Ultrafast Structural Changes Decomposed from Serial Crystallographic Data

Zhong Ren

Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607, USA
Renz Research, Inc., Westmont, IL 60559, USA

zren@uic.edu

Summary

Direct imaging of ultrafast and subtle structural events during a biochemical reaction, such as a single electronic transition from one atomic or molecular orbital to another, is highly desirable but has been beyond the reach of protein crystallography. It entails the capability of observing changes in electronic distributions at both an ultrafast time scale and an ultrahigh spatial resolution (Itatani et al., Nature 432, 867, 2004). The recent developments in femtosecond serial crystallography at X-ray free electron lasers (XFELs) have brought the achievable temporal resolution within a striking distance. This paper presents the electron density map decomposed from the XFEL data that shows the formation of several 3*d* atomic orbitals of the heme iron tens of femtoseconds after the laser-triggered photolysis. A key strategy that has enabled the findings here is a numerical deconvolution to resolve concurrent variations in a series of time-resolved electron density maps so that the image depicting an electron transfer event can be isolated as a partial change from the overwhelming presence of the bulk electrons that are not directly involved in bonding. The subtle changes in electron distribution due to a spin crossover can be decoupled from far greater changes due to atomic displacements. Direct observations of electronic orbitals could offer unprecedented mechanistic insights into a myriad of chemical and biochemical reactions such as electron transfer in redox reactions, and formation, rupture, and isomerization of chemical bonds.

Ligand photodissociation in carbonmonoxy myoglobin (MbCO) has been a benchmark for studying ultrafast protein dynamics in a biological system (from Šrajer et al., Science 274, 1726, 1996 to Barends et al., Science 350, 445, 2015). This paper presents an in-depth analysis of the serial crystallographic datasets of MbCO that Barends & Schlichting et al. (2015) contributed to the Protein Data Bank. First, an atomic resolution image clearly shows the reappearance of the high-spin 3*d* orbitals of the heme iron upon the photodissociation of the CO ligand. Second, the dynamic behaviors of these newly regained 3*d* orbitals within picoseconds after the photolysis provide long-awaited structural validation for previous spectroscopic observations and computational simulations. Specifically, the newly formed densities are oscillating with the heme iron at a high frequency and developing a positional modulation during the first a few picoseconds (Champion, Science 310, 980, 2005). The dominant oscillation at a low frequency previously detected by coherence spectroscopy can be clearly resolved from the time series of electron density maps. The associated changes in electron density during the first cycle of the oscillation are largely located in the solvent rather than on the protein or heme, which suggests that the low frequency oscillations in a number of heme proteins, including MbCO, likely originate from a photolysis triggered pressure wave propagating in the solvated protein. Finally, these findings of chemical signals are isolated from coexisting thermal artifacts also by the numerical deconvolution. It is modeled in this study that the ultrashort XFEL pulses cause a transient spike of the local temperature at the heme site for hundreds of K.

Keywords
ultrafast dynamics
heme protein
X-ray free electron laser
serial crystallography
singular value decomposition

Abstract

Direct visualization of electronic events during chemical and biochemical reactions will offer unprecedented mechanistic insights. This has been long believed to be beyond the reach of protein crystallography, as it requires the ability to capture alterations in electronic distributions at an ultrafast time scale and an ultrahigh spatial resolution. Ultrashort pulses produced by X-ray free electron lasers (XFELs) offer an unprecedented opportunity for direct observations of transient electronic and molecular events as short-lived as tens of femtoseconds. This paper presents an in-depth analysis of the serial crystallographic datasets collected by Barends & Schlichting et al. (2015) that probe the ligand photodissociation in carbonmonoxy myoglobin (MbCO), a long-serving hallmark for observing ultrafast dynamics in a biological system. This analysis reveals orbital densities that depict the formation of high-spin 3*d* atomic orbitals of the heme iron upon the CO departure and the dynamic behaviors of these newly formed orbitals in a time series within the first a few picoseconds. The regained 3*d* orbitals are found vibrating with the heme iron at a high frequency and developing a positional modulation. These new findings here provide long-awaited visual validations for previous works using ultrafast spectroscopy and molecular dynamics simulations. This analysis also extracts electron density changes largely in the solvent during the first period of a low frequency oscillation previously detected by coherence spectroscopy. This work demonstrates the power and importance of the analytical methods in detecting and isolating transient, often very weak signals of electronic changes arising from chemical reactions in proteins.

Introduction

One of the major applications of X-ray free electron lasers (XFELs) is to directly observe ultrafast protein dynamics at atomic resolution during photochemical reactions by serial femtosecond (fs) crystallography (SFX). Time-resolved experiments (Barends et al., 2015; Coquelle et al., 2017; Nango et al., 2016; Nogly et al., 2018; Pande et al., 2016; Tenboer et al., 2014) have been successfully conducted by several groups at XFELs to follow the photodissociation of carbonmonoxy myoglobin (MbCO; Fig. S1), photoisomerization of the *p*-coumaric acid chromophore in photoactive yellow protein (PYP), and the

hydroxybenzylidene imidazolinone chromophore in a photoswitchable fluorescent protein (Levantino et al., 2015a). While the time resolution, that is, the ability to differentiate transient changes, is limited to tens of fs determined by the duration of the XFEL pulses, the reported time resolution for MbCO and PYP was around hundreds of fs due to some timing jitter between a short laser pump and an XFEL probe.

It is well established that the CO ligand parks at a nearby docking site upon photodissociation, while the heme iron moves out of the heme plane towards the proximal side due to a spin crossover resulting in a high-spin, deoxy iron. As the porphyrin structure becomes buckled, several side chains and helices of the globin move accordingly (Fig. 1). These light-induced structural changes are highly reproducible and have been observed by both static and time-resolved crystallography, repeatedly (Bourgeois et al., 2003, 2006; Chu et al., 2000; Kachalova et al., 1999; Schotte et al., 2003; Šrajer et al., 1996, 2001) . Recently, they have been captured once again at a sub-picosecond (sub-ps) time scale in MbCO crystals using XFEL pulses (Barends et al., 2015), a major achievement in SFX. However, no significantly new structural finding is beyond those already obtained by time-resolved Laue diffraction at synchrotrons (Bourgeois et al., 2007; Clifton et al., 1997; Ren et al., 1999), which is one thousand times slower at a time resolution of 100 ps limited by the bunch length of the synchrotrons (Aranda et al., 2006; Bourgeois et al., 2006; Jung et al., 2013; Ren et al., 2012; Schotte et al., 2012).

The signals of ultrafast structural events could be too weak to be observed clearly by the standard crystallographic analyses. In this work, I reexamine the time-resolved XFEL datasets of MbCO deposited in the Protein Data Bank (PDB; 5cmv, 5cn4-9, and 5cnb-g) by Barends & Schlichting et al. (2015) with a previously developed analytical method based on singular value decomposition (SVD; Ren, 2013a; Ren et al., 2012, 2013; Schmidt et al., 2003). This SVD analysis unravels several ultrafast events that have been identified by spectroscopy and computational studies but have not been directly observed as images of electron density distribution. The new findings along with several previously known structural events upon photolysis of MbCO are summarized in a Gantt chart

(Fig. 1).

Heterogeneous signals in dynamic crystallography

Compared to the well-established structural events of MbCO photolysis such as ligand docking and heme doming, the ultrafast signals are small in amplitude and low in population. As a result, they do not stand out at a given time point, thus went unnoticed in the conventional analyses. As demonstrated below, these weak yet consistent signals can be manifested as major components resulting from SVD when multiple datasets collected at different time points are jointly examined. This necessity for decomposition presents both a challenge and a unique opportunity for dynamic crystallography, broadly defined as experimental and analytical approaches that tightly join crystallographic datasets and their metadata (Ren et al., 2013). We have developed a numerical process of deconvolution, which is able to identify and extract the dynamic signals from a large collection of datasets acquired at varying experimental conditions such as time points and temperatures (Bandara et al., 2017; Kim et al., 2017; Ren, 2016; Ren et al., 2012; Yang et al., 2011; Zeng et al., 2015). In essence, this strategy takes advantage of the structural heterogeneity, rather than preemptively avoids them as in the common practice of static protein crystallography. In a nutshell, dynamic crystallography consists of two major steps: first, to experimentally introduce a functionally-relevant structural heterogeneity; and second, to numerically resolve the heterogeneous conformational species as long as their compositions vary as a function of an experimental parameter (Ren et al., 2013). Here, time-resolved serial crystallography, with the variable being the ultrashort time delay, is simply a special case in the broader scope of dynamic crystallography. Therefore, the previously established methodology is largely applicable (Methods).

Ultrafast dynamics during photodissociation of MbCO

The departure of the CO ligand from the iron porphyrin system has been found within the first 15 fs upon photolysis (Dunietz et al., 2003; Falahati et al., 2017). This signal is clear in the earliest time point of ~30 fs (nominally -0.1 ps). A transient charge transfer from the iron to porphyrin subsequently taking place during 50-300 fs has been identified as a key event by resonance Raman spectra

(Franzen et al., 2001). A high-spin iron is resulting in the charge transfer from the porphyrin back to the iron. The transient charge transfer state is not visible in the maps calculated from the datasets of Barends & Schlichting et al. perhaps due to a broad distribution of the transferred electron over the porphyrin. However, the larger, high-spin iron is clearly manifested in the earliest maps. Vibrational signals have been observed by ultrafast spectroscopy, which is attributed to a discontinuity or "seam" at the intersection of two energy landscapes of the photoexcited MbCO and the ligand dissociated product (Champion, 2005). Dissipation of such vibrational energy in the protein matrix and its solvent environment consists of at least two distinct components: localized, high frequency vibration modes of the heme and global, low frequency oscillations of the protein (Fujisaki and Straub, 2005). This analysis provides direct visual evidences to validate both components and to delineate the low frequency oscillations in greater detail.

The most observed high frequency modes range from hundreds to more than a thousand wavenumbers (Armstrong et al., 2003; Kitagawa et al., 2002; Mizutani, 1997). The corresponding vibrational periods of these high frequency modes of 24-50 fs are comparable to the XFEL pulse duration. Hence an individual cycle of such vibrations is beyond what can be traced out by the current XFEL datasets. Nevertheless, the averaged effect of these high frequency vibrations is detectable in the XFEL datasets of Barends & Schlichting et al. Amplitude modulation and damping of the high frequency mode ν_7 was simulated with a period of 350 fs and a time constant of 1.3 ps (Barends et al., 2015). There are no sufficient time points in the existing datasets to determine these amplitude modulation and damping. However, a slower modulation of the high frequency modes is clearly present in these datasets. These in-plane modes are thought to be coupled to out-of-plane motions of the heme and its iron (Nagy et al., 2005), which take place within a fraction of ps and continue to develop non-exponentially for another hundreds of ps according to molecular dynamics simulations (Henry et al., 1985; Kuczera et al., 1993). Recent simulations (Barends et al., 2015; Falahati et al., 2017) suggested that many transient species, including those with iron in or out of the plane, are mixed together before 1 ps. This analysis of the XFEL datasets of Barends & Schlichting

et al. can confirm that the out-of-plane displacement of the iron is a highly asynchronous event among a molecular population, but this event takes place simultaneously with an in-plane shift of the porphyrin. The signal associated with the iron displacement lingers on for many μs before the ligand rebinding as captured by the time-resolved Laue diffraction (Šrajer et al., 2001).

At a low frequency region, the dominant oscillations range between 25 and 45 cm^{-1}, corresponding to a period around 1 ps (Gruia et al., 2008; Rosca et al., 2002). Such a low frequency oscillation is within the reach of the sub-ps time resolution achieved with XFEL pulses. Therefore, each individual cycle of the low frequency oscillation should be, and indeed is, visible in the datasets of Barends & Schlichting et al. Here I demonstrate that such an oscillation occurs more in the solvent channels than in the protein or the heme at a low frequency comparable to those detected by spectroscopy. This analysis supplies the long missing visual evidence to support such a low-frequency oscillation in a solvated protein.

Imaging the geometric shapes of atomic orbitals

The shape of a *d* orbital of copper was first experimentally imaged by Zuo et al. in their difference electron density map (Humphreys, 1999; Zuo et al., 1999), which is not, however, without controversy (Mulder, 2011; Scerri, 2000). Nevertheless, experimenters continued to draw parallels between captured images, sometimes with reconstructions, and atomic or molecular orbitals. Molecular orbital tomography constructs images of highest occupied molecular orbital from many projections of aligned molecules (Itatani et al., 2004; Ogilvie, 2011; Schwarz, 2006; Stapelfeldt, 2004; Vozzi et al., 2011). Both the highest occupied and lowest unoccupied molecular orbitals of flat molecules, such as pentacene, were scanned using scanning tunneling microscopy (Chaika, 2014; Patera et al., 2019; Repp et al., 2005). Field-emission electron microscopy and angle-resolved photoemission spectroscopy can directly capture or indirectly reconstruct orbital images or transformation of orbital configurations (Mikhailovskij et al., 2009; Puschnig et al., 2009). More recently, a photoionization experiment showed the nodal structures of the atomic orbital of excited hydrogen atoms (Stodolna et al., 2013). A charge-density analysis at an

236 ultrahigh resolution showed iron 3*d* and sulfur 3*p* densities (Hirano et al., 2016).
237 In all these achievements, the experimentally captured electron densities with
238 their unique shapes are directly influenced by electrons occupying specific
239 orbitals. Therefore, in this paper, the images and motion pictures of difference
240 electron densities caused by a superposition of several wave functions are
241 attributed to atomic orbital densities and their dynamics.
242
243 **Results**
244 ***Decomposition of difference maps and simulated annealing omit maps***
245 To analyze the time-resolved serial crystallographic data of horse heart (hh)
246 MbCO, one dark and 12 light datasets are obtained from PDB (Fig. 1; Barends et
247 al., 2015), among which three light datasets collected at the shortest time delays
248 of 0 and ±0.1 ps span both sides of time 0 that is centered on the excitation laser
249 flash of 150 fs. These datasets are expected to capture ultrafast signals at delay
250 times as short as a few fs to tens of fs. However, these ultrafast signals are also
251 inevitably mixed with signals at longer delays as far as the X-ray pulses stretch
252 (Haldrup et al., 2011). These sub-ps time points were taken at the same
253 controlled time delay of 0.5 ps. But the diffraction images were binned according
254 to the timestamp records from an experimental timing tool (Bionta et al., 2011,
255 2014), which produced a number of short delays before 1 ps. Several longer time
256 points ranging from 3 to 150 ps are also included in this analysis (Fig. 1) along
257 with a deoxy hhMb dataset collected from a synchrotron beamline at 7°C, which
258 represents the end product of the MbCO photolysis, and two hhMb structures
259 (1dws/t) in the photolyzed state captured by cryo-trapping (Chu et al., 2000).
260
261 Two types of electron density maps, namely difference Fourier map and
262 simulated annealing omit map (SAOM), are produced and subjected to the SVD
263 analyses separately. Each type carries its own advantage (Methods). While
264 difference Fourier maps are sensitive for detecting weak signals, they are also
265 highly susceptible to systematic errors that often overwhelm the signal. It is
266 critically important to choose a proper reference dataset, which is preferably
267 obtained under the identical experimental condition, if not from the very same
268 crystal as the light dataset. An alternative is to use simulated annealing omit
269 maps (SAOMs), where a small portion of a protein structure is set aside from

structure refinement using the simulated annealing algorithm. In such omit maps, the electron densities within the omitted region shall not be biased by the protein model, thus are considered as the most authentic representation of the experimentally observed electron densities (Hodel et al., 1992). A collection of SAOMs within a common omitted region, although less sensitive to subtle changes compared to difference maps, permits a joint analysis of maps derived under very different experimental conditions, which is usually difficult with difference maps.

Difference Fourier maps are synthesized using a Fourier coefficient set of F_{light}-$F_{\text{reference}}$ combined with the best possible phase set. While $F_{\text{reference}}$ is typically the structure factor amplitudes from a dark dataset (5cmv), a light dataset at a given time point can also be used as a reference (Table S1) to reveal changes with respect to that specific time point. Therefore, a collection of difference maps produced with the same reference dataset constitutes a time series that show continuous changes from the reference time point and on. For example, -F_0 series refers to the difference maps with the reference dataset at the nominal 0 ps. For omit maps, the heme and several side chains within its immediate vicinity are omitted to produce SAOMs of the ligand binding site (Methods). A few static structures determined at room temperature of sperm whale (sw) Mb in the aquomet (1bz6), deoxy (1bzp), and CO-bound (1bzr) forms are also included (Kachalova et al., 1999), along with three hhMb structures (1dwr/s/t) (Chu et al., 2000) as control data points in this joint analysis. The resolutions of these static datasets are truncated to 1.75 Å to match the XFEL datasets.

Altogether, a total of 93 difference maps (Table S1 and Fig. S2) and 20 omit maps are obtained for two independent SVD analyses. The electron densities are masked around the heme site unless stated otherwise (Methods). In this work, the significant signals are identified by the top nine components from SVD (Fig. 2e). A numerical procedure called rotation in the SVD space plays an important role to extract chemically sensible components from these (difference) electron density maps. Proper SVD rotations reveal the correlation between the core data, i.e., electron densities, and their metadata, i.e., delay time, sample temperature, and other experimental parameters (Methods). Given their highly dynamic

nature, the ultrafast signals due to multiple causes are often entangled before suitable rotations are found in the multi-dimensional SVD space (Methods). Using this decomposition and rotation strategy, I identify signals that reveal the electronic redistributions between high and low spin of the iron orbitals.

