Application of Singular Value Decomposition to the Analysis of Time-Resolved Macromolecular X-Ray Data

Marius Schmidt,*† Sudarshan Rajagopal,† Zhong Ren,†‡§ and Keith Moffat†‡
*Physik-Department E17, Technische Universitaet Muenchen, 85747 Garching, Germany; †Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637 USA; ‡BioCARS, Argonne National Laboratory, Argonne, Illinois 60439 USA; and §Renz Research, Des Plaines, Illinois 60018 USA

ABSTRACT Singular value decomposition (SVD) is a technique commonly used in the analysis of spectroscopic data that both acts as a noise filter and reduces the dimensionality of subsequent least-squares fits. To establish the applicability of SVD to crystallographic data, we applied SVD to calculated difference Fourier maps simulating those to be obtained in a time-resolved crystallographic study of photoactive yellow protein. The atomic structures of one dark state and three intermediates were used in qualitatively different kinetic mechanisms to generate time-dependent difference maps at specific time points. Random noise of varying levels in the difference structure factor amplitudes, different extents of reaction initiation, and different numbers of time points were all employed to simulate a range of realistic experimental conditions. Our results show that SVD allows for an unbiased differentiation between signal and noise; a small subset of singular values and vectors represents the signal well, reducing the random noise in the data. Due to this, phase information of the difference structure factors can be obtained. After identifying and fitting a kinetic mechanism, the time-independent structures of the intermediates could be recovered. This demonstrates that SVD will be a powerful tool in the analysis of experimental time-resolved crystallographic data.

INTRODUCTION

Characterization of reaction intermediates is critical to understanding the pathways and mechanism by which a protein performs its biological reaction. Direct structural information on these intermediates is difficult to obtain because they are often unstable and have short lifetimes. Thus, methods used to probe them must either be capable of fast time resolution or increase the lifetime of the species to be studied. Chemical or physical trapping methods have been used to increase the lifetime (Moffat and Henderson, 1995; Schlichting and Chu, 2000), always with the caveat that the nature of the trapping might disturb the true structure of the intermediates. In time-resolved crystallography, in contrast, no perturbation of the structure of the intermediates is required (except for those perturbations associated with the crystalline state, which is also true for trapping methods). The relative simplicity of trapping methods is given up in favor of a method that is technically challenging (Ren et al., 1999).

Excellent time resolution as low as 100 ps at third-generation synchrotron x-ray sources is possible using polychromatic Laue crystallography (Bourgeois et al., 1996), which allows the visualization of extremely short-lived intermediates. A number of time-resolved Laue studies have been performed with time resolutions varying from nanoseconds to milliseconds (reviewed in Ren et al., 1999). The most detailed of these are nanosecond pump-probe time-resolved studies on the photolysis of the CO-myoglobin complex (Srajer et al., 1996, 2001) and on the photocycle of the blue-light photoreceptor known as photoactive yellow protein (PYP) (Perman et al., 1998; Ren et al., 2001). In both studies, time-dependent difference Fourier maps are generated in real space from measured structure factor amplitudes as the reaction proceeds. Interpretation of such difference Fourier maps is not trivial. It is hindered by a low signal-to-noise ratio arising from error in the difference structure factor amplitudes and in the phase of the parent structure, and from the difference Fourier approximation itself (Henderson and Moffat, 1971). Signal may be difficult to differentiate from noise by simple visual inspection of the map (Moffat, 2001; Srajer et al., 2001). Furthermore, a difference map that corresponds to a single time point will consist of an admixture of difference features arising from all time-independent, intermediate structures that are significantly populated at that time. ''Deconvolution'', or separation of this mixture into pure, time-independent intermediates, is essential to determine the chemical reaction mechanism and the structure of each intermediate (Moffat, 1989).

Both issues, the differentiation of signal from noise and the separation of intermediates, may be addressed by a mathematical procedure commonly used in the analysis of time-resolved data, singular value decomposition (SVD) (Golub and Reinsch, 1970). SVD takes data—e.g., a set of optical absorption spectra or electron density obtained under different conditions, such as time, pH, or voltage—and represents it by two sets of vectors, which are weighted by their corresponding singular values. In time-resolved spectroscopy, for example, the ''left'' set of singular vectors (lSVs) constitute a time-independent orthonormal basis set from which all time-dependent difference spectra in the data matrix are constructed. The ''right'' singular vectors (rSVs)

Submitted June 10, 2002, and accepted for publication November 4, 2002.

