can determine the optimum kinetic model with much higher fidelity. Ref. 7 (Cho *et al.*) studied the structural dynamics of wild-type sperm whale MbCO in the time range from 100 ps to 10 ms with four time points per decade, and therefore is equivalent to our present work in terms of experimental technique, sample, time range, and number of time points. However, the key difference between ref. 7 and our present work is that ref. 7 used only a sequential kinetic model that involves geminate CO recombination on microsecond time scale but no biphasic transition. In fact, the kinetic model employed in ref. 7 is identical to one of the candidate



models (model **r**) examined in this work (see Figure S6). However, all the evidences presented in this work demonstrate that model **r** is not an optimum model. First of all, model **r** gives the second largest  $\chi^2$  value among a total of eighteen candidate models. Also, model **r** has three intermediates in the time range up to 3.16 ns and therefore is not consistent with the result of our SVD analysis that suggests the existence of two intermediates in that time range.

**Table S1.** Reduced chi-square values ( $\chi_{red}^2$ ) obtained from the kinetic analysis using various models as shown in Figure S6.

|                |          |          |          |          |          |          |
|----------------|----------|----------|----------|----------|----------|----------|
| <b>Model</b>   | <b>a</b> | <b>b</b> | <b>c</b> | <b>d</b> | <b>e</b> | <b>f</b> |
| $\chi_{red}^2$ | 1.72     | 1.66     | 1.54     | 1.72     | 1.58     | 1.68     |
| <b>Model</b>   | <b>g</b> | <b>h</b> | <b>i</b> | <b>j</b> | <b>k</b> | <b>l</b> |
| $\chi_{red}^2$ | 1.64     | 1.60     | 1.68     | 1.64     | 1.92     | 1.68     |
| <b>Model</b>   | <b>m</b> | <b>n</b> | <b>o</b> | <b>p</b> | <b>q</b> | <b>r</b> |
| $\chi_{red}^2$ | 1.62     | 1.84     | 1.72     | 1.66     | 1.64     | 1.84     |

**Table S2.** Comparison with previous studies using time-resolved X-ray solution scattering

|                                   | <b>Ahn <i>et al.</i></b> <sup>a)</sup> | <b>Kim <i>et al.</i></b> <sup>b)</sup> | <b>Cho <i>et al.</i></b> <sup>c)</sup> | <b>This work</b>               |
|-----------------------------------|--|--|--|--------------------------------|
| <b>Sample</b>                     | wild-type swMbCO <sup>d)</sup>         | wild-type hhMbCO <sup>e)</sup>         | wild-type swMbCO <sup>d)</sup>         | wild-type swMbCO <sup>d)</sup> |
| <b>Time resolution</b>            | 10 ns                                  | 100 ps                                 | 100 ps                                 | 100 ps                         |
| <b>Time range</b>                 | 10 ns only                             | 100 ps – 1 $\mu$ s                     | 100 ps – 10 ms                         | 100 ps – 10 ms                 |
| <b>Number of time points</b>      | 10 ns only                             | 2 time points per decade               | 4 time points per decade               | 4 time points per decade       |
| <b>Number of candidate models</b> | --                                     | 1                                      | 1                                      | 18                             |
| <b>Optimum kinetic model</b>      | --                                     | sequential model                       | model <b>r</b> in Figure S6            | model <b>c</b> in Figure S6    |