Molecular image of low-spin to high-spin crossover

The ultrafast signals within several fs to a few tens of fs are manifested in the fourth component, denoted as $\boldsymbol{U}_4$, of 93 difference maps. The component $\boldsymbol{U}_4$ exists only in the map series with the dark dataset as reference (red, $-F_{\mathrm{dark}}$ series in Fig. 2). That is to say, all signals in this component have occurred by the time of the very first time point at ~30 fs (nominally -0.1 ps), therefore the map series $-F_{-0.1}$ shows the cancellation of this component (orange series in Fig. 2). Some of these signals start to decay slightly before 1 ps and diminish significantly after 3 ps (Fig. 2cd). The corresponding map of $\boldsymbol{U}_4$ reveals the previously well-established events arising from the CO docking, out-of-plane motion of the iron, and an uneven heme doming upon photolysis of MbCO (Fig. 3). The negative densities associated with the pyrrole rings and their propionic acid side chains increase in intensity from ring A to B to C to D (Fig. 3c). Positive densities are found at a cross section 1.1 Å above from the heme plane towards the distal side (Fig. 3b).

More remarkably, $\boldsymbol{U}_4$ captures some hitherto unseen signals characterized by a cubic-shaped network of positive densities that cage the iron atom. To facilitate a more concise description, a cross section of $Z = 0$ is defined as the heme plane and the cross section of $X = 0$ is the dividing plane between two propionic acid side chains, from which the third orthogonal cross section of $Y = 0$ can be derived (Fig. S1). The eight corners of the positive cube are oriented along the X and Y cross sections. Strong positive densities appear in the middle of four six-membered rings encircled by adjacent pyrroles, the connecting methane bridge, and the heme iron (marked 1-4 in Fig. 3c). These positive features reach their four minima when they cross the heme plane. These positive densities become more prominent above and below the heme plane (Fig. 3). The eight corners of this cube are located at 1.7±0.2 Å away from the iron. This cubic-shaped network is asymmetric with a notable skew towards CD rings (Fig. 3fh).

Meanwhile, the iron skews in the opposite direction towards the AB rings with an out-of-plane displacement towards the proximal side (Fig. 3e). But the major features of this cube seem more symmetric in the Y cross section.

Similar features are also evident in the seventh component $\boldsymbol{U}_7$ from an independent SVD analysis of 20 SAOMs. It is clear that the seventh dimension after a proper rotation (Methods) represents the reaction coordinate of the ligand dissociation. Mb structures along this reaction coordinate form a lineup on c_7 dimension regardless of sample temperatures: from the ligated structures of MbCO, to several photoexcited species including the static structures Mb*CO and **CO, to the time-resolved structures, eventually to deoxy Mb (Fig. 4e). $\boldsymbol{U}_7$ of the omit maps contains the usual signals of ligand dissociation, such as strong negative densities on the ligand indicating dissociation and a positive layer on the distal side of the heme indicating its doming (Fig. 5), except that the quality of these signals are not as good as those from difference maps because the SAOMs are derived from datasets in very different experiments. Despite the systematic errors, the seventh component $\boldsymbol{U}_7$ of the omit maps again clearly shows the cubic network of positive densities around the iron. This cubic network also skews towards the CD rings of the heme. This is a reoccurrence of the same signals shown in $\boldsymbol{U}_4$ of the difference maps (compare Figs. 3 and 5). The eight corners of the positive cube are 1.9±0.2 Å away from the iron.

It is clear that the positive densities arranged in the cubic network around the iron, either in the difference maps or in the omit maps, originate from the intrinsic signals induced by light. While multiple interpretations of these distinct cube-shaped signals may apply, this positive network clearly indicates a larger, high-spin iron as an immediate overall effect upon photolysis. Here I attribute the distinct shape of the positive feature to instantaneous regaining of the high-spin 3*d* orbitals of the heme iron. The theoretical radii of the 3*d* orbitals for an isolated Fe(II) cation are about 0.36 Å at the maximum probability of these orbitals (Waber and Cromer, 1965). Iron 3*d* orbital densities in a Fe_4S_4 cluster span no further than 1 Å away from the iron as measured in a protein (Hirano et al., 2016). The observed distances between the iron and the corners of the cube are significantly larger. Therefore, the observed cubic densities here are not an

image of a probability distribution of a 3*d* electron. They are resulted from a net gain of electron densities as a fully ligated iron transitions to the high-spin deoxy state when several 3*d* orbitals are reoccupied by unpaired electrons. Second, the deoxy iron remains ligated to the five nitrogen atoms after the ligand dissociation, which is very different from an isolated iron. A calculated image of a combined probability distribution of the textbook $3d_{xz}$ and $3d_{yz}$ with $3d_{xy}$ or $3d_{x^2-y^2}$ orbitals show features highly similar to the observed cubic-shaped map (Fig. 3lm). The strongest electron densities are located at eight corners of the cube where the lobes of 3*d* orbitals intercept. It is entirely possible that the captured positive cube is due to a greater influence of the regained high-spin electrons in several 3*d* orbitals at greater distances from the iron. The observed cubic densities result from not only a superposition of several 3*d* orbitals but also an average of multiple electronic configurations. Such orbital changes have not been captured in protein crystals by X-ray crystallography. Since the non-bonding inner electrons remain largely constant in their distributions, their presence as a majority can easily overwhelm small changes in the minority of electrons involved in bonding. Therefore, it usually requires ultrahigh resolution to observe orbital densities (Hirano et al., 2016). It is very difficult, if not impossible, to detect small orbital changes upon bond rupture using regular electron density maps such as a 2Fo-Fc map. This difference map derived from an SVD component represents a partial, variable signal in addition to the total signal that is largely constant regardless of the bonding change. After the decomposition and a proper rotation, it is highly sensitive to small but consistent changes such as the low-spin to high-spin transition of the heme iron.

Interestingly, the signals in $\boldsymbol{U}_4$ of difference maps decay after a few ps leading up to the static structures Mb*CO, **CO, and the deoxy Mb (Fig. 2cd). Given in-plane vibration modes (Abe et al., 1978; Nagy et al., 2005) with time periods comparable to the XFEL pulse duration, the averaged electron density between vibrating atoms is expected to increase, as shown in the six-membered rings. The observed decay of $\boldsymbol{U}_4$ can be interpreted as the damping of these vibrations at high frequencies (see below). These in-plane vibrations trigger the iron to pop out of the heme plane (Franzen et al., 1995). However, it remains unclear from this observation which, or a combination of which vibration modes, plays the

dominant role. Importantly, this decay does not reach zero even for the static deoxy Mb (Fig. 2de), which suggests that the above interpretation of the cubic network of positive densities as the consequence of the spin crossover largely holds.

Development, modulation, and damping after 1 ps

If the cubic network of positive densities is indeed influenced by the newly formed 3*d* orbitals, the dynamic behaviors of these high-spin electrons should be visible in the time points shortly after the initial formation. The observed evolution of the cubic network provides the necessary validation. Along the dimension c_1 in the SVD space of 93 difference maps, all time points after 1 ps, including the deoxy Mb (5d5r, representing a time point of ∞), are well separated from the sub-ps time points around $c_1 = 0$ (Fig. 2ac). Therefore, the first component $\boldsymbol{U}_1$ of the difference maps represents more permanent changes. The cubic positive densities in $\boldsymbol{U}_4$ have no sign in $\boldsymbol{U}_1$. Among the major features captured by $\boldsymbol{U}_1$ (Fig. S3), the most noticeable is a rather skewed displacement of the iron towards the AB rings of the heme. In the meanwhile, the motion of His93 corroborates the skewed displacement of the iron (Fig. S3ae). It is also shown below that the entire heme moves in its plane towards the AB rings as well. Such long-lived signals in $\boldsymbol{U}_1$ are in good agreement with a wide distribution of individual trajectories and interquartile range of iron displacement predicted by molecular dynamics simulations (Barends et al., 2015; Henry et al., 1985; Kuczera et al., 1993). However, the cryo-trapped structures, Mb*CO and **CO, are two exceptions. They exhibit very little characteristics in $\boldsymbol{U}_1$ (Fig. 2ac), even though they presumably represent long time delays. It can be reasoned that the skewed displacement of iron requires concerted motions in the proximal His93 and helix F, which are evidently forbidden at cryogenic temperatures.

Another major signal revealed after 1 ps is a modulation of the new formation of 3*d* orbitals imaged by the cubic-shaped positive densities around the iron. This is a positional modulation manifested in the sixth and eighth components of the difference maps. While the corresponding coefficients c_6 and c_8 are nearly constant in the sub-ps range, they start to ramp up after 1 ps and oscillate at a

period of 220±20 ps (Fig. 2f). The $\boldsymbol{U}_6$ features a distorted network of positive densities around the heme iron (Fig. S4) similar to $\boldsymbol{U}_4$ but with an inclination towards ring B. In $\boldsymbol{U}_6$, the strong negative density on iron is skewed towards the proximal side in contrast to $\boldsymbol{U}_4$ (compare Figs. 3e and S4c). In a linear combination sense, $\boldsymbol{U}_6$ and $\boldsymbol{U}_8$ can be interpreted as variations that are combined into $\boldsymbol{U}_4$ resulting in an oscillation of a period at a few hundreds of ps (Fig. 2f). Specifically, the cubic-shaped densities lean towards ring D from 0.5 to 3 ps, return to the mean position after 50 ps, then sway towards ring B in the opposite direction (Fig. 2f and Movie S1). Such oscillating behavior suggests modulation and damping of the vibrational modes. In other words, the cubic network of positive densities is not only observed in the earliest time points. The motions of this network are captured in later time points as well. This 220-ps oscillation is largely along BD cross section (Fig. S1), while damping of the vibrational amplitude was only recorded up to 150 ps thus insufficient to depict a complete picture. Two other more significant components $\boldsymbol{U}_2$ and $\boldsymbol{U}_3$ also show an oscillation of a distinct type as discussed below (Fig. 2b).

Vibration induced by XFEL pulses and instantaneous local temperature jump

An independent SVD analysis of 20 simulated annealing omit maps (SAOMs) masked around the heme site (Methods) offers a direct comparison of the XFEL data with previous synchrotron data. Among the top eight components of the SAOMs, the first component reveals a well-defined molecular image of the heme and its surrounding (Fig. S5); and the second and fourth components apparently account for differences between hhMb and swMb. It is the third component $\boldsymbol{U}_3$ that clearly distinguishes the XFEL datasets from the synchrotron datasets in a temperature-dependent manner (Fig. 4ace). In other words, the cryo structures (1dwr-t), ambient temperature structures (1bz6/p/r) including the deoxy hhMb at 7°C (5d5r), and the room temperature XFEL structures line up in an ascending order along the dimension c_3. Not incidentally, the photolyzed structures Mb*CO and **CO have the most negative coefficient c_3 because they were obtained at temperatures even lower than 100 K (Chu et al., 2000). The corresponding decomposed map $\boldsymbol{U}_3$ reveals the thermal effects in electron densities of the heme as a nearly symmetric shell of spherical positive densities around the iron with very little positive or negative density on the iron itself (Fig.

6), which transitions into a square shape with four corners centered at the six-membered rings in the heme plane (marked 1-4 in Fig. 6e). This symmetric $\boldsymbol{U}_3$ of the omit maps shows no usual signals for ligand dissociation and heme doming therefore contrasts with the asymmetric $\boldsymbol{U}_4$ of the difference maps characterized by the distinct cube-shaped positive density maxima at the eight corners. Two sheets of positive densities are found at the cross sections parallel to, and 1.3 Å away from, the heme plane on both sides. They are connected via pillar-shaped positive densities that run through the six-membered rings (marked 1-4 in Fig. 6e) and the pyrrole rings (marked 5-8 in Fig. 6e) along with some other positive densities outside the heme (marked 9-20 in Fig. 6e). It is noteworthy that such thermal effects do not stand out as a major component in the SVD analysis of difference maps because they are largely self-canceled during the subtraction of the reference.

The temperature dependency and the symmetry of the signals in $\boldsymbol{U}_3$ strongly suggest that these features are related to the local temperature at which the MbCO molecules were probed, rather than artifacts arising from structural refinement or difference in spatial resolutions among datasets. All datasets used for the analysis are unified to a similar spatial resolution (Methods). Both the positional and thermal parameters of the model within the omitted region are intentionally "forgotten" in the calculation of the omit maps. The large c_3 coefficients of the XFEL datasets are due to instantaneous temperature elevation at the heme site upon exposure to each ultrashort XFEL pulse, which introduces isotropic vibrations at high frequencies to the entire tetrapyrrole cofactor as captured by $\boldsymbol{U}_3$. This molecular image of $\boldsymbol{U}_3$ shows an increasing intensity from cryo to ambient temperatures and to a very high local temperature induced by XFEL pulses (Movie S2). Based on an estimate by linear extrapolation, the instantaneous temperature of the heme may exceed 500 K during an 80-fs XFEL pulse (Fig. 6i).

Identification of such symmetric and temperature-dependent signals is important for interpretation of the highly dynamic ultrafast signals. The fact that the symmetric $\boldsymbol{U}_3$ and the asymmetric $\boldsymbol{U}_7$ of the omit maps are orthogonal to each other mathematically guarantees that these coexisting components are not

interchangeable and not cross contaminated between them (Methods). The cubic shaped $\boldsymbol{U}_7$ of the omit maps and its counterpart $\boldsymbol{U}_4$ of the difference maps are the chemical consequence of ligand dissociation; the symmetrical and spherical shaped $\boldsymbol{U}_3$ originates from thermal consequence. One cannot compensate for the other and both coexist in all maps. The equivalents can be stated: The cubic positive densities influenced by the 3*d* orbitals are not due to thermal vibrations; the symmetrically enlarged iron and the thickened heme due to an instantaneous temperature elevation have been isolated from the chemically relevant signals. It is worth noting that the ultrafast onset of $\boldsymbol{U}_4$ of the difference maps (Fig. 2cd) is not in contradiction with $\boldsymbol{U}_7$ of the omit maps contained not only in time-resolved XFEL datasets but also in the static structures of deoxy Mb and photolyzed Mb*CO and **CO at different temperatures (Fig. 4e). Instead, regaining of the high-spin 3*d* orbitals of the iron occurs before the first time point at tens of fs; this permanent change lasts throughout the dissociation reaction into the deoxy product.

The asymmetric $\boldsymbol{U}_7$ of the omit maps also describes an important but unfamiliar structural event upon the ligand dissociation – the heme in-plane movement. This type of motion is harder to capture in difference maps due to little net gain or loss of electron density. With respect to the protein framework, the heme moves towards its AB rings in addition to the other changes (Movie S3). Heme in-plane sliding is consistent with the strong signals of the skewed iron displacement and the proximal His93 (Fig. S3) that are closely related to structural responses in the globin. Similar in-plane sliding is even more significant in tetrameric hemoglobins (Baldwin and Chothia, 1979), which is the origin of the cooperative oxygen affinity (Ren, 2013b). However, the functional role of the heme in-plane sliding in Mb is not yet clear and this discussion branches away from the topic here on ultrafast structural changes.

Global oscillation at low frequency

A recent solution scattering experiment using XFEL pulses showed that the radius of gyration R_g first increases within 1 ps upon photodissociation of MbCO and then oscillates with a period of 3.6 ps, equivalent to 9 cm^{-1}, a frequency lower than those observed by coherence spectroscopy (Levantino et al., 2015b). The

molecular volume increases afterward with a phase shift. These observations indicate a wavelike motion propagating outward in the protein. Recent molecular dynamics simulations have confirmed that a pressure wave generated by photolysis reaches the molecular surface at 300 fs along the direction perpendicular to the heme plane (Brinkmann and Hub, 2016), which corresponds to a frequency of 28 cm^{-1} if the wave is reflected by the surface and oscillates. It was found by the simulations that the oscillation in solvent would be overdamped thus dissipate quickly.

Some SVD components derived here display strict correlation suggesting their dependency on each other. The $\boldsymbol{U}_2$-$\boldsymbol{U}_3$ pair of the difference maps (Fig. 2b) is evidently equivalent to the $\boldsymbol{U}_5$-$\boldsymbol{U}_8$ pair of the omit maps (Fig. 4b). When the difference maps of the entire globin are analyzed, the same signals emerge as the top-ranked $\boldsymbol{U}_1$-$\boldsymbol{U}_3$ pair suggesting that the local signals at the heme site concur with global structural changes (Fig. 7). Interestingly, this pair of components displays a near-perfect circular correlation at sub-ps delays but is almost absent at longer delays > 1 ps (Fig. 7afgh). Such an oscillation at a low frequency is depicted by at least two orthogonal components. The oscillating signals have spread over the entire globin at sub-ps delays while their amplitudes decay to nearly zero after a few ps. Based on least-squares fittings of the corresponding coefficients (Methods), the frequency of the oscillation is determined to be 36±1 cm^{-1}, corresponding to a period of 0.93 ps. The equivalent oscillation frequencies determined from the difference maps and omit maps at the heme site are 42±2 cm^{-1} (a period of 0.8 ps) and 37±1 cm^{-1} (a period of 0.89 ps), respectively.

The low frequency oscillation determined from the XFEL datasets of Barends & Schlichting et al. is highly comparable to those measured by fs vibrational coherence spectroscopy (Gruia et al., 2008; Rosca et al., 2002). I postulate that this low frequency reflects the molecular size of the solvated Mb rather than the property of the heme. First, the corresponding period of 0.93 ps is about the time that a mechanical wave takes to traverse 14 Å, that is, an average radius of Mb, given a speed of sound at 1.5 km/s or nm/ps (Fig. S1). Second, the molecular sizes of several heme proteins, such as cytochrome c, Mb, and cystathionine β-synthase, show good correlation with their dominant low frequency of 25, 40,

and 44 cm^{-1}, respectively (Gruia et al., 2008; Karunakaran et al., 2010, 2014). Third, no association between the oscillating signals and specific structural features cab be found, such as the heme, ligand, surrounding side chains, or helices (Fig. 7ik). These oscillating signals sharply contrast with strong, non-oscillating signals associated with the heme and various helices evidenced by the second and fourth components of the difference maps (Fig. 7jl), where helices E, G, and H move away from the heme in response to ligand dissociation (Barends et al., 2015). These motion signals of various structural elements are much enhanced by SVD.