Address reprint requests to Marius Schmidt, E-mail: marius@hexa.e17.physik.tu-muenchen.de.

0006-3495/03/03/2112/18 $2.00

describe the time-dependent variations of the corresponding lSVs. The singular values correspond to the degree to which their respective lSVs and rSVs contribute, in a least-squares sense, to the data matrix. Because the vectors that model the data matrix are weighted by singular values, the data matrix can be approximated by a subset of singular values and vectors that contains primarily signal, thus reducing the noise present in the data. This procedure acts as a mechanism-independent filter of noise that is objective (up to the point at which the particular subset of significant singular values and vectors is chosen). The reduced representation of the data facilitates the interpretation of the rSVs with a chemical kinetic mechanism by means of a least-squares fit. This then allows the condition-independent (here, time-independent) intermediates to be obtained. The only requirement for the application of SVD is that the observable varies linearly with the concentration of the intermediates, which is the case with difference spectra and electron density.

SVD has been successfully used in a number of areas such as the analysis of spectroscopic data (Henry, 1997; Henry and Hofrichter, 1992, and references therein), of molecular dynamics simulations (Romo et al., 1995; Doruker et al., 2000) and the temporal variation of genome-wide expression (Alter et al., 2000). However, time-resolved crystallographic data differs substantially from, for example, time-resolved spectroscopic data. The signal-to-noise is much lower at a typical grid point in a difference electron density map than at a typical wavelength in a difference absorption spectrum; the signal tends to be concentrated at a subset of grid points, usually near the active site or chromophore of the protein, rather than distributing over a wide wavelength range, with the consequence that the majority of grid points exhibit only noise; and the difference electron density calculated from thousands of reflections is distributed over 10^5–10^6 grid points, rather than at 10^2 wavelengths. Further, time-resolved crystallographic data exhibit systematic as well as random errors, arising, for example, from the difference Fourier approximation (Henderson and Moffat, 1971) and from variation in the extent of reaction initiation from time point to time point.

It was therefore not obvious that SVD would be immediately applicable to time-resolved crystallographic data, nor how it should be implemented, and what the limitations on its successful use might be. We evaluate these issues here by preparing mock data that simulate a time-resolved crystallographic experiment on PYP. These data were generated for several chemical kinetic mechanisms of varying complexity in which the level of noise on the structure amplitudes, the number of measured time points, and the extent of reaction initiation were varied to simulate a realistic time-resolved experiment as closely as possible. Analysis of these mock data by SVD reveals that the method is powerful and can be successfully applied, within certain limits. These limits in turn suggest which types of noise are most important and how experiments should be conducted, if SVD is to be subsequently applied. The noise-reduced data can be further analyzed to identify the chemical kinetic mechanism and determine the time-independent structures of the intermediates in that mechanism.

A GENERAL GUIDE TO THE SVD ANALYSIS

The following paragraphs explore the applicability of SVD to time-resolved x-ray data by using mock data. In Section 1, Application of of SVD to time-dependent difference electron density, we present the principles. This paragraph mainly describes how the time-dependent difference electron density maps can be related to the columns of a single large matrix, which can be decomposed by SVD. In Section 2, Generation of the mock data, the generation of realistic mock data is outlined in detail. Special attention is given to the incorporation of noise of different origins. Section 3, Application of SVD to mock data, details the outcome of the application of SVD to the mock data. It shows the influence of noise on the magnitude and shape of the singular values and vectors, respectively. In Section 4, SVD as a noise filter, we demonstrate how SVD can contribute to the phasing of, and to a substantial reduction of noise in, the crystallographic difference maps. We call this method SVD flattening. Section 5, Extraction of mechanism, is the key section of this paper. Time-independent difference electron density maps are extracted by modeling or fitting the rSVs by several candidate, chemical kinetic mechanisms. Here, we pinpoint the limits of the SVD-driven analysis arising from different sources of noise. We demonstrate that an SVD analysis of experimental time-resolved crystallographic data is indeed feasible. In a last step, we elaborate a method to discriminate between compatible and incompatible candidate, chemical kinetic mechanisms. This method depends on representation of the time-independent difference maps on an absolute scale. This scale is not available in the maps extracted from the singular vectors. However, absolute scale may be restored after intermediate structures are determined. Therefore, we call this scheme ''posterior analysis''.