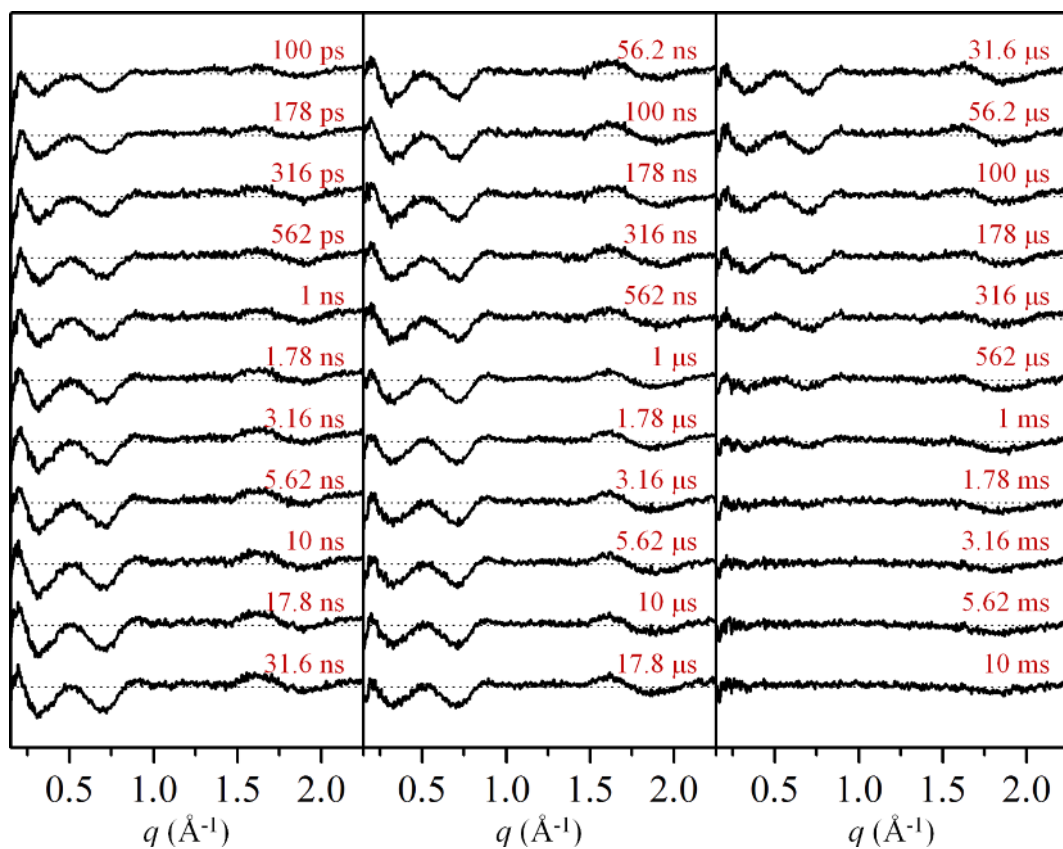
<sup>a)</sup> Ref. 5 of the main text

<sup>b)</sup> Ref. 6 of the main text

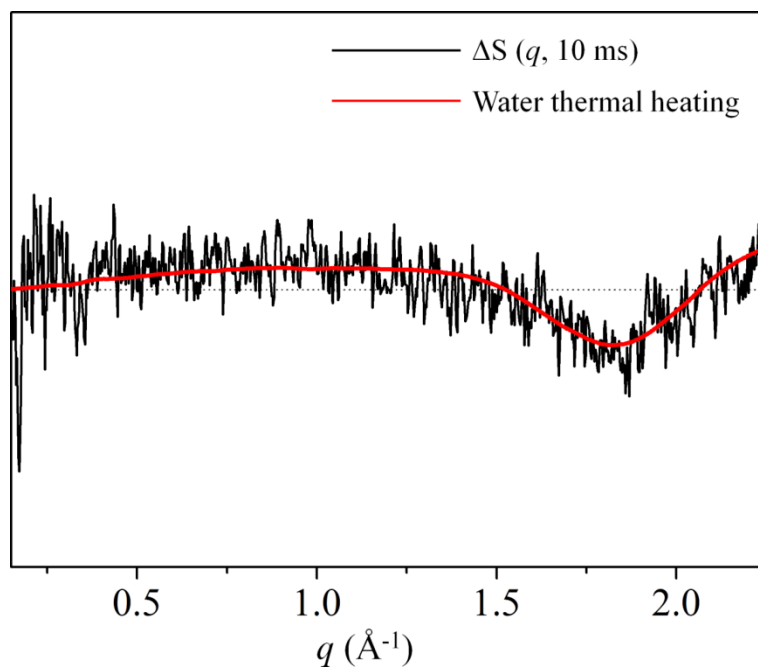
<sup>c)</sup> Ref. 7 of the main text

<sup>d)</sup> swMbCO stands for sperm whale MbCO.