The oscillating signals at low frequency seem to propagate via solvent within the protein matrix rather than the globin. Strong positive and negative peaks in $\boldsymbol{U}_5$, $\boldsymbol{U}_6$, and $\boldsymbol{U}_8$ of the omit maps are consistently found in the solvent region away from the heme and protein. However, a specific description on how the solvent responses to ligand dissociation is difficult, since the solvent structure is far less well determined than the globin structure. The total contribution of $\boldsymbol{U}_5$, $\boldsymbol{U}_6$, and $\boldsymbol{U}_8$ can be judged by a sum of the absolute values $|\boldsymbol{U}_5|+|\boldsymbol{U}_6|+|\boldsymbol{U}_8|$ that enhances the distribution of signals. The heme and globin are located in low intensities of these oscillating components (Fig. 7b-e). These global oscillations occur only a few times before dissipation (Gruia et al., 2008; Rosca et al., 2002). However, there are not enough time points to determine the damping time constant, which is roughly estimated to be 1-4 ps. Contrary to the common belief that low frequency oscillations are due to the properties of the heme, here I attribute the low frequency oscillation to the global response of protein and solvent as a whole to a sudden molecular event – photodissociation of the ligand, where the vibrational energy released from the photodissociation is quickly dissipated into the environment.

Discussion

Spatial and temporal resolutions vs. resolution of heterogeneity

Ultrashort XFEL pulses improve the time resolution to sub-ps or even tens of fs compared to 100 ps at synchrotrons. This ultrafast time resolution, when combined with the spatial resolution of crystallography, offers an unprecedented opportunity for direct observation of transient strucutral changes in a

photochemical reaction as detail as individual electrons. However, due to their highly dynamic and heterogeneous nature, it is difficult to experimentally isolate short-lived structural species along the temporal dimension regardless of the ultrafast time resolution. As demonstrated by this work, analytical methods such as SVD are needed to deconvolute them from mixtures of intermediate species. The strategy we proposed for resolution of structural heterogeneity (Ren et al., 2013) is highly effective for separating independent electronic and atomic events that often coexist during a chemical reaction or a biological process.

Protein crystallography is proven to be a powerful tool for capturing atomic motions involving displacements of tens of electrons such as side chain rotation or heme iron movement by a few tenths of an Å. However, it has been long believed that ultrahigh spatial resolution is required for direct observation of changes involving fewer electrons such as a single electron transfer. Such a notion contradicts the fact that there is no discontinuity in the probability distribution of any electronic orbital. Therefore, a reasonably high spatial resolution shall be sufficient to analytically describe the smooth function of a gain or loss of an electron. In reality, any gain or loss of a few electrons would be overwhelmed by the strong presence of static distributions of the bulk electrons. Difference Fourier maps must be used to reveal such small changes. Furthermore, difference signals are often dominated by displacements of the nuclei that carry the displacements of many electrons, which in turn conseals subtle changes of electronic configuration. Therefore, the small amplitude of the desired changes over a much greater known and variable quantity hinders a reliable observation more than the lack of high spatial frequencies does. This work presents the effectiveness of the strategy that enables numerical deconvolution to resolve concurrent events so that the image of a partial change, such as a single electron transfer, can be isolated from the total change.

This work shows that it is possible to observe structural changes as small as a gain or loss of individual electrons by protein crystallography. Small structural signals can be correlated with their corresponding metadata, thus allowing the origin of these signals be revealed. I posit that all observations so far are the consequence, rather than the cause, of the ligand dissociation. The proposed

transient event of metal-to-ligand charge transfer is not imaged here, presumably because the transferred electron is delocalized all over the porphyrin (Franzen et al., 2001). Therefore, the fundamental questions remain for photolysis of MbCO: Upon the absorption of a quantum of energy from a green photon, what initial electronic events lead to a partitioning of the bonding electrons in two separate atoms, that is, the bond rupture? What electronic events cause spin crossover? To image these events directly, even higher time resolution is required to capture electron density changes prior to and during bond rupture.

Future outlook of ultrafast dynamic crystallography

Despite the recent developments in SFX and its applications to time-resolved studies of a few model systems, it is yet premature to declare that the current SFX platform applies to other systems of more significant biological importance such as visual and microbial rhodopsins, plant and bacterial phytochromes, photosynthetic reaction centers, and light-dependent DNA repair enzymes. There are at least three major challenges. First, it is not practical to produce and inject an astronomical number of nano or microcrystals of these proteins into an XFEL beam to support data collections of a time series. The current serial protocol used to obtain the MbCO and other data has a diminishingly small yield of useful diffraction data, based on the ratio between the number of diffracting crystals and the total number of crystals produced (Ren et al., 2018). That is to say, the vast majority of the crystals have no chance to diffract X-ray even when they are injected into the X-ray beam. Such a yield, defined differently from the hit rate, is not sustainable for other photosensitive protein crystals. Several "fixed target" implementations of crystal delivery systems have been reported recently, including one that we have recently proposed (Ren et al., 2018). However, it remains to be seen how effective they are in achieving a higher yield while maintaining the diffraction quality.

Second, since the XFEL pulses usually carry insufficient bandwidths to fully integrate the Bragg reflections and in the meanwhile angular rotation of a crystal during an ultrashort XFEL pulse is not attainable, tens to hundreds of thousands of diffraction images are required to obtain integrated intensities for all Bragg reflections in a complete sampling of the reciprocal space. A possibility to solve

this partiality problem is to use an X-ray beam with a significant convergence of about a degree (Ho et al., 2002; Spence et al., 2014), which will produce completely integrated intensities with a far fewer number of diffraction images, thereby alleviating the demand for a large quantity of purified proteins.

Third, data merging from a large crystal pool is a major source of error in producing difference signals for photosensitive crystals, particularly those with much lower diffraction power than Mb crystals. To address this challenge, we also proposed a protocol to obtain difference signals that bypasses the integrated intensities of Bragg reflections. We reason that integrated intensities are not necessary if the dark and light diffraction images can be obtained within a single exposure of an XFEL pulse because the percentage change in integrated intensities of a given reflection, that is, the time-resolved signal, is already captured even if this reflection is partially measured in both dark and light conditions. A plan has been laid to implement this protocol, based on a dichromator instrument that achieves angular split and temporal delay at the same time (Ren and Yang, 2016). When each of these technical hurdles is overcome, direct observations of atomic and electronic movements during ultrafast processes will be closer to broad applications for many light sensitive crystals of great biological and medical significance.

Methods

Difference Fourier maps

Difference Fourier maps are commonly used in protein crystallography. A difference Fourier map is synthesized from a Fourier coefficient set of F_{light}-$F_{reference}$ with the best available phase set. Before Fourier synthesis, F_{light} and $F_{reference}$ must be properly scaled to the same level so that the distribution of difference values is centered at 0 and not skewed either way. All available datasets from Barends & Schlichting et al. (2015) are carefully scaled and difference Fourier maps are calculated using a software package dynamiX™ according to previously published protocols including a weighting scheme under the assumption that a greater amplitude of a difference Fourier coefficient F_{light}-$F_{reference}$ is more likely caused by noise than by signal (Ren et al., 2001, 2013; Šrajer et al., 2001; Ursby and Bourgeois, 1997).

Both the dark and light datasets can serve as a reference in difference maps. If a light dataset at a certain delay is chosen as a reference, the difference maps show the changes since that delay time but not the changes prior to that delay. These 91 maps in a full matrix provide independent observations for all changes (Table S1). This strategy is important to the subsequent singular value decomposition (SVD) analysis. In addition, two more difference maps of the photolyzed Mb*CO and **CO (1dws/t) are also included in the -F_{dark} series for SVD analysis.

Each difference map is masked around the heme site. All densities at those grid points are set to zero, if they are greater than 5 Å away from the heme, the bound or docked CO, the side chain of the proximal His93, and that of the distal His64. Difference maps around the heme site include the most significant signals and exclude noises in the protein and solvent. However, the unmasked difference maps in the entire globin are also used a separate SVD analysis to include the protein and solvent regions for observation of global changes (Fig. 7i-l).

Simulated annealing omit maps

The heme, the ligand, the side chains of proximal His93 and distal His64 are removed from all structures listed in Table S2. The remaining structures are refined using simulated annealing method with phenix.refine starting from a simulated temperature of 5,000 K. Fourier coefficients are obtained from the differences between the structure factor amplitudes $F_{observed}$ observed in the time-resolved experiments and the structure factor amplitudes $F_{calculated}$ calculated from the omitted structures. The resulting $F_{observed}$-$F_{calculated}$ maps, known as simulated annealing omit maps (SAOMs), reveal the unbiased electron densities in the omitted region. Since the sperm whale (sw) Mb structures are in a different unit cell compared to the horse heart (hh) Mb structures, the SAOMs are aligned together according to a least-squares fitting of the main chain atoms in residues 2-151. The heme and ligand are not used in the alignment procedure. Therefore, the motion of heme found in these omit maps is with respect to the globin

framework (Fig. 5 and Movie S3). 20 aligned omit maps are subjected to the subsequent SVD analysis.

Singular value decomposition of (difference) electron density maps

The principle of SVD can be found in texts on linear algebra. SVD analysis of (difference) electron density maps was previously described (Ren et al., 2013; Schmidt et al., 2003). Here I briefly recap this numerical procedure with an emphasis on the physical meaning of the outcome.

An electron density map, sometimes a difference map, consists of density values on a collection of grid points within a mask (see above), also known as voxels that are equivalent to pixels in a digitized image. All M voxels in a three-dimensional map can be serialized into a one-dimensional sequence of density values according to a specific protocol. It is not important what the protocol is as long as a consistent protocol is used to serialize all maps of the same grid setting and size, and a reverse protocol is available to erect a three-dimensional map from a sequence of M densities. Therefore, a set of N serialized maps, also known as vectors in linear algebra, can fill the columns of a data matrix $\mathbf{A}$ with no specific order so that the width of $\mathbf{A}$ is N columns and the length is M rows. Often, $M >> N$, thus $\mathbf{A}$ is an elongated matrix. In this work, $N = 93$ for the difference maps and $N = 20$ for the omit maps. If a consistent protocol of serialization is used, the corresponding voxel in all N maps occupies a single row of matrix $\mathbf{A}$. For example, the voxels at the heme iron in all maps could occupy the jth row of matrix $\mathbf{A}$. This strict correspondence in a row of matrix $\mathbf{A}$ is important. Changes of the density values in a row are due to either signals, systematic errors, or noises.

SVD of the data matrix $\mathbf{A}$ results in $\mathbf{A} = \mathbf{UWV}^{\mathbf{T}}$, also known as matrix factorization. Matrix $\mathbf{U}$ has the same shape as $\mathbf{A}$, that is, N columns and M rows. The N columns contain decomposed components $\boldsymbol{U}_k$, known as left singular vectors of length M, where $k = 1, 2, \ldots, N$. Therefore, each component $\boldsymbol{U}_k$ can be erected using the reverse protocol to form a three-dimensional map. This decomposed elemental map can be presented in the same way as the original maps, for example, rendered in molecular graphics software such as Coot and

PyMol. The second matrix **W** is a square matrix that contains all zeros except for N positive values on its major diagonal, known as singular values w_k. The magnitude of w_k is considered as a weight or significance of its corresponding component $\boldsymbol{U}_k$. The third matrix **V** is also a square matrix of $N \times N$. Each column of **V** or row of its transpose $\mathbf{V}^{\mathrm{T}}$, known as a right singular vector, contains the relative compositions of $\boldsymbol{U}_k$. That is to say, the N values in a column $\boldsymbol{V}_k$ describe the relative compositions of the component $\boldsymbol{U}_k$ in the N original maps. A singular triplet denotes 1) a decomposed component $\boldsymbol{U}_k$, 2) its singular value w_k, and 3) the composition function $\boldsymbol{V}_k$. Singular triplets are often sorted in a descending order of their singular values w_k. Only a small number of n significant singular triplets identified by the greatest singular values w_1 through w_n can be used in a linear combination to reconstitute a set of composite maps that closely resemble the original ones in matrix **A**, where $n < N$. For example, the original map in the ith column of matrix **A** can be closely represented by the ith composite map $w_1 v_{1i} \boldsymbol{U}_1 + w_2 v_{2i} \boldsymbol{U}_2 + \ldots + w_n v_{ni} \boldsymbol{U}_n$, where $(v_{1i}, v_{2i}, \ldots)$ is the ith row of matrix **V**. $c_k = w_k v_{ki}$ in front of $\boldsymbol{U}_k$ is the coefficients for the linear combination. Excluding the components after $\boldsymbol{U}_n$ in this linear combination is based on an assumption that the singular values after w_n are very small relative to those from w_1 through w_n. As a result, the structural information evenly distributed in all N original maps is effectively concentrated into a far fewer number of n significant components, known as information concentration or dimension reduction. In this study, the difference maps seem to contain 15 major components; the top nine are the most significant (Fig. 2e). The components 10 through 15 may very likely harbor less and less signals or systematic differences.

On the other hand, the trailing components in matrix **U** contain inconsistent fluctuations and random noises. Excluding these components effectively rejects noises. However, no clear boundary is guaranteed between signals, systematic errors, and noises. Systematic errors could be more significant than the desired signals, for example, the second and fourth components $\boldsymbol{U}_2$ and $\boldsymbol{U}_4$ of the omit maps describe significant differences between hhMb and swMb, a valid scientific point to investigate but outside the scope of this paper.

The orthonormal property of SVD

The solution set of SVD must guarantee that the columns in **U** and **V**, the left and right singular vectors $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$, are orthonormal, that is, $\boldsymbol{U}_h \bullet \boldsymbol{U}_k = \boldsymbol{V}_h \bullet \boldsymbol{V}_k = 0$ (ortho) and $\boldsymbol{U}_k \bullet \boldsymbol{U}_k = \boldsymbol{V}_k \bullet \boldsymbol{V}_k = 1$ (normal), where $h \neq k$ but both are from 1 to N. The orthonormal property also holds for the row vectors. As a result, each component $\boldsymbol{U}_k$ is independent of the other components. In other words, a component cannot be represented by a linear combination of any other components. For example, the symmetrical, spherical, and thermal component $\boldsymbol{U}_3$ (Fig. 6) and the asymmetrical, cubic component of $\boldsymbol{U}_7$ (Fig. 5) of the omit maps are orthonormal. Both components show positive densities around the iron. But the cubic positive densities in $\boldsymbol{U}_7$ are associated with the strong negative densities on the ligand depicting dissociation, which is a direct link from ligand dissociation to regaining of the high-spin 3*d* orbitals of the iron. On the other hand, no sign of ligand dissociation is found in the component $\boldsymbol{U}_3$ that features the spherical densities, which strongly suggests that this component describes the thermal effect. These two components are not interchangeable despite some similarity in the arrangement of positive densities, that is, one of the physical meanings of the orthonormal property of SVD.

Due to the orthonormal property of SVD, an N-dimensional Euclidean space is established, and the first n dimensions define its most significant subspace. Each coefficient set $c_k = w_k v_{ki}$ of the ith composite map is located in this subspace. All coefficient sets for $i = 1, 2, \ldots, N$ in different linear combinations to approximate the N original maps can be represented by N points in the n-dimensional Euclidean subspace. In this work, many two-dimensional orthographical projections of this n-dimensional subspace are presented as scatter plots. These scatter plots are highly informative to reveal the relationship between the (difference) electron density maps and their metadata. For example, the difference maps at short time delays < 1 ps are consistently different from those at long time delays > 1 ps in a certain aspect (Fig. 2ac) and that two major components are circularly correlated but only before 1 ps, that is, a damped oscillation (Fig. 2b).

Rotation in SVD space

Dimension reduction is indeed effective in data analysis of dynamic crystallography when many datasets are evaluated at the same time. However, the interpretation of a given component $\boldsymbol{U}_k$, that is, "what-does-it-mean", requires a clear demonstration of the relationship between the core data, here (difference) electron density maps, and their metadata, such as time, temperature, and other experimental conditions. The outcome of SVD does not provide a direct answer to "what-does-it-mean". The factorized set of matrices **U**, **W**, and **V** from SVD is not a unique solution. That is to say, they are not the only solution to factorize matrix **A**. The default solution set from SVD may not carry an obvious physical meaning to depict "what-does-it-mean", thus its usefulness is very limited to merely a mathematical construction. This is also the origin of the debate on whether atomic orbitals as a basis set of mathematical constructions are physical and observable (Scerri, 2000). It is very important to find one or more alternative solution sets that can elucidate a structural interpretation. This process, called rotation in SVD space, also known as Givens rotation, is a connection between the analytical tool SVD and scientific findings. This rotation shall not be confused with a rotation in the three-dimensional real space, in which a molecular structure resides.