Fig. 1 presents a road map to analyze crystallographic difference electron density maps by means of SVD. It is intended as a reminder to a reader who is already familiar with the general concepts of this paper. The upper panel (for details read Sections 1–4) describes the largely objective mathematical decomposition of the data into singular vectors and values followed by noise reduction by means of SVD flattening. A successful fit of the right singular vectors by a sum of exponentials determines the minimum number of intermediates (Section 5). The middle panel describes the fit of the rSVs by chemical kinetic mechanisms and the assessment of the extracted time-independent maps by the criteria specified in Section 5. The scheme of ''posterior analysis'' in the lower panel discriminates further among the remaining mechanisms until the best mechanism(s) compat-

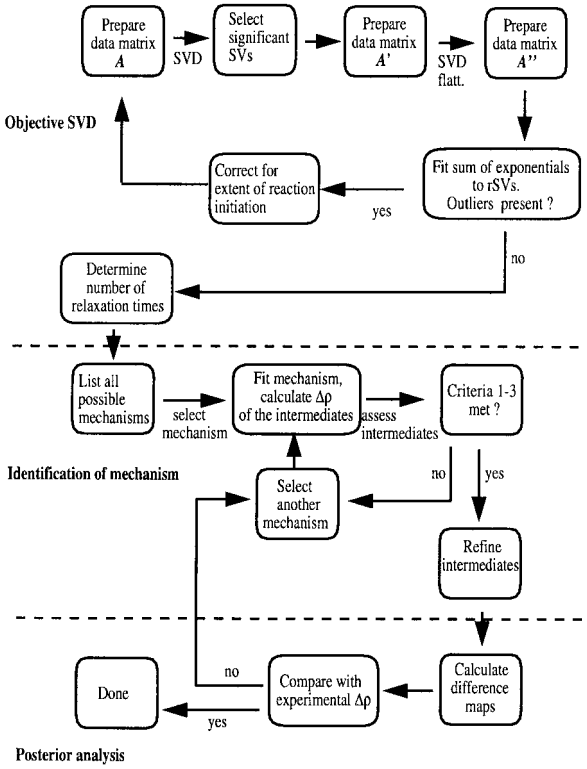

FIGURE 1 Road map for analysis of time-dependent crystallographic difference maps by SVD. (*Upper panel*) The data matrix $\boldsymbol{A}$ is decomposed by SVD and reconstituted with the most significant rSVs and lSVs. The resultant matrix $\boldsymbol{A}'$ can be SVD-flattened and fit by a sum of exponentials. If outliers are apparent, the original data matrix must be corrected; if not, determine the number of relaxation times. (*Middle panel*) List all chemical kinetic mechanism that are consistent with the number of relaxation times, fit a candidate mechanism, extract time-independent difference electron density maps, and assess the quality of such density using the three criteria mentioned in the text. If any of the maps fails a criterion, select another mechanism. (*Bottom panel*) If all criteria are passed, refine atomic structures of the intermediates. Calculate time-dependent difference maps using concentrations from the mechanism and the refined structures of the intermediates and the ground state. Compare these maps with the experimental difference maps by the mean-square deviation in the difference electron density. Choose the mechanism that exhibits the lowest deviation.

ible with the data is (are) found. We will refer to this figure throughout the paper as different steps in the analysis are reached.