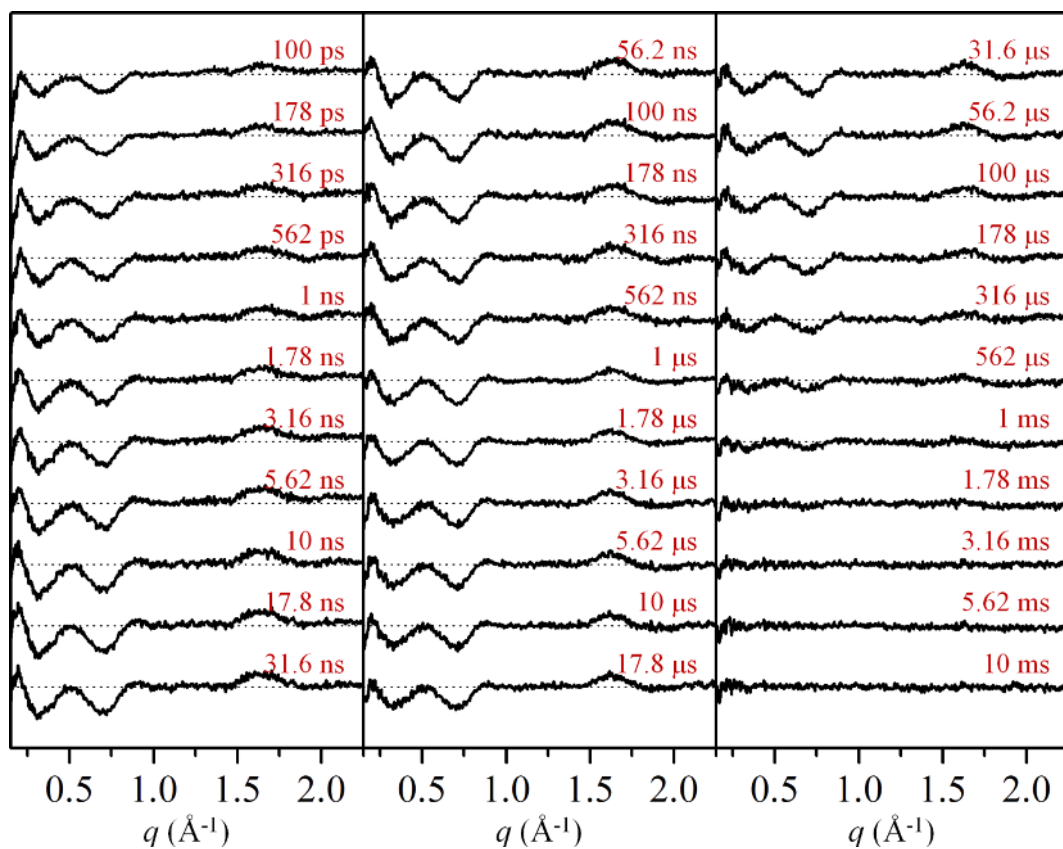
<sup>e)</sup> hhMbCO stands for horse heart MbCO.