A rotation in the n-dimensional Euclidean subspace is necessary to change the perspective before a clear relationship emerges to elucidate scientific findings. It can be shown that two linear combinations are identical before and after a rotation of an angle θ in a two-dimensional subspace of h and k. That is,

$$c_h\boldsymbol{U}_h + c_k\boldsymbol{U}_k = f_h\boldsymbol{R}_h + f_k\boldsymbol{R}_k, \tag{1}$$

where c_h and c_k are the coefficients of the components $\boldsymbol{U}_h$ and $\boldsymbol{U}_k$ before the rotation as defined above for the ith map; and f_h and f_k are the coefficients of the rotated components $\boldsymbol{R}_h$ and $\boldsymbol{R}_k$, respectively. The same Givens rotation is applied to both the components and their coefficients:

$$\begin{cases} \boldsymbol{R}_h = \boldsymbol{U}_h\cos\theta - \boldsymbol{U}_k\sin\theta; \\ \boldsymbol{R}_k = \boldsymbol{U}_h\sin\theta + \boldsymbol{U}_k\cos\theta. \end{cases} \tag{2}$$

Obviously, the rotated components $\boldsymbol{R}_h$ and $\boldsymbol{R}_k$ remain mutually orthonormal and orthonormal to other components. And

$$\begin{cases} f_h = s_h t_h = c_h \cos\theta - c_k \sin\theta = w_h v_h \cos\theta - w_k v_k \sin\theta; \\ f_k = s_k t_k = c_h \sin\theta + c_k \cos\theta = w_h v_h \sin\theta + w_k v_k \cos\theta. \end{cases} \tag{3}$$

Here $s_{h|k} = \sqrt{\sum f_{h|k}^2}$ are the singular values that replace w_h and w_k, respectively, after the rotation. They may increase or decrease compared to the original singular values so that the descending order of the singular values no longer holds. $\boldsymbol{T}_{h|k} = (t_{h|k1}, t_{h|k2}, \ldots, t_{h|kN}) = (f_{h|k1}, f_{h|k2}, \ldots, f_{h|kN})/s_{h|k}$ are the right singular vectors that replace $\boldsymbol{V}_h$ and $\boldsymbol{V}_k$, respectively. $\boldsymbol{T}_h$ and $\boldsymbol{T}_k$ remain mutually orthonormal after the rotation and orthonormal to other right singular vectors that are not involved in the rotation.

A rotation in two-dimensional subspace of h and k has no effect in other dimensions, as the orthonormal property of SVD guarantees. Multiple Givens rotations can be carried out in many two-dimensional subspaces consecutively to achieve a multi-dimensional rotation. Rotation in the SVD space converts one solution set $\mathbf{A} = \mathbf{UWV}^{\mathrm{T}}$ to other alternative solutions $\mathbf{A} = \mathbf{RST}^{\mathrm{T}}$ so that an appropriate perspective can be found to elucidate the relationship between the core data and metadata clearly and concisely. A solution derived from a rotation retains the orthonormal property of SVD. SVD analyses presented in this paper employ rotations extensively except that no distinction is made in the symbols of components and coefficients before and after a rotation.

Least-squares fittings of oscillating components

Oscillating components are repeatedly found in the difference maps and omit maps. Usually, a pair of components are oscillating with the same frequency, which does not exclude the possibility that a third weaker component is also oscillating at the same frequency. A pair of coefficients in a two-dimensional damping oscillation of a period p is least-squares fitted as x and y:

$$\begin{cases} x = x_0 + a e^{-t/\tau} \sin(2\pi t/p + \varphi)\,; \\ y = y_0 + b e^{-t/\tau} \sin(2\pi t/p + \psi)\,. \end{cases} \tag{4}$$

x_0 and y_0 are the center of the oscillation. a and b are the amplitudes of the elliptical oscillation in two dimensions. They are damped according to an exponential decay with a time constant τ. φ and ψ are the phase angles of the oscillation in two dimensions. They are not strictly $\pi/2$ apart.

The sixth and eighth components of the difference maps appear to gain a two-dimensional oscillation after 1 ps (Fig. 2f). Their coefficients are least-squares fitted as x and y with an exponential onset:

$$\begin{cases} x = x_0 + a\left(1 - e^{-t/\tau}\right)\cos(2\pi t/p + \varphi)\,; \\ y = y_0 + a\left(1 - e^{-t/\tau}\right)\sin(2\pi t/p + \varphi)\,. \end{cases} \tag{5}$$

Due to the lack of sufficient time points, the oscillation is forced to take a circular form.

Code availability

Computer software dynamiX™ is available from Renz Research, Inc.

Acknowledgements

The following database and software are used in this work: CCP4 (ccp4.ac.uk), Coot (www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot), dynamiX™ (Renz Research, Inc.), gnuplot (gnuplot.info), Mathematica (wolfram.com), PDB (rcsb.org), PHENIX (phenix-online.org), PyMOL (pymol.org), Python (python.org), and SciPy (scipy.org).

Competing interests

ZR is the founder of Renz Research, Inc. that currently holds the copyright of the computer software dynamiX™.

References

Abe, M., Kitagawa, T., and Kyogoku, Y. (1978). Resonance Raman spectra of octaethylporphyrinato-Ni(II) and meso-deuterated and 15N substituted derivatives. II. A normal coordinate analysis. J. Chem. Phys. *69*, 4526–4534.

Aranda, R., Levin, E.J., Schotte, F., Anfinrud, P.A., and Phillips, G.N. (2006). Time-dependent atomic coordinates for the dissociation of carbon monoxide from myoglobin. Acta Crystallogr. D Biol. Crystallogr. *62*, 776–783.

Armstrong, M.R., Ogilvie, J.P., Cowan, M.L., Nagy, A.M., and Miller, R.J.D. (2003). Observation of the cascaded atomic-to-global length scales driving protein motion. Proc. Natl. Acad. Sci. *100*, 4990–4994.

Baldwin, J., and Chothia, C. (1979). Haemoglobin: the structural changes related to ligand binding and its allosteric mechanism. J. Mol. Biol. *129*, 175–200.

Bandara, S., Ren, Z., Lu, L., Xiaoli, Z., Shin, H., Zhao, K.-H., and Yang, X. (2017). Photoactivation mechanism of a carotenoid-based photoreceptor. Proc. Natl. Acad. Sci. U. S. A. *114*, 6286–6291.

Barends, T.R.M., Foucar, L., Ardevol, A., Nass, K., Aquila, A., Botha, S., Doak, R.B., Falahati, K., Hartmann, E., Hilpert, M., et al. (2015). Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. Science *350*, 445–450.

Bionta, M.R., Lemke, H.T., Cryan, J.P., Glownia, J.M., Bostedt, C., Cammarata, M., Castagna, J.-C., Ding, Y., Fritz, D.M., Fry, A.R., et al. (2011). Spectral encoding of x-ray/optical relative delay. Opt. Express *19*, 21855.

Bionta, M.R., Hartmann, N., Weaver, M., French, D., Nicholson, D.J., Cryan, J.P., Glownia, J.M., Baker, K., Bostedt, C., Chollet, M., et al. (2014). Spectral encoding method for measuring the relative arrival time between x-ray/optical pulses. Rev. Sci. Instrum. *85*, 083116.

Bourgeois, D., Vallone, B., Schotte, F., Arcovito, A., Miele, A.E., Sciara, G., Wulff, M., Anfinrud, P., and Brunori, M. (2003). Complex landscape of protein structural dynamics unveiled by nanosecond Laue crystallography. Proc. Natl. Acad. Sci. *100*, 8704–8709.

Bourgeois, D., Vallone, B., Arcovito, A., Sciara, G., Schotte, F., Anfinrud, P.A., and Brunori, M. (2006). Extended subnanosecond structural dynamics of myoglobin revealed by Laue crystallography. Proc. Natl. Acad. Sci. *103*, 4924–4929.

Bourgeois, D., Schotte, F., Brunori, M., and Vallone, B. (2007). Time-resolved methods in biophysics. 6. Time-resolved Laue crystallography as a tool to investigate photo-activated protein dynamics. Photochem. Photobiol. Sci. *6*, 1047.

Brinkmann, L.U.L., and Hub, J.S. (2016). Ultrafast anisotropic protein quake propagation after CO photodissociation in myoglobin. Proc. Natl. Acad. Sci. *113*, 10565–10570.

Chaika, A.N. (2014). Visualization of electron orbitals in scanning tunneling microscopy. JETP Lett. *99*, 731–741.

Champion, P.M. (2005). Chemistry: Following the flow of energy in biomolecules. Science *310*, 980–982.

Chu, K., Vojtchovský, J., McMahon, B.H., Sweet, R.M., Berendzen, J., and Schlichting, I. (2000). Structure of a ligand-binding intermediate in wild-type carbonmonoxy myoglobin. Nature *403*, 921–923.

Clifton, I.J., Duke, E.M.H., Wakatsuki, S., and Ren, Z. (1997). Evaluation of Laue diffraction patterns. Meth Enzym. *277*, 448–467.

Coquelle, N., Sliwa, M., Woodhouse, J., Schirò, G., Adam, V., Aquila, A., Barends, T.R.M., Boutet, S., Byrdin, M., Carbajo, S., et al. (2017). Chromophore twisting in the excited state of a

photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. Nat. Chem. *10*, 31–37.

Dunietz, B.D., Dreuw, A., and Head-Gordon, M. (2003). Initial steps of the photodissociation of the CO ligated heme group. J. Phys. Chem. B *107*, 5623–5629.

Falahati, K., Tamura, H., Burghardt, I., and Huix-Rotllant, M. (2017). Quantum dynamics of ultrafast heme-carbon monoxide photolysis and spin-crossover. ArXiv170903402 Cond-Mat Physicsphysics.

Franzen, S., Bohn, B., Poyart, C., DePillis, G., Boxer, S.G., and Martin, J.-L. (1995). Functional aspects of ultra-rapid heme doming in hemoglobin, myoglobin, and the myoglobin mutant H93G. J. Biol. Chem. *270*, 1718–1720.

Franzen, S., Kiger, L., Poyart, C., and Martin, J.-L. (2001). Heme photolysis occurs by ultrafast excited state metal-to-ring charge transfer. Biophys. J. *80*, 2372–2385.

Fujisaki, H., and Straub, J.E. (2005). Vibrational energy relaxation in proteins. Proc. Natl. Acad. Sci. *102*, 6726–6731.

Gruia, F., Kubo, M., Ye, X., and Champion, P.M. (2008). Investigations of vibrational coherence in the low-frequency region of ferric heme proteins. Biophys. J. *94*, 2252–2268.

Haldrup, K., Harlang, T., Christensen, M., Dohn, A., van Driel, T.B., Kjær, K.S., Harrit, N., Vibenholt, J., Guerin, L., Wulff, M., et al. (2011). Bond shortening (1.4 Å) in the singlet and triplet excited states of [Ir2(dimen)4]2+ in solution determined by time-resolved X-ray scattering. Inorg Chem *50*, 9329–9336.

Henry, E.R., Levitt, M., and Eaton, W.A. (1985). Molecular dynamics simulation of photodissociation of carbon monoxide from hemoglobin. Proc. Natl. Acad. Sci. U. S. A. *82*, 2034–2038.

Hirano, Y., Takeda, K., and Miki, K. (2016). Charge-density analysis of an iron–sulfur protein at an ultra-high resolution of 0.48 Å. Nature *534*, 281–284.

Ho, J.X., Ruble, J.R., McInnis, T.R., Carter, D.C., Huang, H., and Gibson, W.M. (2002). Convergent-beam method in macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *58*, 2087–2095.

Hodel, A., Kim, S.H., and Brünger, A.T. (1992). Model bias in macromolecular crystal structures. Acta Crystallogr. A *48*, 851–858.

Humphreys, C.J. (1999). Electrons seen in orbit. Nature *401*, 21–22.

Itatani, J., Levesque, J., Zeidler, D., Niikura, H., Pépin, H., Kieffer, J.C., Corkum, P.B., and Villeneuve, D.M. (2004). Tomographic imaging of molecular orbitals. Nature *432*, 867–871.

Jung, Y.O., Lee, J.H., Kim, J., Schmidt, M., Moffat, K., Šrajer, V., and Ihee, H. (2013). Volume-conserving trans–cis isomerization pathways in photoactive yellow protein visualized by picosecond X-ray crystallography. Nat. Chem. *5*, 212–220.

Kachalova, G.S., Popov, A.N., and Bartunik, H.D. (1999). A steric mechanism for inhibition of CO binding to heme proteins. Science *284*, 473–476.

Karunakaran, V., Benabbas, A., Sun, Y., Zhang, Z., Singh, S., Banerjee, R., and Champion, P.M. (2010). Investigations of low-frequency vibrational dynamics and ligand binding kinetics of cystathionine β-synthase. J. Phys. Chem. B *114*, 3294–3306.

Karunakaran, V., Sun, Y., Benabbas, A., and Champion, P.M. (2014). Investigations of the low frequency modes of ferric cytochrome c using vibrational coherence spectroscopy. J. Phys. Chem. B *118*, 6062–6070.

Kim, T.H., Mehrabi, P., Ren, Z., Sljoka, A., Ing, C., Bezginov, A., Ye, L., Pomès, R., Prosser, R.S., and Pai, E.F. (2017). The role of dimer asymmetry and protomer dynamics in enzyme catalysis. Science *355*, eaag2355.

Kitagawa, T., Haruta, N., and Mizutani, Y. (2002). Time-resolved resonance Raman study on ultrafast structural relaxation and vibrational cooling of photodissociated carbonmonoxy myoglobin. Biopolymers *67*, 207–213.

Kuczera, K., Lambry, J.-C., Martin, J.-L., and Karplus, M. (1993). Nonexponential relaxation after ligand dissociation from myoglobin: A molecular dynamics simulation. Proc. Natl. Acad. Sci. U. S. A. *90*, 5805–5807.

Levantino, M., Yorke, B.A., Monteiro, D.C., Cammarata, M., and Pearson, A.R. (2015a). Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules. Curr. Opin. Struct. Biol. *35*, 41–48.

Levantino, M., Schirò, G., Lemke, H.T., Cottone, G., Glownia, J.M., Zhu, D., Chollet, M., Ihee, H., Cupane, A., and Cammarata, M. (2015b). Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. Nat. Commun. *6*, 6772.

Mikhailovskij, I.M., Sadanov, E.V., Mazilova, T.I., Ksenofontov, V.A., and Velicodnaja, O.A. (2009). Imaging the atomic orbitals of carbon atomic chains with field-emission electron microscopy. Phys. Rev. B *80*.

Mizutani, Y. (1997). Direct observation of cooling of heme upon photodissociation of carbonmonoxy myoglobin. Science *278*, 443–446.

Mulder, P. (2011). Are orbitals observable? Int. J. Philos. Chem. *17*, 24–35.

Nagy, A.M., Raicu, V., and Miller, R.J.D. (2005). Nonlinear optical studies of heme protein dynamics: Implications for proteins as hybrid states of matter. Biochim. Biophys. Acta BBA - Proteins Proteomics *1749*, 148–172.

Nango, E., Royant, A., Kubo, M., Nakane, T., Wickstrand, C., Kimura, T., Tanaka, T., Tono, K., Song, C., Tanaka, R., et al. (2016). A three-dimensional movie of structural changes in bacteriorhodopsin. Science *354*, 1552–1557.

Nogly, P., Weinert, T., James, D., Carbajo, S., Ozerov, D., Furrer, A., Gashi, D., Borin, V., Skopintsev, P., Jaeger, K., et al. (2018). Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. Science eaat0094.

Ogilvie, J.F. (2011). Is a molecular orbital measurable by means of tomographic imaging? Found. Chem. *13*, 87–91.

Pande, K., Hutchison, C.D.M., Groenhof, G., Aquila, A., Robinson, J.S., Tenboer, J., Basu, S., Boutet, S., DePonte, D.P., Liang, M., et al. (2016). Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. Science *352*, 725–729.

Patera, L.L., Queck, F., Scheuerer, P., and Repp, J. (2019). Mapping orbital changes upon electron transfer with tunnelling microscopy on insulators. Nature *566*, 245–248.

Puschnig, P., Berkebile, S., Fleming, A.J., Koller, G., Emtsev, K., Seyller, T., Riley, J.D., Ambrosch-Draxl, C., Netzer, F.P., and Ramsey, M.G. (2009). Reconstruction of molecular orbital densities from photoemission data. Science *326*, 702–706.

Ren, Z. (2013a). Reaction trajectory revealed by a joint analysis of Protein Data Bank. PLoS ONE *8*, e77141.

Ren, Z. (2013b). Reverse engineering the cooperative machinery of human hemoglobin. PLoS ONE *8*, e77363.

Ren, Z. (2016). Molecular events during translocation and proofreading extracted from 200 static structures of DNA polymerase. Nucleic Acids Res. *6*, 1–13.

Ren, Z., and Yang, X. (2016). Angular-split/temporal-delay approach to ultrafast protein dynamics at XFELs. Acta Crystallogr. Sect. Struct. Biol. *72*, 871–882.

Ren, Z., Bourgeois, D., Helliwell, J.R., Moffat, K., Srajer, V., and Stoddard, B.L. (1999). Laue crystallography: coming of age. J Synchrotron Rad *6*, 891–917.

Ren, Z., Perman, B., Srajer, V., Teng, T.-Y., Pradervand, C., Bourgeois, D., Schotte, F., Ursby, T., Kort, R., Wulff, M., et al. (2001). A molecular movie at 1.8 Å resolution displays the photocycle of photoactive yellow protein, a eubacterial blue-light receptor, from nanoseconds to seconds. Biochemistry *40*, 13788–13801.

Ren, Z., Šrajer, V., Knapp, J.E., and Royer, W.E. (2012). Cooperative macromolecular device revealed by meta-analysis of static and time-resolved structures. Proc. Natl. Acad. Sci. U. S. A. *109*, 107–112.