APPLICATION OF SVD TO TIME-DEPENDENT DIFFERENCE ELECTRON DENSITY: PRINCIPLES

Most time-resolved experiments are of the pump-probe type (Moffat, 2002). After reaction initiation by a brief laser pulse, the structure is probed at a time delay t with an intense polychromatic x-ray beam. The results of such experiments are N data sets of time-dependent x-ray structure amplitudes. By combining these with the structure factors of the dark/unperturbed state, a time-dependent difference electron density map $\Delta\rho(t_i)$ is calculated for each time point t_i ($I = 1..N$) at M equidistant grid points in the asymmetric unit. By assigning grid point m ($m = 1..M$) of map t_i to the element a ($a = 1..M$) of vector $\mathbf{a_i}$, the data matrix $\boldsymbol{A}$ is generated. The assignment of grid point m on to element a must be consistent throughout the entire set of maps and the maps must be sorted in temporal order. Each column vector $\mathbf{a_i}$ of data matrix $\boldsymbol{A}$ contains an entire difference electron density map for time point t_i. By tracing the content of equivalent grid points in consecutive maps along the time axis, a time course of difference electron density is generated. The time courses are represented in the row vectors of the data matrix $\boldsymbol{A}$. The SVD procedure then decomposes data matrix $\boldsymbol{A}$ according to Eq. 1 into an $(M \times N)$ matrix $\boldsymbol{U}$, each of whose N columns are called lSVs; the $(N \times N)$ diagonal matrix $\boldsymbol{S}$ whose diagonal elements are the singular values; and the transpose of the square $(N \times N)$ matrix $\boldsymbol{V}$, $\boldsymbol{V}^{\mathbf{T}}$, each of whose rows is an rSV:

$$\boldsymbol{A} = \boldsymbol{U} \cdot \boldsymbol{S} \cdot \boldsymbol{V}^{\mathbf{T}}. \tag{1}$$

The data matrix $\boldsymbol{A}$ can be best approximated in a least-squares sense by matrix $\boldsymbol{A}'$ obtained by using the $S < N$ most significant columns of $\boldsymbol{U}$ (denoted $\boldsymbol{U}'$), rows of $\boldsymbol{V}^{\mathbf{T}}$ (denoted $\boldsymbol{V}'^{\mathbf{T}}$), and elements of $\boldsymbol{S}$ (denoted $\boldsymbol{S}'$) (Eq. 2) for the reconstruction:

$$\boldsymbol{U}' \cdot \boldsymbol{S}' \cdot \boldsymbol{V}'^{\mathbf{T}} = \boldsymbol{A}' \approx \boldsymbol{A}. \tag{2}$$

In this way SVD acts as an effective noise filter by separating the signal contained in the first few lSVs used for reconstruction from the noise in the remaining insignificant singular vectors, omitted from the reconstruction. The number S of significant singular values/vectors can be judged by examination of the magnitude of the singular values and the magnitude of the autocorrelation of each row of $\boldsymbol{V}^{\mathbf{T}}$. For data that contain systematic noise across the time axis, due to, for example, fluctuating laser power and extent of reaction initiation, the effectiveness of this noise filter can be enhanced by improving the autocorrelation, a measure of the function's smoothness, through the ''rotation'' algorithm described in detail by Henry and Hofrichter (1992), who also provide a comprehensive review of the SVD procedure.

GENERATION OF THE MOCK DATA

PYP is a small (14 kDa), water soluble blue-light receptor of the photoautotrophic purple bacterium *Ectothiorhodospira halophila*. It undergoes a fully reversible photocycle characterized by several spectroscopically distinct intermediates (Hoff et al., 1994; Ujj et al., 1998; Ren et al., 2001). The ultimate goal of a time-resolved crystallographic experiment on PYP is to identify the mechanism of this photocycle and to determine the structure of each intermediate. The number of spectroscopically distinct intermediates may not necessarily match the number of structurally distinct intermediates