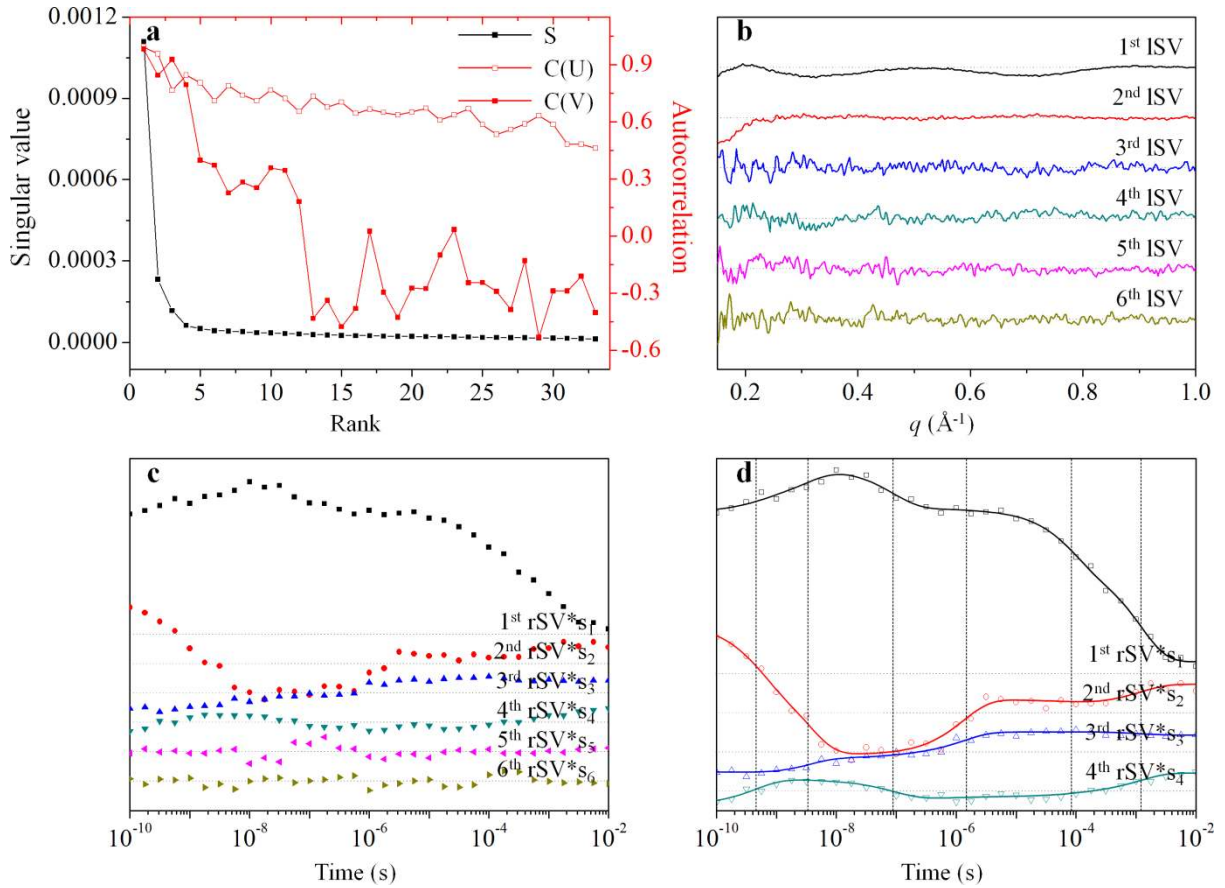
**Figure S1.** Time-resolved X-ray solution scattering data of MbCO containing solvent heating at all the time delays.