Ren, Z., Chan, P.W.Y., Moffat, K., Pai, E.F., Royer, W.E., Šrajer, V., and Yang, X. (2013). Resolution of structural heterogeneity in dynamic crystallography. Acta Cryst *D69*, online.

Ren, Z., Ayhan, M., Bandara, S., Bowatte, K., Kumarapperuma, I., Gunawardana, S., Shin, H., Wang, C., Zeng, X., and Yang, X. (2018). Crystal-on-crystal chips for in situ serial diffraction at room temperature. Lab. Chip 2246–2256.

Repp, J., Meyer, G., Stojković, S.M., Gourdon, A., and Joachim, C. (2005). Molecules on insulating films: Scanning-tunneling microscopy imaging of individual molecular orbitals. Phys. Rev. Lett. *94*.

Rosca, F., Kumar, A.T.N., Ionascu, D., Ye, X., Demidov, A.A., Sjodin, T., Wharton, D., Barrick, D., Sligar, S.G., Yonetani, T., et al. (2002). Investigations of anharmonic low-frequency oscillations in heme proteins. J. Phys. Chem. A *106*, 3540–3552.

Scerri, E.R. (2000). Have orbitals really been observed? J. Chem. Educ. *77*, 1492.

Schmidt, M., Rajagopal, S., Ren, Z., and Moffat, K. (2003). Application of singular value decomposition to the analysis of time-resolved macromolecular X-ray data. Biophys. J. *84*, 2112–2129.

Schotte, F., Lim, M., Jackson, T.A., Smirnov, A.V., Soman, J., Olson, J.S., Phillips, G.N., Wulff, M., and Anfinrud, P.A. (2003). Watching a protein as it functions with 150-ps time-resolved X-ray crystallography. Science *300*, 1944–1947.

Schotte, F., Cho, H.S., Kaila, V.R.I., Kamikubo, H., Dashdorj, N., Henry, E.R., Graber, T.J., Henning, R., Wulff, M., Hummer, G., et al. (2012). Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography. Proc. Natl. Acad. Sci. *109*, 19256–19261.

Schwarz, W.H.E. (2006). Measuring Orbitals: Provocation or Reality? Angew. Chem. Int. Ed. *45*, 1508–1517.

Spence, J.C.H., Zatsepin, N.A., and Li, C. (2014). Coherent convergent-beam time-resolved X-ray diffraction. Philos. Trans. R. Soc. B Biol. Sci. *369*, 20130325–20130325.

Šrajer, V., Teng, T. -y., Ursby, T., Pradervand, C., Ren, Z., Adachi, S. -i., Schildkamp, W., Bourgeois, D., Wulff, M., and Moffat, K. (1996). Photolysis of the carbon monoxide complex of myoglobin: nanosecond time-resolved crystallography. Science *274*, 1726–1729.

Šrajer, V., Ren, Z., Teng, T.-Y., Schmidt, M., Ursby, T., Bourgeois, D., Pradervand, C., Schildkamp, W., Wulff, M., and Moffat, K. (2001). Protein conformational relaxation and ligand migration in myoglobin: A nanosecond to millisecond molecular movie from time-resolved Laue X-ray diffraction. Biochemistry *40*, 13802–13815.

Stapelfeldt, H. (2004). Electrons frozen in motion. Nature *432*, 809–810.

Stodolna, A.S., Rouzée, A., Lépine, F., Cohen, S., Robicheaux, F., Gijsbertsen, A., Jungmann, J.H., Bordas, C., and Vrakking, M.J.J. (2013). Hydrogen atoms under magnification: Direct observation of the nodal structure of Stark states. Phys. Rev. Lett. *110*.

Tenboer, J., Basu, S., Zatsepin, N., Pande, K., Milathianaki, D., Frank, M., Hunter, M., Boutet, S., Williams, G.J., Koglin, J.E., et al. (2014). Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. Science *346*, 1242–1246.

Ursby, T., and Bourgeois, D. (1997). Improved estimation of structure-factor difference amplitudes from poorly accurate data. Acta Crystallogr. A *53*, 564–575.

Vozzi, C., Negro, M., Calegari, F., Sansone, G., Nisoli, M., De Silvestri, S., and Stagira, S. (2011). Generalized molecular orbital tomography. Nat. Phys. *7*, 822–826.

Waber, J.T., and Cromer, D.T. (1965). Orbital radii of atoms and ions. J. Chem. Phys. *42*, 4116–4123.

Yang, X., Ren, Z., Kuk, J., and Moffat, K. (2011). Temperature-scan cryocrystallography reveals reaction intermediates in bacteriophytochrome. Nature *479*, 428–432.

Zeng, X., Ren, Z., Wu, Q., Fan, J., Peng, P.-P., Tang, K., Zhang, R., Zhao, K.-H., and Yang, X. (2015). Dynamic crystallography reveals early signaling events in ultraviolet photoreceptor UVR8. Nat Plants *1*, 14006.

Zuo, J.M., Kim, M., O'Keeffe, M., and Spence, J.C.H. (1999). Direct observation of d-orbital holes and Cu–Cu bonding in Cu2O. Nature *401*, 49–52.

Figures and Legends

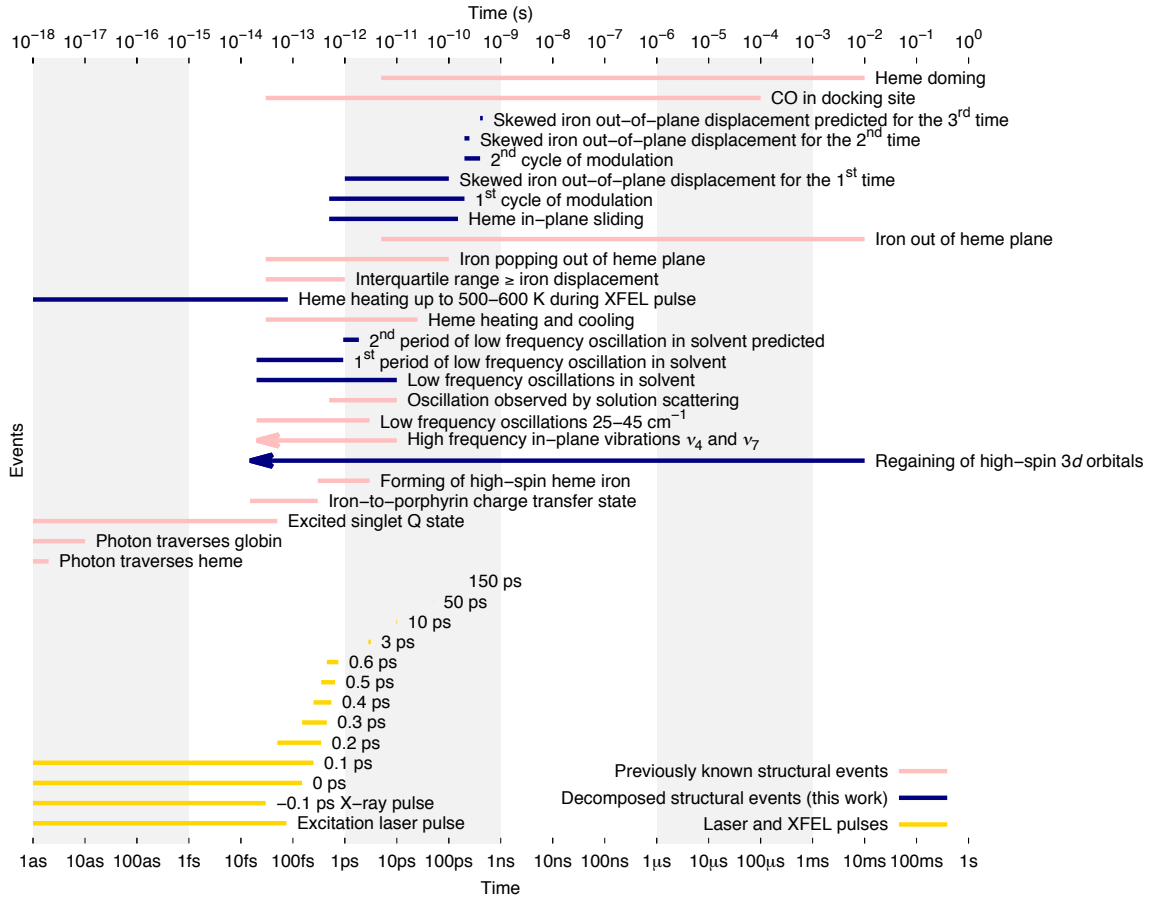


Figure 1. A Gantt chart summarizing structural events during photodissociation of MbCO. Previously known structural events are in pink. The excitation laser pulse and XFEL pulses are in yellow. The newly observed structural events from decomposition of the serial crystallographic datasets are in dark blue.

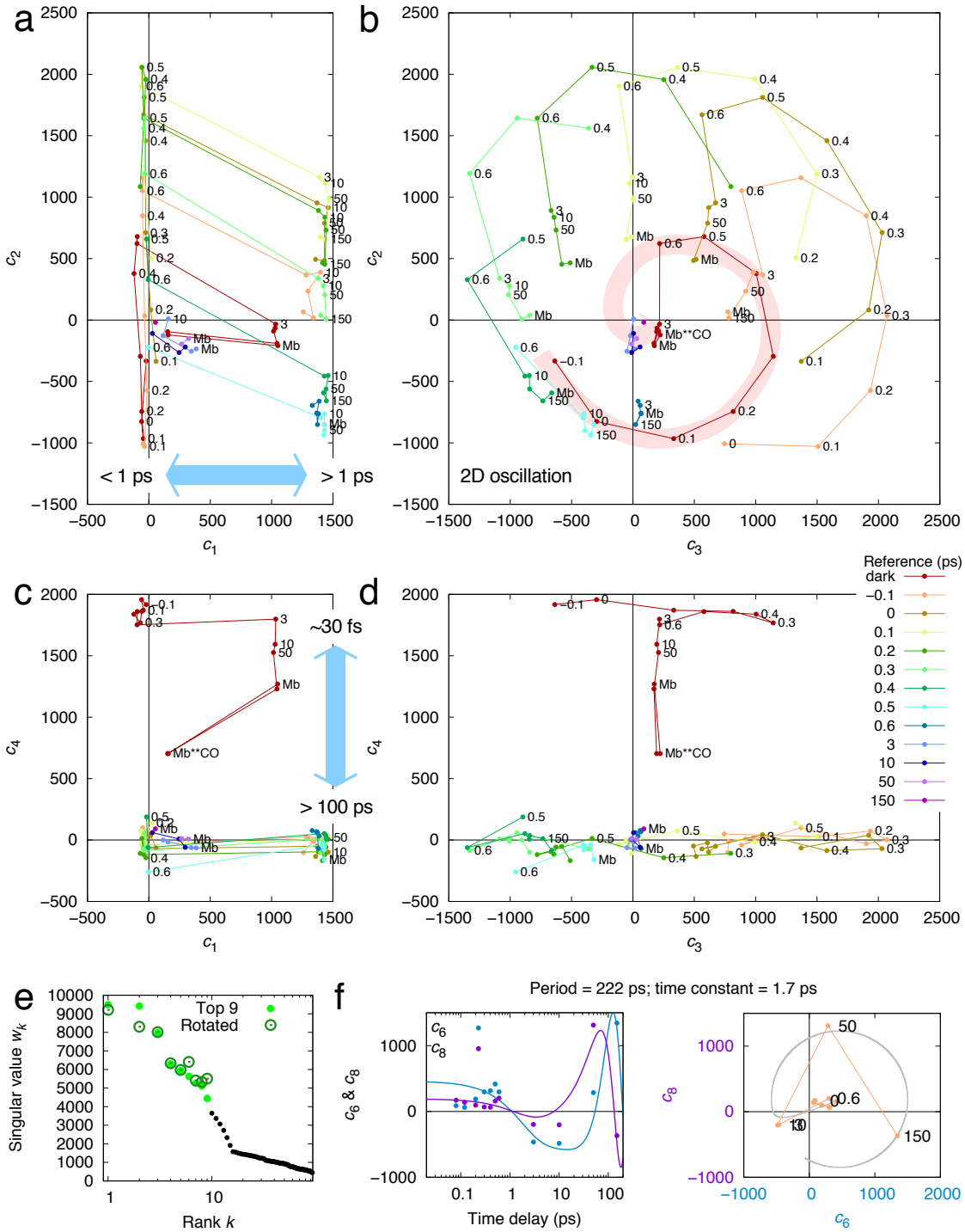


Figure 2. SVD of 93 difference maps around the heme site. These difference maps are divided into several series. Each series consists of the difference maps produced from subtracting the same reference (Methods and Table S1). Each of

the 93 difference maps can be expressed as a linear combination of many components $c_1\boldsymbol{U}_1 + c_2\boldsymbol{U}_2 + \ldots$ after the decomposition. Among a common set of components $\boldsymbol{U}_k$ ($k = 1, 2, \ldots$) shared by all difference maps, also known as left singular vectors in linear algebra, the first a few components are the most significant. Each of the difference maps requires a coefficient set c_k for the linear combination. This coefficient set varies from one map to the other, that is, a function of the delay time or other parameters associate with the datasets. Each coefficient c_k indicates how much of the corresponding component $\boldsymbol{U}_k$ exists in the linear combination. Each coefficient set is derived from a right singular vector as traditionally known in linear algebra (Methods). These coefficients produced by SVD should not be confused with the Fourier coefficients used to synthesize each map. (a-d) Orthographic projections of coefficients in some of the significant dimensions. Each difference map is represented by a dot at the location that marks the coefficients associated with the major components $\boldsymbol{U}_1$ through $\boldsymbol{U}_4$. Difference maps in a series are connected by lines in a corresponding color. The multi-dimensional space is projected to various planes and tiled together so that any two adjacent panels from (a) to (d), horizontal or vertical, can be imagined erecting a three-dimensional subspace. (e) Singular values indicating the significance of the components are sorted in a descending order of their values before a rotation. A multi-dimensional rotation alters their values. (f) Coefficients c_6 and c_8 of the $-F_{-0.1}$ series in yellow are least-squares fitted with a two-dimensional oscillation (Methods).

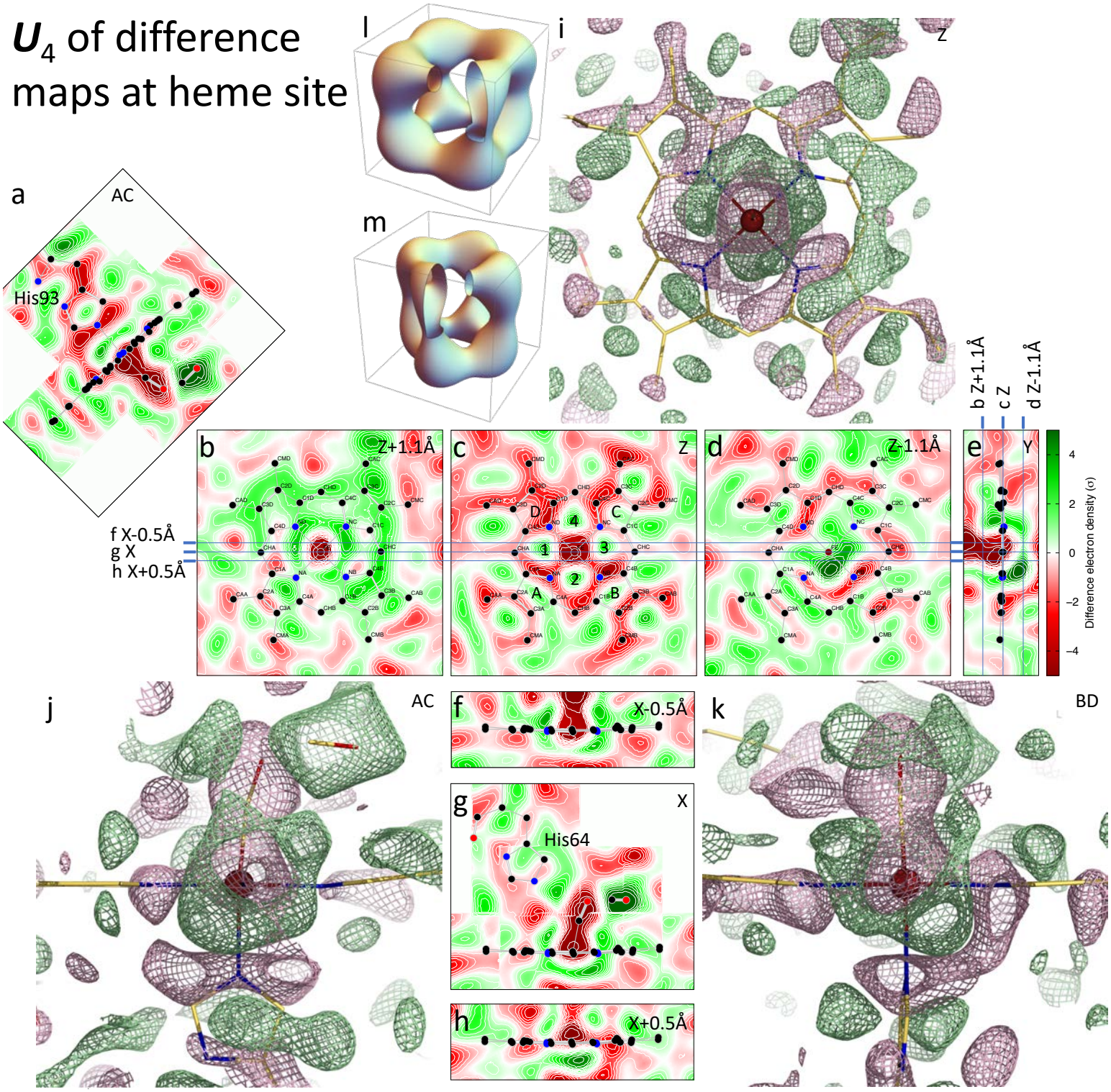


Figure 3. Ultrafast formation of high-spin 3*d* orbitals of the heme iron. The fourth component $\boldsymbol{U}_4$ of the difference maps captures a cubic shaped positive density network. Each component in the linear combination resulting from SVD,

that is, a left singular vector, can also be presented as a difference map. The fourth component $\boldsymbol{U}_4$ is rendered as a difference map in two styles, two-dimensional cross sections and three-dimensional contours. The corresponding coefficient c_4 is plotted in Fig. 2cd. (a-h) Images of two-dimensional cross sections through the component difference map $\boldsymbol{U}_4$. Cross sections defined in Fig. S1 are labeled in the top right corner of each panel. Non-centric cross sections are marked to indicate small offsets from the centric cross sections. The structure of the heme, CO ligand in both bound and docked positions, and some side chains are also projected onto the images. C, N, O, and Fe atoms are in black, blue, red, and rusty dots, respectively. The imidazole rings of the distal His64 and proximal His93 are approximately coplanar with *X* and AC cross sections, respectively. Two propionic acid side chains of AD rings extend to the left in (bcd). But they are not coplanar with these cross sections. These two-dimensional cross sections are able to reveal more details in difference Fourier maps. (i-k) Three-dimensional mesh contours superimposed over a structure are traditionally rendered to display electron density maps. Green and red meshes are positive and negative contours at $\pm 7\sigma$, respectively. (l and m) Theoretical contours of probability distributions of electrons in 3*d* orbitals of iron. The combined probability distributions of $3d_{xz}$ and $3d_{yz}$ with $3d_{xy}$ (l) or with $3d_{x^2-y^2}$ (m) are both in the shape of a cube with the largest probability at the eight corners. A 1/16 cutaway in each contour shows the cross sections. This characteristic shape of electron distribution is observed in $\boldsymbol{U}_4$ of the difference maps.