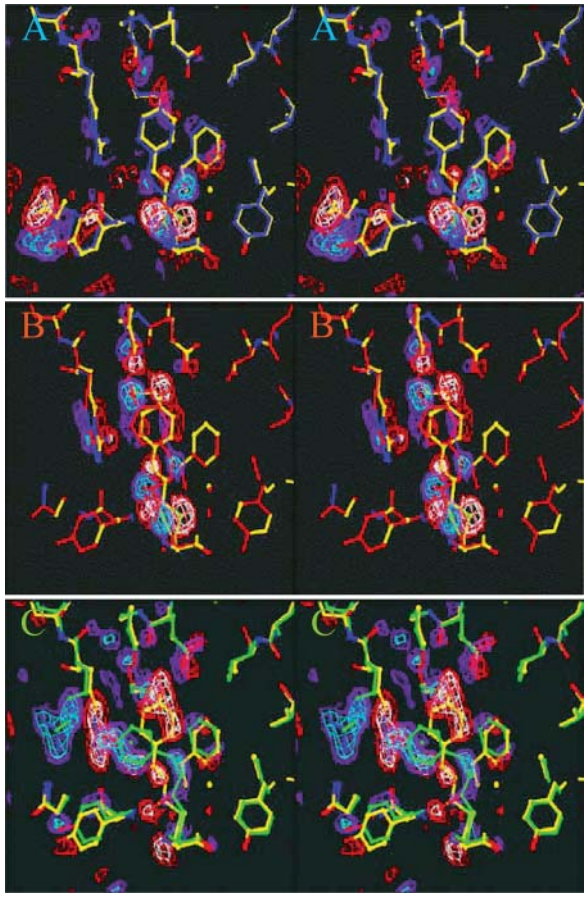

FIGURE 2 Stereo representation of the atomic structures of the mock intermediates and the corresponding reference, time-independent difference electron density maps $\Delta\rho_{Rj}$. Contour level of difference maps: red/white $-4\ \sigma/-5\ \sigma$; blue/cyan $4\ \sigma/5\ \sigma$. Atomic structure of the dark state shown in yellow in all panels. (*A*) Blue structure: intermediate I1; (*B*) red structure: intermediate I2; (*C*) green structure: intermediate I3.

(Ng et al., 1995). Nevertheless, for the generation of realistic mock data, we use a dark state, I0, and three intermediates, I1, I2, and I3 (see Fig. 2). The photocycle is completed by the recovery of the dark state. For the atomic structure of the dark state, I0, we used the Protein Data Bank (Berman et al., 2000) entry 2phy (Borgstahl et al., 1995). The first intermediate, I1, was constructed from Protein Data Bank entry 3pyp (Genick et al., 1998) by depleting this structure of all hydrogens and superimposing it on I0. The entry 2pyr (Perman et al., 1998) and a modified PYP structure from entry 2pyp (Genick et al., 1997) were used as the second and third intermediate, I2 and I3, respectively. After removing water, all structures were transferred to the room temperature unit cell of wild-type PYP, which was used throughout the simulation (see Tables 1 and 2). From the dark state structure and the intermediate structures, we calculated structure factors $\mathbf{F}_{I0}(\mathbf{h})$ and $\mathbf{F}_{I1}(\mathbf{h})..\mathbf{F}_{I3}(\mathbf{h})$ to a resolution of 1.9 Å.

We selected two mechanisms to generate the mock data from a generic chemical mechanism for interconversion of three intermediates in a branched reaction (Fig. 3 *A*). These are the irreversible sequential mechanism S (Fig. 3 *B*) and the dead-end mechanism DE with I3 in a side path (Fig. 3 *C*). For each mechanism, we chose rate coefficients to create a simple and a more complicated case, for a total of four kinetic mechanisms. Tables 3 and 4 (below) list the rate coefficients used to integrate the coupled differential equations to determine the time-dependent fractional concentrations for each of the four mechanisms, as shown in Fig. 4. In sequential mechanism S1 (Fig. 4 *A*), the time at which the peak concentration of each intermediate is reached is very well separated, whereas in S2 (Fig. 4 *B*), the rate coefficient for the decay of I2 is chosen to be larger than that of I1, such that I2 is completely buried within I1 and attains only a low maximum concentration. If noise is present, the recovery of this intermediate from the data will be a particularly challenging test for the analysis. The mechanism DE employs two irreversible steps with rate coefficients k_{+1} and k_{+4} and a reversible reaction with the rate coefficients k_{+3} and k_{-3}. This leads to the formation of two transients for I2. In mechanism DE1 (Fig. 4 *C*), the second transient for I2 starting at ~4 ms is barely visible, whereas in mechanism DE2 (Fig. 4 *D*), the second transient decays in parallel with I3. The relaxation time for the decay of the first transient of I2 is close to $1/k_{-3}$, whereas the second transient approximately parallels the decay of the equilibrium between I2 and I3, whose relaxation time is close to $1/k_{+4}$.