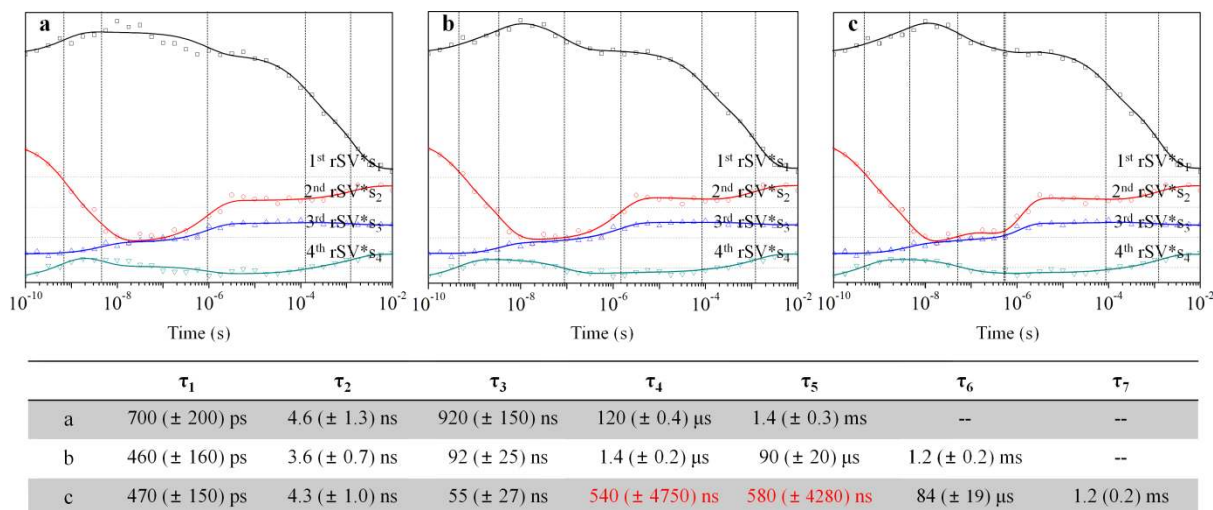
**Figure S2.** The scattering data at 10 ms time delay (black). At this time delay, the signal at low  $q$  values ( $< 1.0 \text{ \AA}^{-1}$ ) is very small, indicating that the contribution from transiently generated species is negligible. Difference curve at 10 ms can be represented by a sum of contributions from thermal heating of bulk water (red), that is, the temperature change at constant density ( $(\partial S(q)/\partial T)_\rho$ ) and the density change at constant temperature ( $(\partial S(q)/\partial \rho)_T$ ). This observation confirms that the difference signals at late time delays mainly arise from solvent heating.