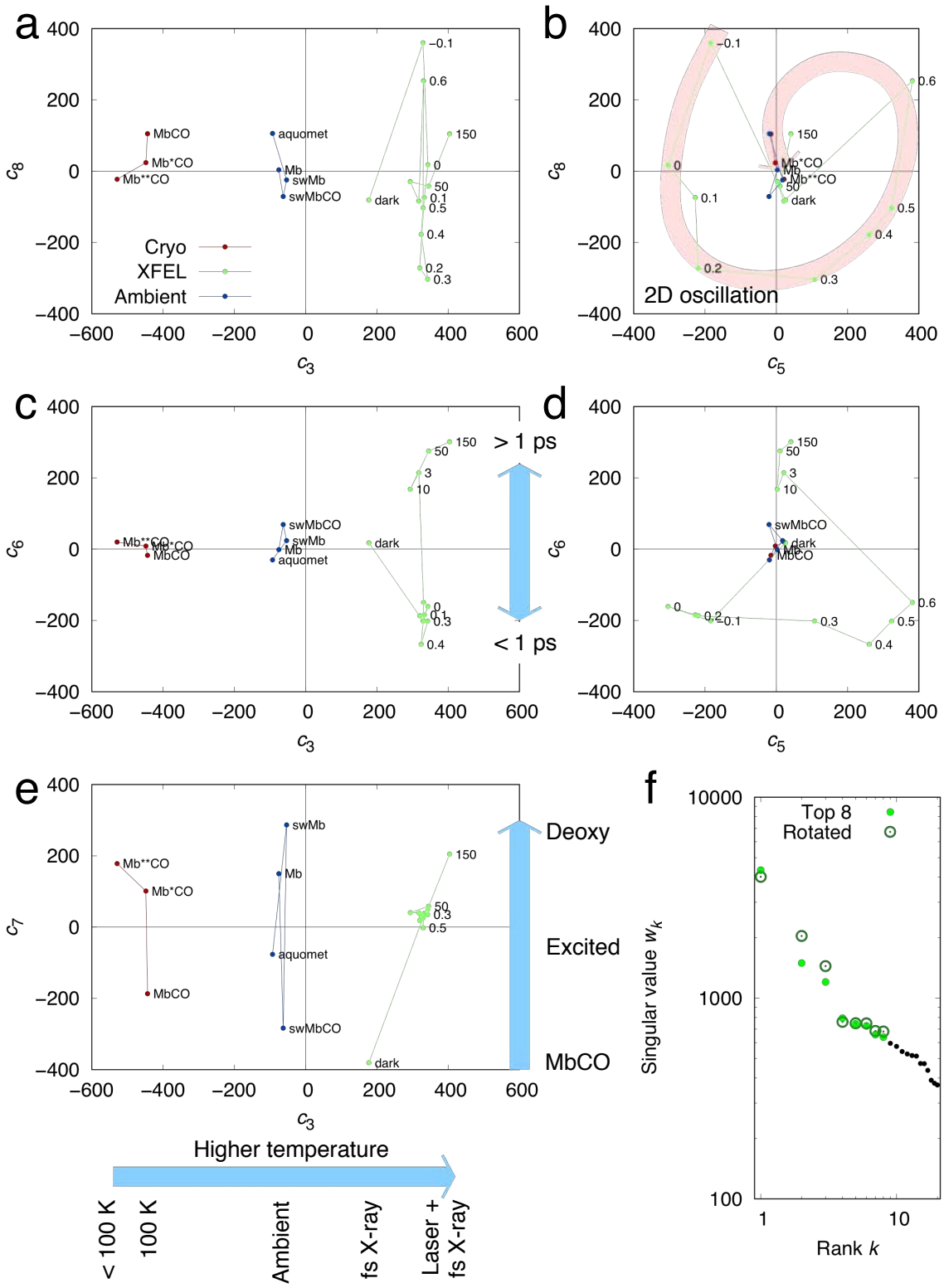


Figure 4. SVD of 20 simulated annealing omit maps (SAOMs) around the heme site. The first component is the average of all omit maps (Fig. S5). (a-e) Each

omit map is represented by a dot at the location that marks the coefficients associated with the major components. SAOMs derived from the structures at cryo and ambient temperatures are in red and blue, respectively. SAOMs from the XFEL datasets are in green. The multi-dimensional space is projected to various planes and tiled together so that any two adjacent panels from (a) to (e), horizontal or vertical, can be imagined erecting a three-dimensional subspace. (b) Coefficients c_5 and c_8 show a damped two-dimensional oscillation as marked by the curved arrow. (f) Singular values indicating the significance of the components are sorted in a descending order of their values before a multi-dimensional rotation.

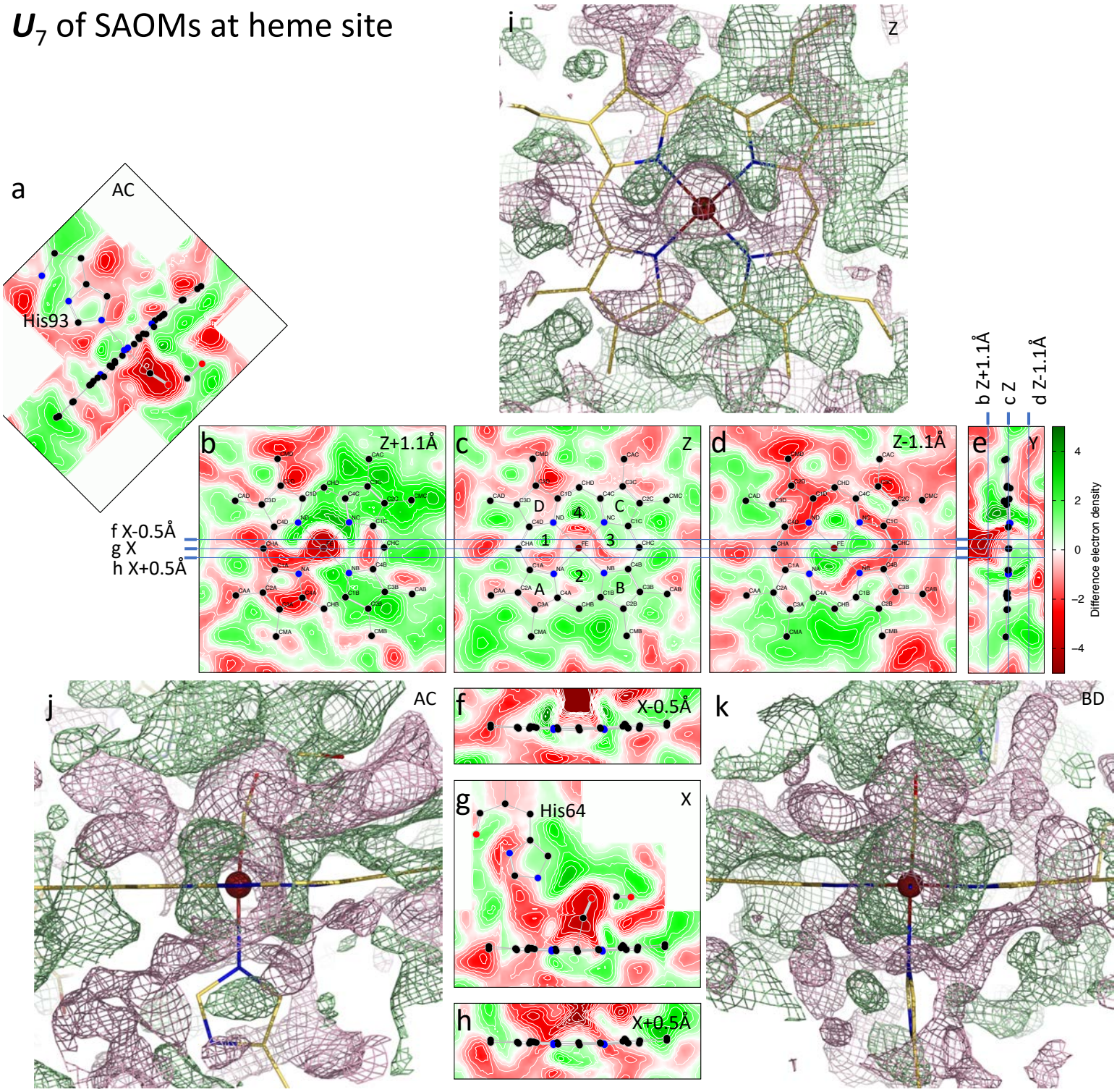
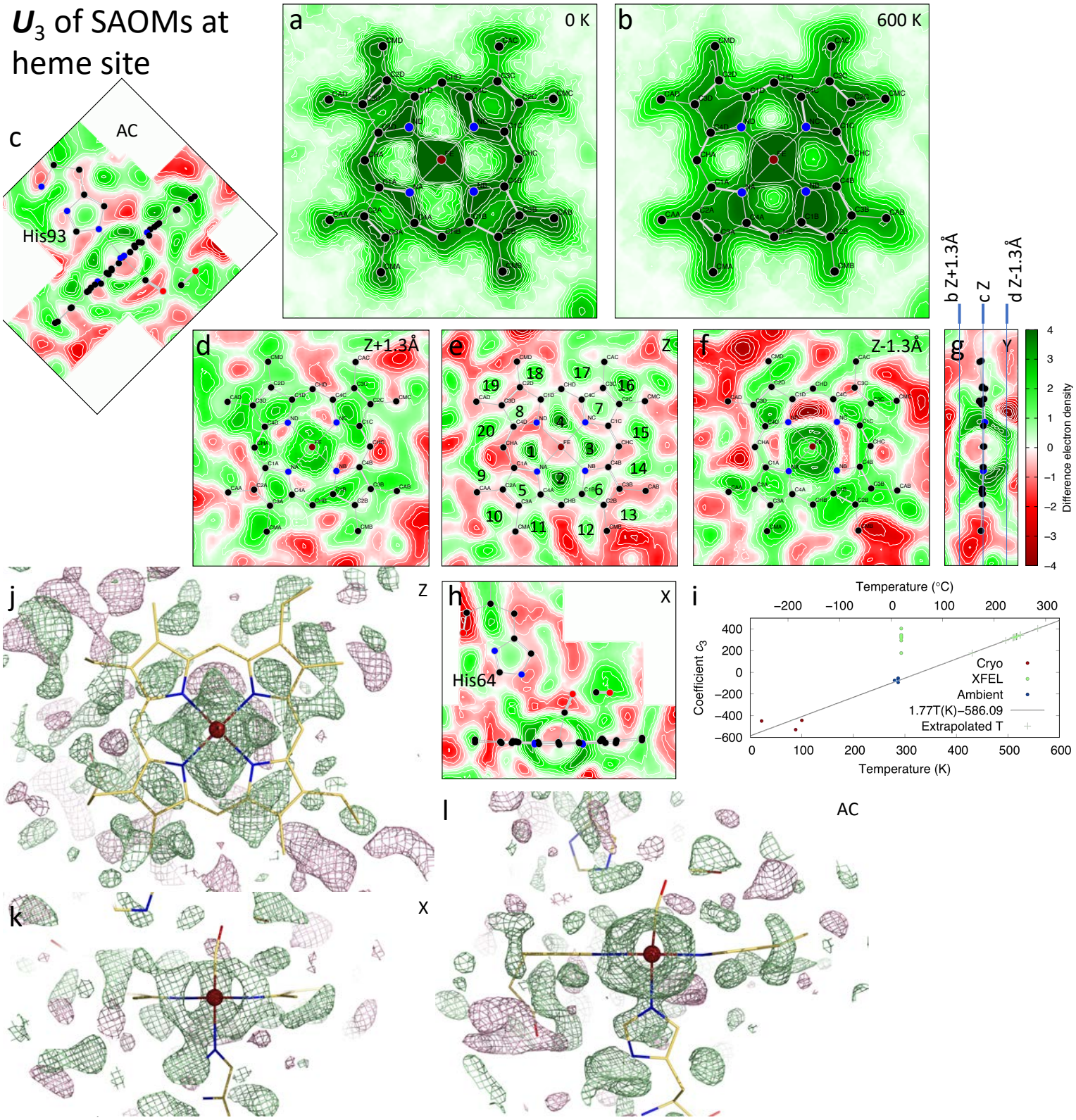


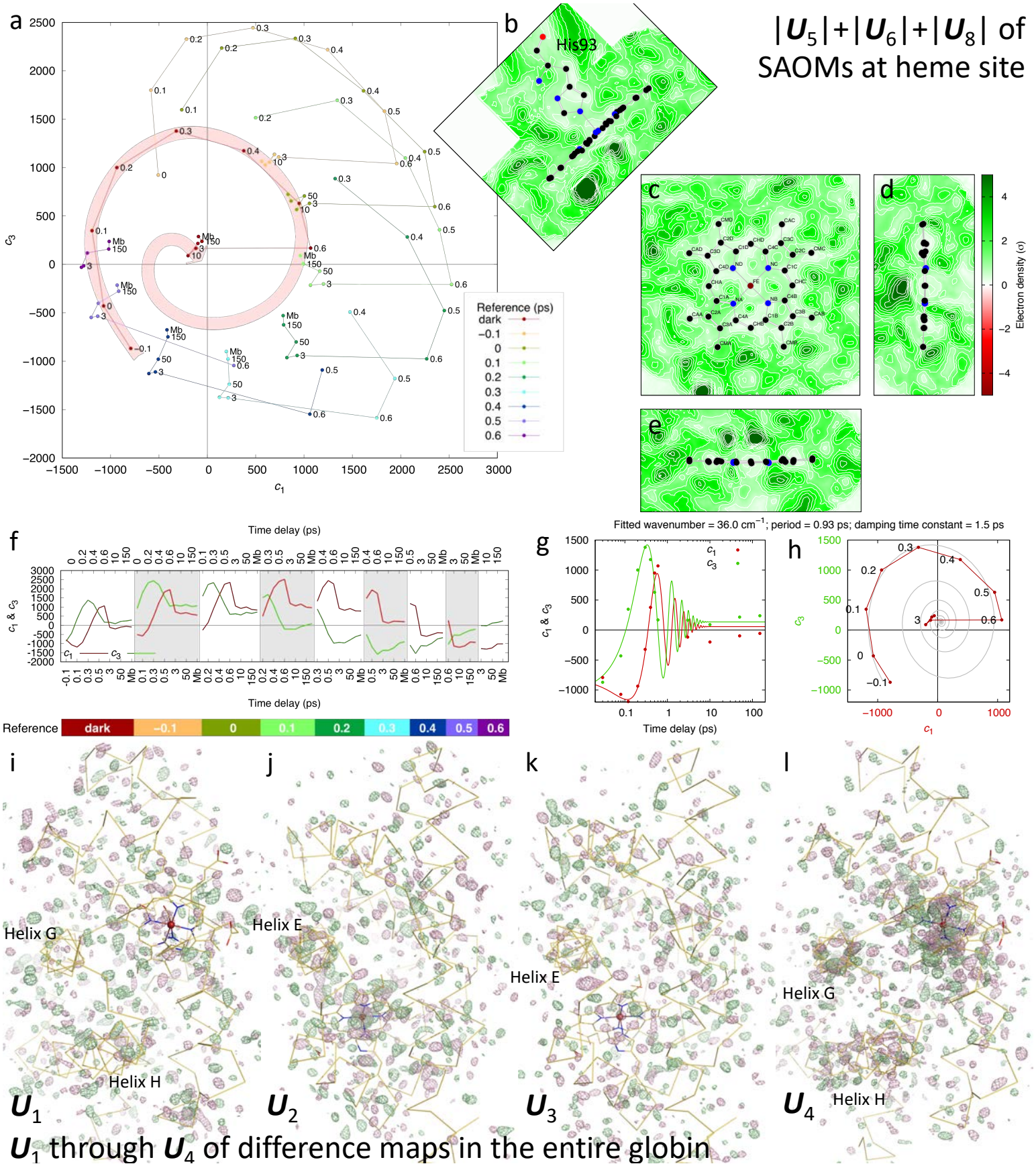
Figure 5. The seventh component of SAOMs correlated with photolysis at all temperatures. The coefficient corresponding to $\boldsymbol{U}_7$ of SAOMs increases as the reaction of ligand dissociation proceeds from MbCO to deoxy Mb in static, cryo-

trapping, and time-resolved XFEL experiments (Fig. 4e). The hypothetical transition from MbCO to deoxy Mb at 0 K is shown in Movie S3. (a-h) Cross sections through the seventh component $\boldsymbol{U}_7$. A cubic network of positive densities is marked 1-4 in (c). (i-k) Three-dimensional meshes of this component in green and red are contoured at $\pm 15\sigma$, respectively. See Fig. 3 legend for more detail.


U_3 of SAOMs at heme site
c
AC
His93
a
0 K
b
600 K
b Z+1.3Å
c Z
d Z-1.3Å
d
Z+1.3Å
e
Z
f
Z-1.3Å
g
Y
Difference electron density
j
Z
h
X
His64
i
Temperature (°C)
Coefficient c_3
Cryo
XFEL
Ambient
1.77T(K)−586.09
Extrapolated T
Temperature (K)
k
X
l
AC

1228

Figure 6. Thermal vibrations isolated from the chemical signals of photolysis. The third component of the SAOMs $\boldsymbol{U}_3$ is highly correlated with the sample temperature (Fig. 4ace). A linear extrapolation based on the datasets available at cryo and ambient temperatures predicts that the instantaneous local temperature jump of the heme may have exceeded 500 K in the XFEL datasets (i). (a) and (b) Linear combinations of $900\boldsymbol{U}_1 - 586\boldsymbol{U}_3$ and $900\boldsymbol{U}_1 + 476\boldsymbol{U}_3$ reconstitute the omit maps of the heme at 0 and 600 K, respectively, excluding all chemical signals related to photolysis, where the coefficient $c_1 = 900$ of $\boldsymbol{U}_1$ is nearly constant for all omit maps. The transition from 0 to 600 K free of the chemical signals from photolysis is shown in Movie S2. (c-h) Cross sections through the third component $\boldsymbol{U}_3$. Positive densities in 20 pillars connecting two sheets of positive densities above and below the heme in the cross sections of $Z\pm1.3$ Å are marked in (e). (j-l) Three-dimensional meshes of this component in green and red are contoured at $\pm15\sigma$, respectively. See Fig. 3 legend for more detail.



U_1 through U_4 of difference maps in the entire globin

Figure 7. Low frequency oscillations in the solvent. (a) SVD analysis of the difference maps in the entire globin produces the first and third components circularly oscillating before 1 ps and quickly diminishing during the first a few ps. Similar oscillations are also observed only around the heme site in both the difference maps (Fig. 2b) and the omit maps (Fig. 4b). (b-e) Cross sections through an averaged map of both the positive and negative halves of $\boldsymbol{U}_5$, $\boldsymbol{U}_6$, and $\boldsymbol{U}_8$ of the SAOMs. Low densities in white and light green are located consistently on the heme and protein, which suggests that the oscillating signals are more associated with the solvent rather than the globin. (f) The circularly oscillating coefficients c_1 and c_3 are plotted as function of delay time in many series. The offset sinusoidal curves indicate a circular oscillation. Each map series is derived from subtracting a different reference dataset as indicated by the color bar below. (g) and (h) c_1 and c_3 of the $-F_{\mathrm{dark}}$ series are least-squares fitted with a two-dimensional damped oscillation (Methods). (i-l) Comparison of several components of difference maps in the entire globin. $\boldsymbol{U}_1$ and $\boldsymbol{U}_3$ contain widespread signals, as they are ranked the most significant components. But no signal is concentrated on the heme or the protein. Instead, $\boldsymbol{U}_2$ and $\boldsymbol{U}_4$ clearly contain strong signals associated with the heme and secondary structures of the protein.

Ultrafast Structural Changes Decomposed from Serial Crystallographic Data

Zhong Ren

Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607, USA
Renz Research, Inc., Westmont, IL 60559, USA

zren@uic.edu

Supplementary Table

Table S1. The full matrix of difference Fourier maps

Reference (ps)	Dark	Time points (ps)												
[1]Dark (5cmv)	[2].	-0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	3	10	50	150	deoxy
-0.1 (5cn4)	[3]-	.	0	0.1	0.2	0.3	0.4	0.5	0.6	3	10	50	150	deoxy
0 (5cn5)	-	-	.	0.1	0.2	0.3	0.4	0.5	0.6	3	10	50	150	deoxy
0.1 (5cn6)	-	-	-	.	0.2	0.3	0.4	0.5	0.6	3	10	50	150	deoxy
0.2 (5cn7)	-	-	-	-	.	0.3	0.4	0.5	0.6	3	10	50	150	deoxy
0.3 (5cn8)	-	-	-	-	-	.	0.4	0.5	0.6	3	10	50	150	deoxy
0.4 (5cn9)	-	-	-	-	-	-	.	0.5	0.6	3	10	50	150	deoxy
0.5 (5cnb)	-	-	-	-	-	-	-	.	0.6	3	10	50	150	deoxy
0.6 (5cnc)	-	-	-	-	-	-	-	-	.	3	10	50	150	deoxy
3 (5cnd)	-	-	-	-	-	-	-	-	-	.	10	50	150	deoxy
10 (5cne)	-	-	-	-	-	-	-	-	-	-	.	50	150	deoxy
50 (5cnf)	-	-	-	-	-	-	-	-	-	-	-	.	150	deoxy
150 (5cng)	-	-	-	-	-	-	-	-	-	-	-	-	.	deoxy
Deoxy (5d5r)	-	-	-	-	-	-	-	-	-	-	-	-	-	.
PDB entry	5cmv	5cn4	5cn5	5cn6	5cn7	5cn8	5cn9	5cnb	5cnc	5cnd	5cne	5cnf	5cng	5d5r

[1]Traditionally, the dark dataset is chosen as the reference in calculation of difference Fourier maps.
[2]Dot (.) indicates that these are empty maps.
[3]Dash (-) indicates that these are negated maps.

Table S2. Dataset labeling

PDB entry	Label	Description
1bz6	aquomet	Aquomet Mb from sperm whale at room temperature
1bzp	swMb	Deoxy Mb from sperm whale at room temperature
1bzr	swMbCO	MbCO from sperm whale at room temperature
1dwr	MbCO	MbCO from horse heart at 100 K
1dws	Mb*CO	Photolyzed MbCO from horse heart at 22 K
1dwt	Mb**CO	Photorelaxed Mb from horse heart at 88 K
5cmv	dark	XFEL dark dataset of MbCO from horse heart
5cn4	-0.1	XFEL light datasets binned by delay time < 1 ps
5cn5	0	
5cn6	0.1	
5cn7	0.2	
5cn8	0.3	
5cn9	0.4	
5cnb	0.5	
5cnc	0.6	
5cnd	3	XFEL light datasets with delay time > 1 ps
5cne	10	
5cnf	50	
5cng	150	
5d5r	Mb	Deoxy Mb from horse heart at 7°C

Supplementary Figures and Legends

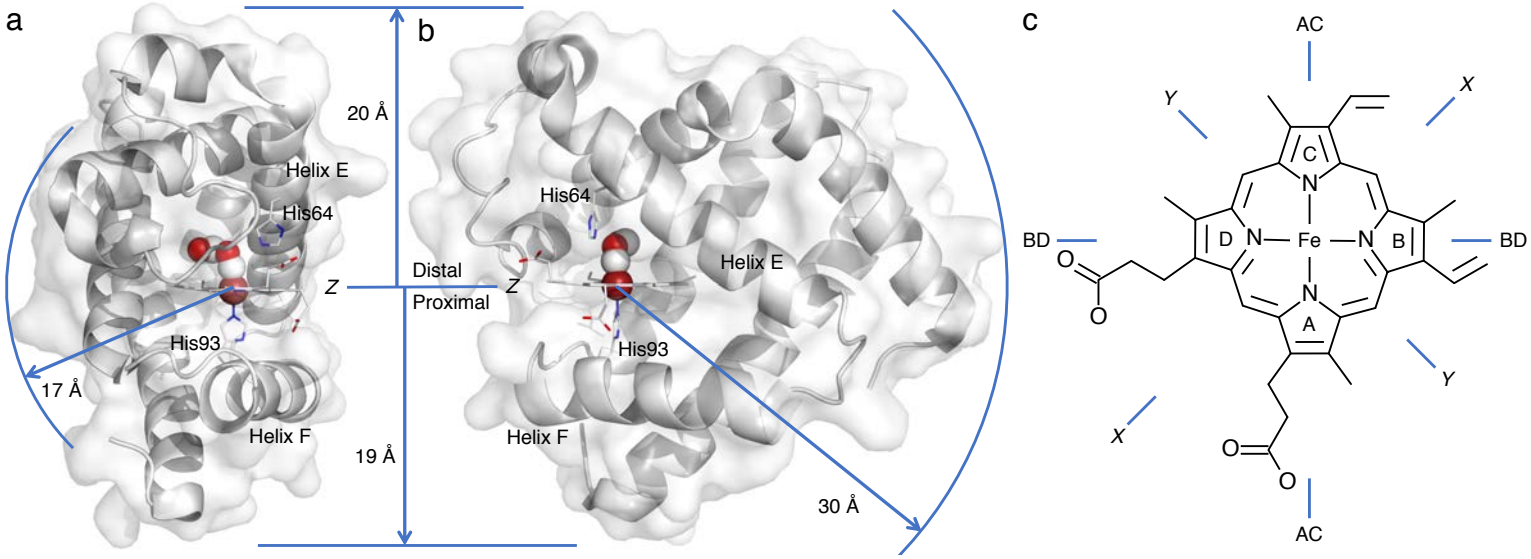


Figure S1. Orientations in myoglobin. Myoglobin (Mb) is a 17 kDa protein in muscle that carries molecular oxygen O_2. The protein structure is rendered in ribbon and surface models in (a) and (b). The non-protein prosthetic group heme in stick model is anchored to helix F of the protein through His93 also in sticks. Two propionic acid side chains of the heme are facing the molecular surface. The sixth coordination site of the heme iron in sphere could bind O_2, CO, and NO ligands also in spheres. Both a bound CO and a dissociated CO docked nearby are shown. Traditionally, the anchor side of the heme is called proximal with respect to the opposite distal side, where the ligands bind. The most remote part of the protein is 30 Å away from the heme iron. The proximal and distal directions extend 19 and 20 Å in the protein from the heme iron, respectively. The tetrapyrrole rings are traditionally named ABCD as marked in (c). In this paper, the heme plane is called *Z* cross section. Other cross sections *X*, *Y*, AC, and BD are defined in (c).

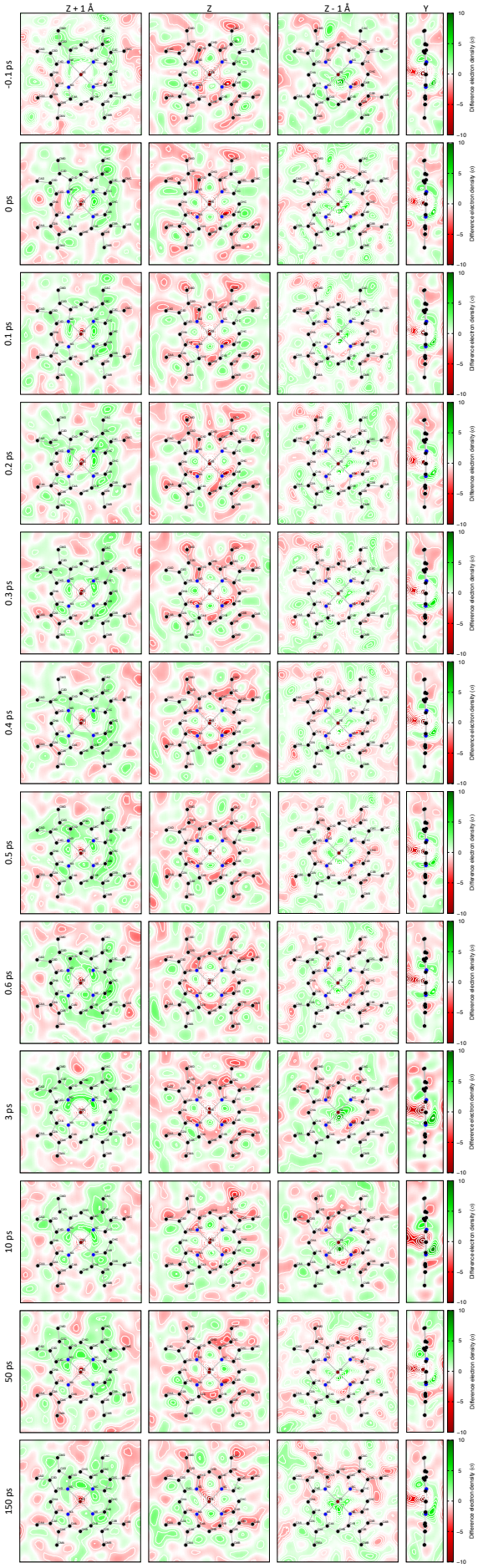

Z + 1 Å
Z
Z - 1 Å
Y
-0.1 ps
0 ps
0.1 ps
0.2 ps
0.3 ps
0.4 ps
0.5 ps
0.6 ps
3 ps
10 ps
50 ps
150 ps
Difference electron density (e)
10
5
0
-5
-10

Figure S2. The difference Fourier maps for SVD analysis. Difference maps are calculated according to the protocol described in Methods. Only the series from -0.1 to 150 ps with the dark dataset (5cmv) as the reference is displayed here. Cross sections are marked on top. See the legends of Figs. S1 and 3 for definition of cross sections. See Table S1 for the full matrix of difference maps. A visual inspection of these difference maps at different delay times may conclude that they are largely identical. Very subtle variations could be seen by a careful comparison of these maps in a time series. A numerical deconvolution is necessary to extract the inherent components.

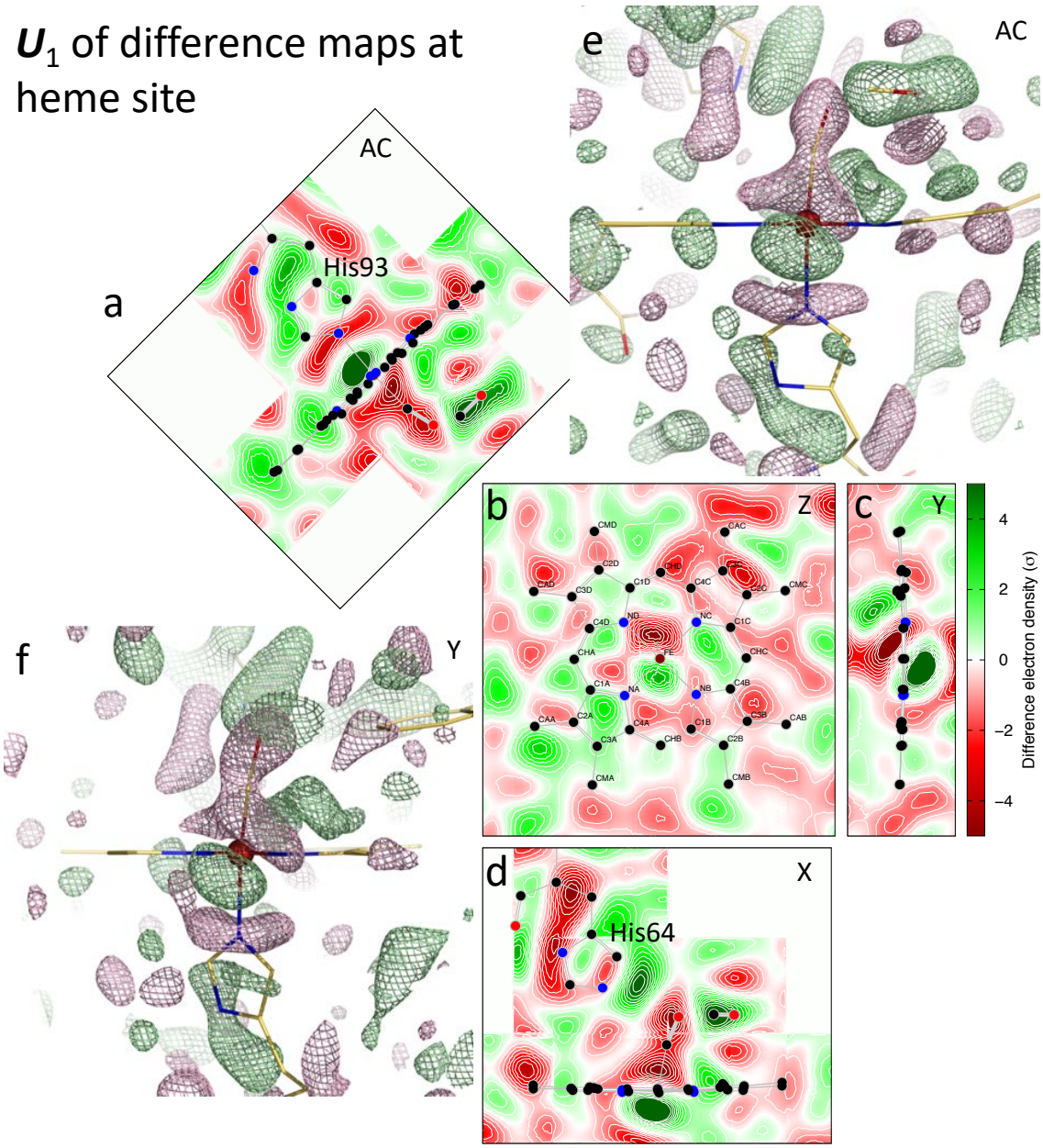


Figure S3. The most significant component $\boldsymbol{U}_1$ of the difference maps. No doubt the most significant event shown in the difference maps is the out-of-plane displacement of the iron. This component shows the severely skewed iron displacement. The corresponding coefficient c_1 is plotted in Fig. 2ac. This component clearly separates the delays > 1 ps from those < 1 ps. (a-d) Images of two-dimensional cross sections through the first component. (e) and (f) Three-dimensional mesh contours of the component difference map. Green and red meshes are positive and negative contours at $\pm 7\sigma$, respectively.