**Figure S3.** Heat-free, time-resolved X-ray solution scattering data of MbCO at all the time delays.

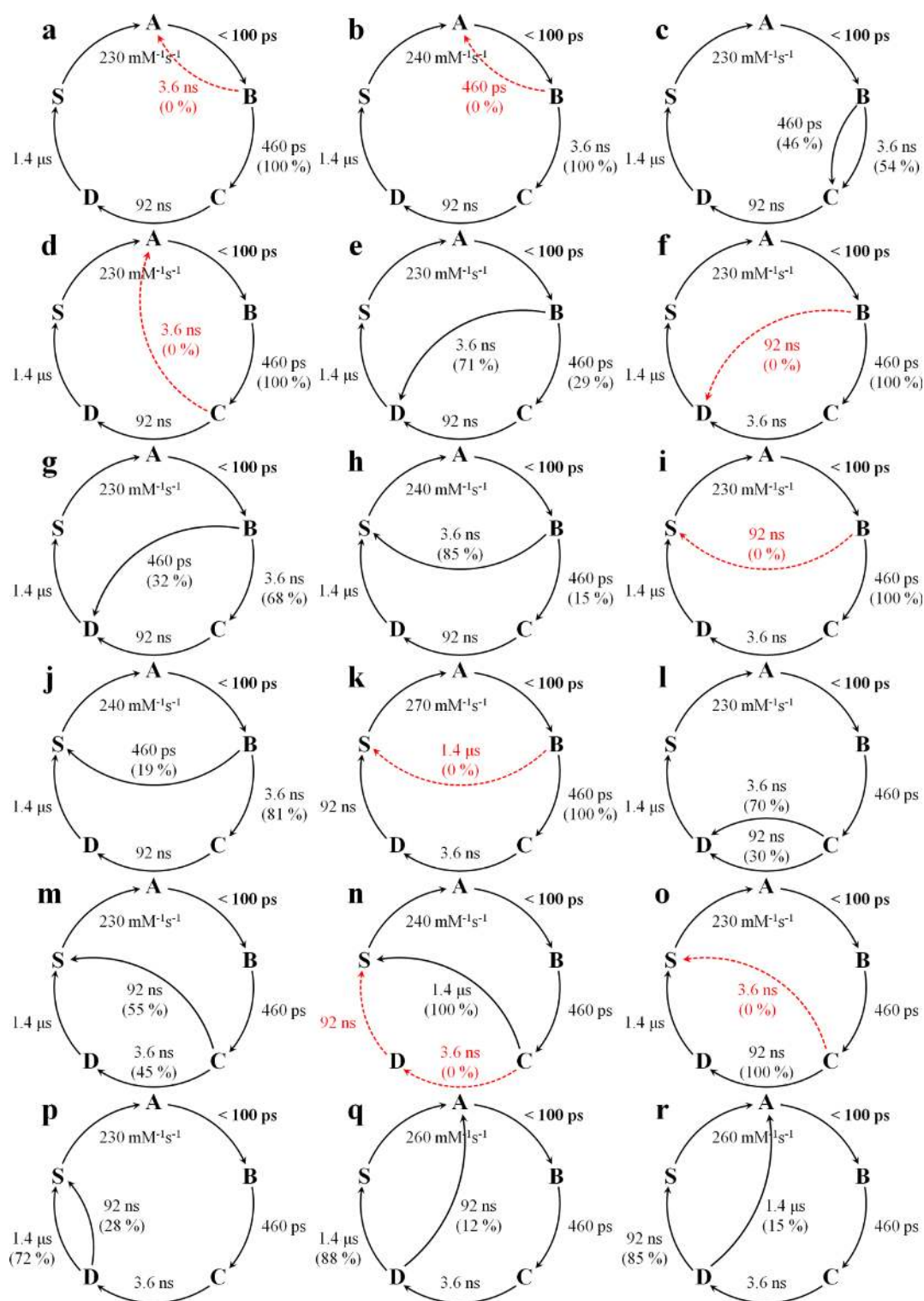


**Figure S4.** (a) Singular values (S, black solid squares), autocorrelations of left singular vectors (C(U), red open squares) and autocorrelations of right singular vectors (C(V), red solid squares) obtained from time-resolved X-ray solution scattering data of MbCO. The contribution of each singular vector (ISV or rSV) to the data is proportional to its corresponding singular value. Meanwhile, the autocorrelation of U or V matrix can serve as a good measure of signal-to-noise ratio of the singular vectors. Judging by the amplitudes of the singular values and the autocorrelations, we employed the first four ISVs and rSVs for our kinetic analysis. (b) The first six left singular vectors. (c) The first six right singular vectors multiplied by singular values. (d) The first four right singular vectors multiplied by singular values were fit by six exponentials sharing common relaxation times, yielding the relaxation times of  $460 \pm 160$  ps,  $3.6 \pm 0.7$  ns,  $92 \pm 25$  ns,  $1.4 \pm 0.2$   $\mu$ s,  $90 \pm 20$   $\mu$ s, and  $1.2 \pm 0.2$  ms indicated by the vertical dotted lines.



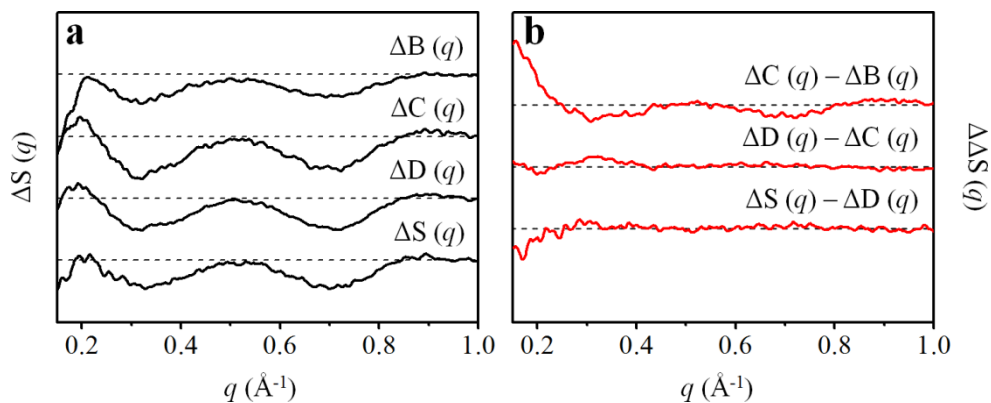
**Figure S5.** In order to verify the validity of using six exponentials for the fit, the first four right singular vectors (rSVs) multiplied by singular values were fit by (a) five exponentials, (b) six exponentials, (c) and seven exponentials sharing common relaxation times. The vertical dotted lines indicate the time constants of the used exponentials. The complete fit parameters are listed in the table at the bottom. When we use five exponentials, as shown in (a), the fit quality of each rSV component is clearly worse than when using six exponentials, even for the most dominant first rSV component. When we use seven exponentials, the fit quality does not improve much for any rSV component compared with when using six exponentials. Although we can see some improvement of the fit quality for the second rSV (red points), two (of the seven) exponentials have similar time constants ( $\tau_4$  and  $\tau_5$ ) to each other and their errors are extremely large. Based on this observation, we consider that the data are overfitted by seven exponentials. Thus, we conclude that it is valid to use six exponentials for fitting the data.





**Figure S6.** All possible candidate models considered in this work. The models **a – d** are consistent with the SVD analysis in the entire time range (100 ps – 10 ms) as well as in the reduced time range (100 ps – 3.16 ns). For example, model **c** has four structurally distinct intermediates in the entire time range and two intermediates in the reduced time range, which is in agreement with the results of the SVD analyses in the two time ranges. In contrast, the models **e – r** are consistent only with the SVD analysis in the entire time range. For example,

model **e** has four intermediates in the entire time range and three intermediates in the reduced time range, contradicting the result of the SVD analysis in the reduced time range. Red dashed lines indicate the parallel pathways that make zero contribution to the photocycle.



**Figure S7.** (a) Species-associated difference scattering curves for the four intermediates. These curves correspond to the constituent of the matrix **E**. (b) Double difference scattering curves that highlight subtle changes at small angles ( $q = 0.15 - 1.0 \text{\AA}^{-1}$ ). The double difference scattering curves were calculated by taking the difference between the temporally adjacent species-associated difference scattering curves shown in (a).

## References

- (1) Smith, H. O.; Hutchison, C. A.; Pfannkoch, C.; Venter, J. C. Generating a Synthetic Genome by Whole Genome Assembly:  $\phi$ X174 Bacteriophage from Synthetic Oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15440-15445.
- (2) 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*, 658-670.
- (3) Kim, J.; Kim, K. H.; Kim, J. G.; Kim, T. W.; Kim, Y.; Ihee, H. Anisotropic Picosecond X-Ray Solution Scattering from Photoselectively Aligned Protein Molecules. *J. Phys. Chem. Lett.* **2011**, *2*, 350-356.
- (4) Ahn, S.; Kim, K. H.; Kim, Y.; Kim, J.; Ihee, H. Protein Tertiary Structural Changes Visualized by Time-Resolved X-Ray Solution Scattering. *J. Phys. Chem. B* **2009**, *113*, 13131-13133.
- (5) Kim, K. H.; Oang, K. Y.; Kim, J.; Lee, J. H.; Kim, Y.; Ihee, H. Direct Observation of Myoglobin Structural Dynamics from 100 Picoseconds to 1 Microsecond with Picosecond X-Ray Solution Scattering. *Chem. Commun.* **2011**, *47*, 289-291.
- (6) Henry, E. R.; Hofrichter, J. Singular Value Decomposition: Application to Analysis of Experimental Data. *Method. Enzymol.* **1992**, *210*, 129-192.
- (7) Jun, S.; Lee, J. H.; Kim, J.; Kim, J.; Kim, K. H.; Kong, Q. Y.; Kim, T. K.; Lo Russo, M.; Wulff, M.; Ihee, H. Photochemistry of HgBr<sub>2</sub> in Methanol Investigated using Time-Resolved X-Ray Liquidography. *Phys. Chem. Chem. Phys.* **2010**, *12*, 11536-11547.