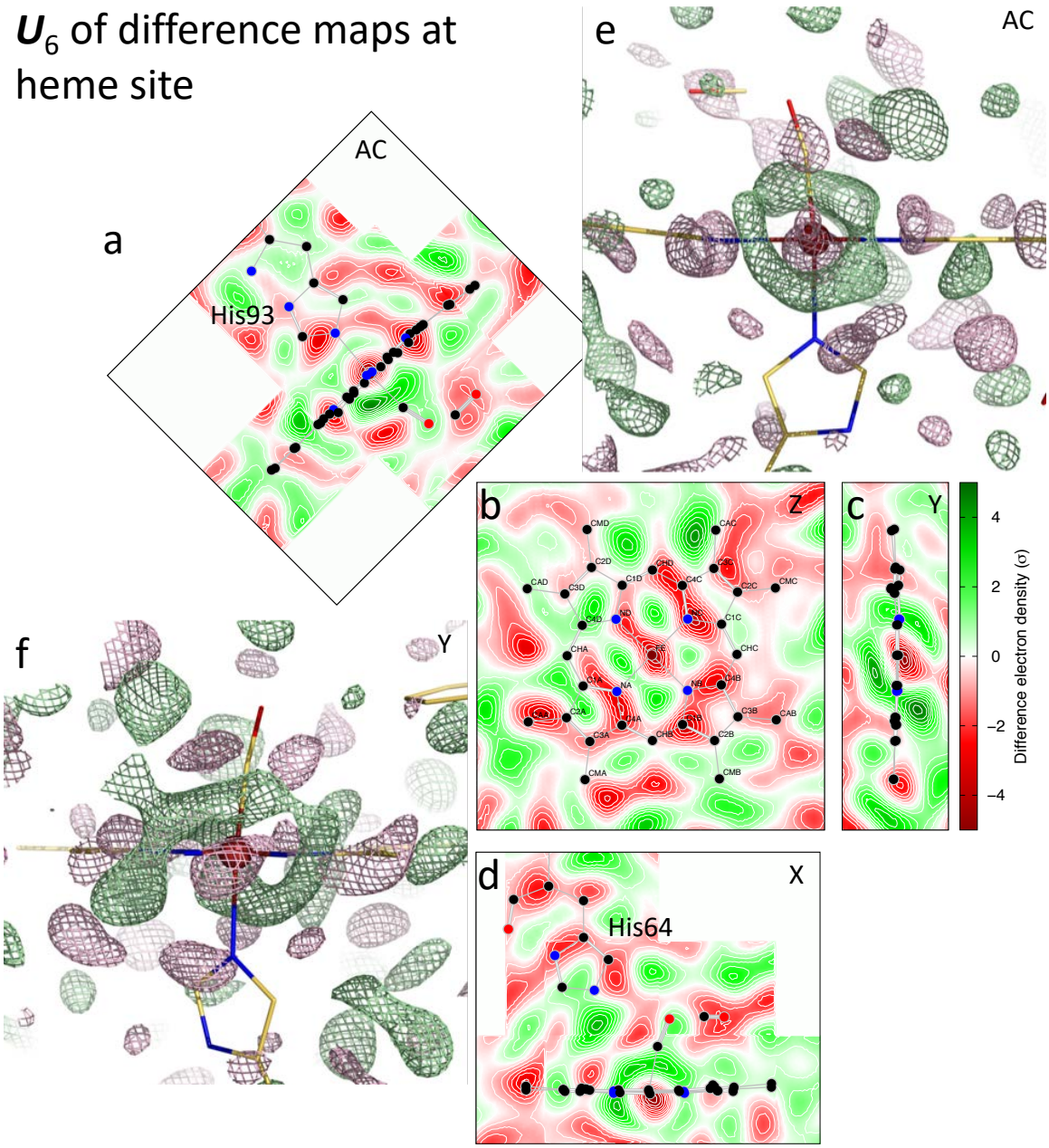


Figure S4. Modification to the cubic positive densities around the iron in the sixth component $\boldsymbol{U}_6$ of the difference maps. The corresponding coefficient c_6 is plotted in Fig. 2f. This component and the eighth component $\boldsymbol{U}_8$ form a two-dimensional oscillation that ramps up in the delays > 1 ps (Fig. 2f). The effect of modulation and oscillation is shown in Movie S1. (a-d) Images of two-dimensional cross sections through the sixth component. (e) and (f) Three-dimensional mesh contours of the component difference map. Green and red meshes are positive and negative contours at $\pm 7\sigma$, respectively.

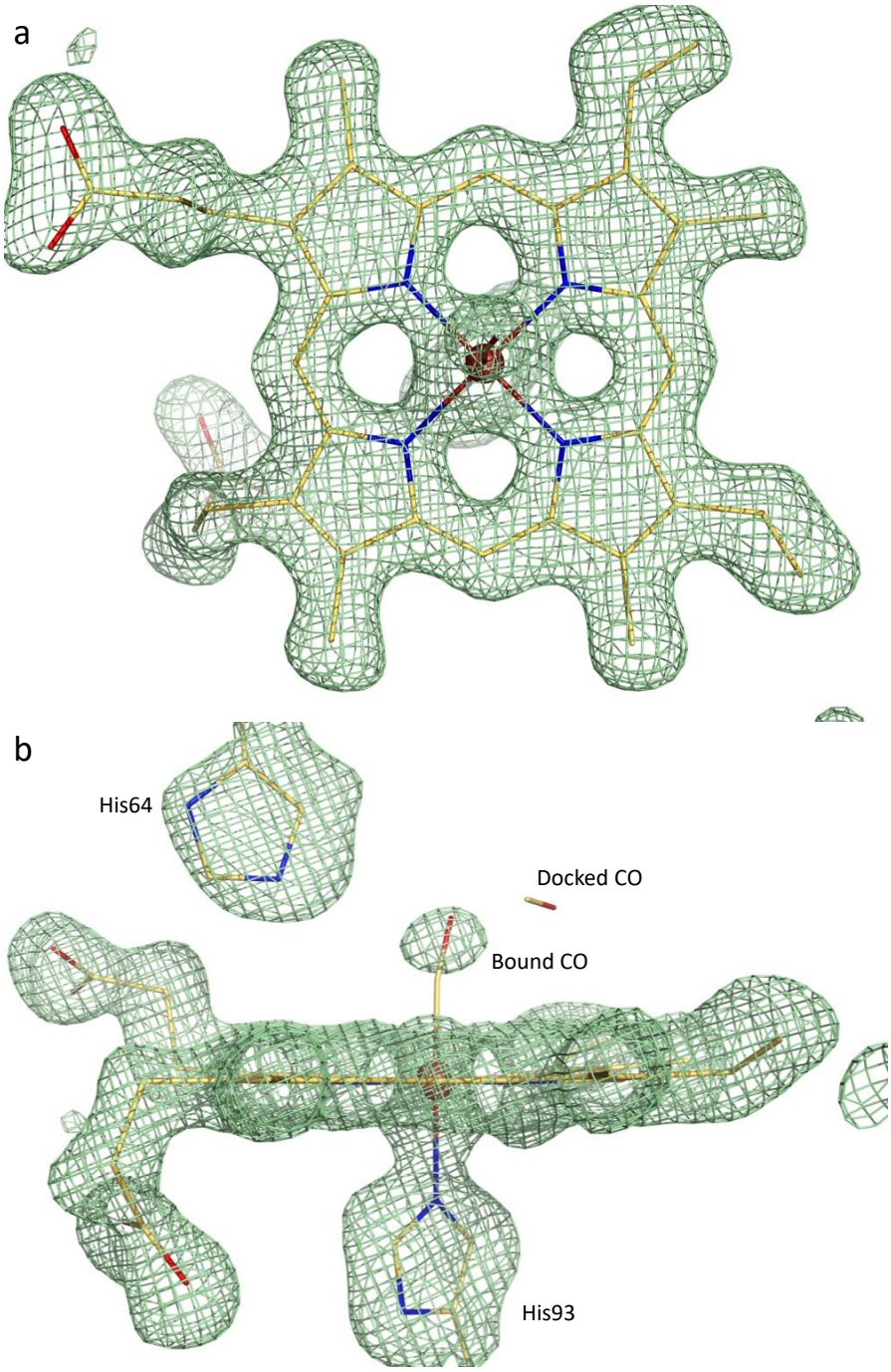
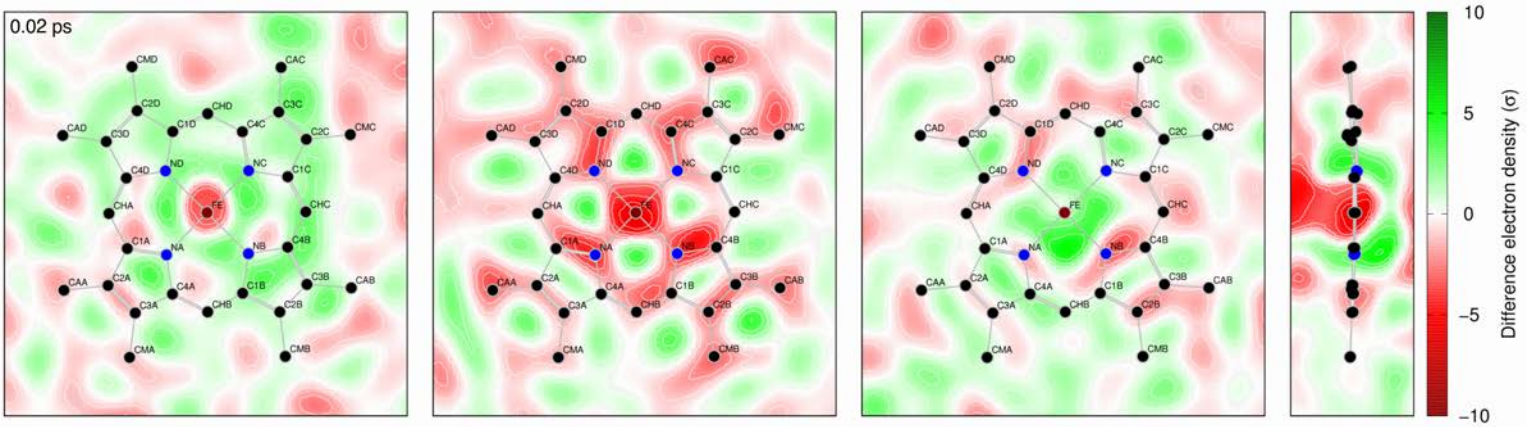
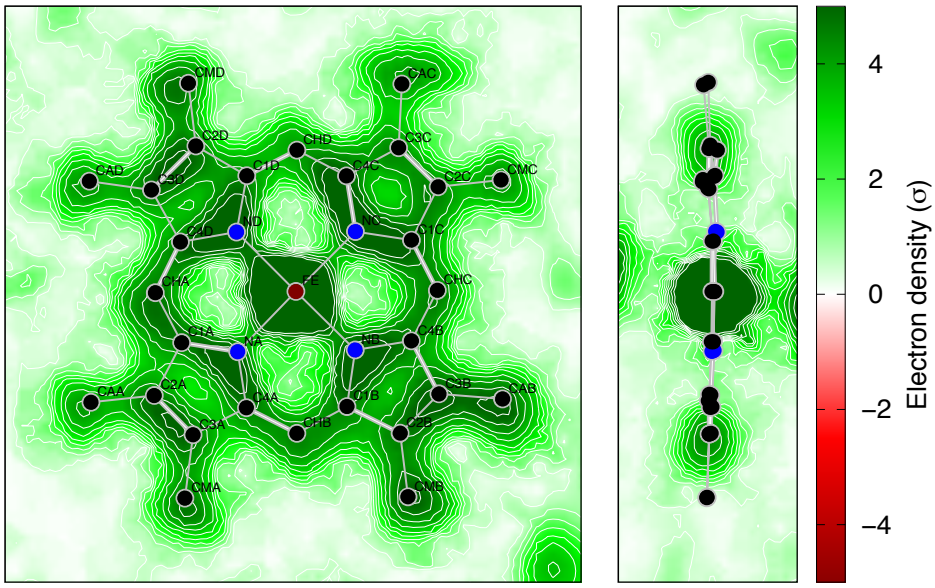


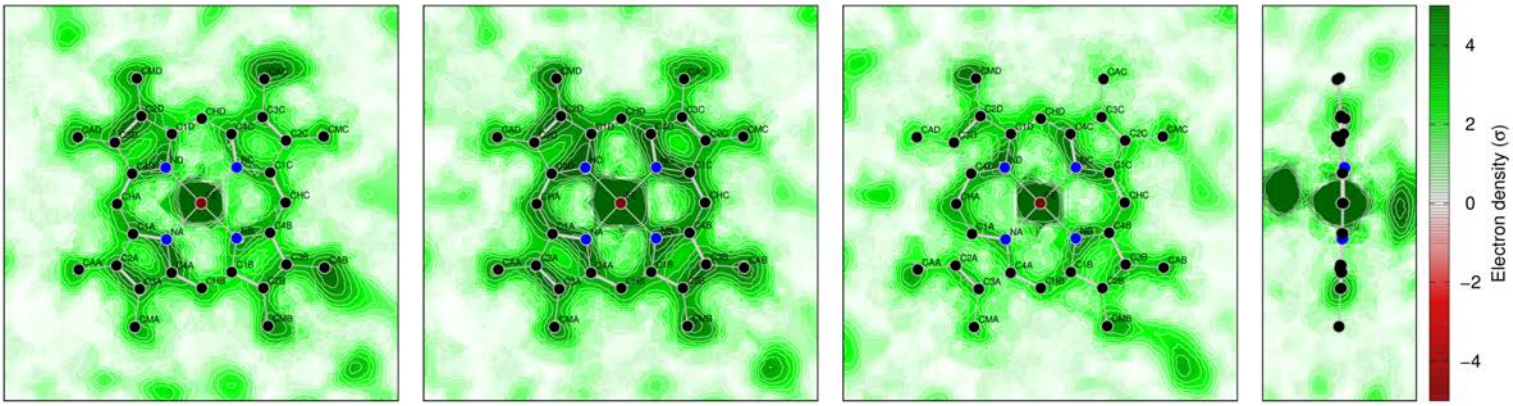
Figure S5. The most significant component of SAOMs. Since an omit map is the least biased experimental electron density map, the most significant component of the omit maps, by far (Fig. 4f), is an averaged experimental electron density map of 20 omit maps. Only the ligand is not clear due to its variation. All other components will be added onto or subtracted from this average.



Movie S1. Reconstituted difference maps showing modulation and oscillation. Linear combination of components $\boldsymbol{U}_1$, $\boldsymbol{U}_4$, $\boldsymbol{U}_6$, and $\boldsymbol{U}_8$ of the difference maps span from 0.02 to 500 ps with some extrapolation on both ends. Other major components carrying systematic difference from the deoxy Mb or featuring oscillations in the solvent are not included in the linear combination to produce this mivie. The four panels are cross sections of Z+1.1 Å, Z, Z-1.1 Å, and Y, respectively. See Figs. S1 and 3 for the definition of cross sections. The cubic positive densities first lean towards ring D from 0.5 to 3 ps. Then the position returns back to the mean position after 50 ps, sways in the opposite direction towards ring B, and ramps up after 1 ps into an oscillation of a period of 220 ps (Fig. 2f).



Movie S2. Temperature effect on electron density map of the heme. The first and third components of the SAOMs are recombined to show the temperature effect only. The components related to the chemical signals of the photolysis are not included. Therefore, this movie shows the pure thermal effect. Slight extrapolation at both ends covers from 0 to 600 K. The two panels are cross sections Z and Y, respectively. See also Fig. 6ab for electron density maps at 0 and 600 K.



Movie S3. Comparison of electron density of the heme in ligated and unligated deoxy states. The first, third, and seventh components of the SAOMs are recombined to directly contrast the electron density maps of the heme in ligated and deoxy states. The coefficients for the first and third components are constant in the reconstituted maps at 0 K, that is to say, the thermal effect is excluded. The coefficient c_7 for the seventh components is set to -400 for the ligated state and 300 for the deoxy state (Fig. 4e). The panels are cross sections of *Z*+1.1 Å, *Z*, *Z*-1.1 Å, and *Y*, respectively. The ligation state and the iron out-of-plane motion are shown in cross section *Y*. Cross section *Z*+1.1 Å on the distal side clearly shows greater densities of the heme in deoxy state, signaling the heme doming. Most importantly, the heme is sliding towards the AB rings in deoxy state. This is consistent with the skewed displacement of the iron and the motion of the proximal His93.