For each of the four mechanisms, time-dependent structure factors $\mathbf{F}(\mathbf{h},t)$ were calculated by the vector addition of time-independent structure factors of the intermediates $\mathbf{F}_{Ij}(\mathbf{h})$ and the dark-state structure factor weighted by their individual time-dependent concentrations $c_j(t)$ according to Eq. 3:

$$\mathbf{F}(\mathbf{h},t) = \left(\sum_{j=1}^{J} c_{Ij}(t)\cdot\mathbf{F}_{Ij}(\mathbf{h})\right) + \left(1-\sum_{j=1}^{J} c_{Ij}(t)\right)\cdot\mathbf{F}_{I0}(\mathbf{h}). \tag{3}$$

Note that Eq. 3 is general for any kinetic mechanism. J is the total number of intermediates and equals 3 in this instance. Due to stoichiometric constraints, the concentration of the dark state can be calculated from the concentrations of the intermediates. In typical simulations, 20% reaction initiation with $c_{I1}(0) = 0.2$ was employed, which is a realistic experimental value for PYP entering the photocycle in the crystal. However, this value can vary from crystal to crystal, and hence in some simulations the initial concentration of I1 was randomly selected from time point to time point to be between 14% and 26% or between 5% and 17%. Values of t in Eq. 3 were chosen equidistant in logarithmic time to cover the entire time course.

Noise on the structure amplitudes $|\mathbf{F}(\mathbf{h},t)|$ was based on experimental standard deviations (σ-values) measured in one particular reference Laue data set collected from a PYP

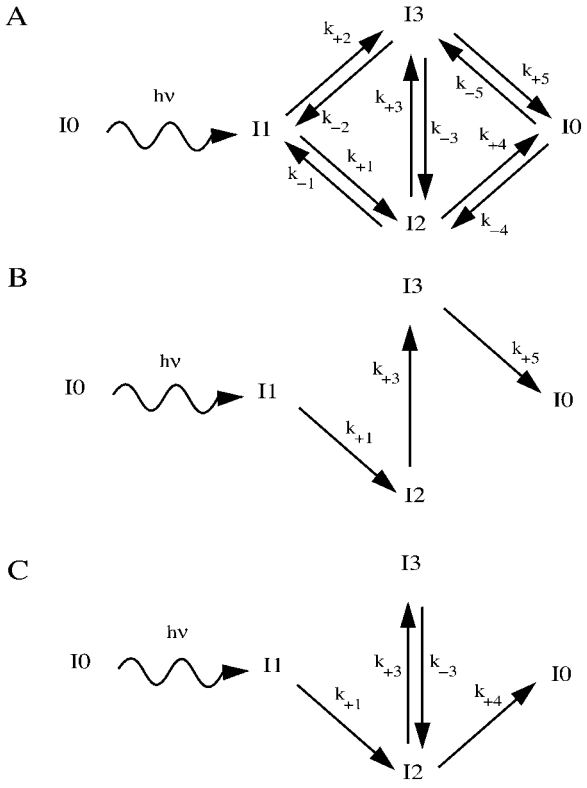

FIGURE 3 (*A*) General chemical kinetic mechanism for a branched reaction with three intermediate states I1, I2, and I3. The dark state I0 is excited by light. The first intermediate I1 relaxes back to the dark state via two other intermediates, I2 and I3. The direct path from I1 to I0 is not considered. (*B*) Irreversible sequential kinetic mechanism S; (*C*) Dead-end kinetic mechanism DE. An equilibrium between I2 and I3 is allowed as a side path or dead end; I2 relaxes back to the dark state.

crystal in the dark. To preserve the correlation between the magnitude of the structure amplitude and the magnitude of the σ-value, we defined 10 equally spaced bins between the minimum and maximum value of the structure amplitudes

TABLE 2 Difference amplitudes of the expermental and the mock PYP data sets

		Difference amplitudes				
Noise added	Experiment	None	1 s/0.5 s	2 s/1 s	5 s/3 s	10 s/5 s
Mean ΔF [el]	−0.002	0.002	0.001	−0.002	−0.005	−0.007
Mean abs(ΔF) [el]	23.3	4.8	11.2	16.6	29.4	45.1
Mean $\sigma_{\Delta F}$ [el]	5.4	–	4.7	12.2	25.4	64.9
Mean weight	0.49	–	0.50	0.48	0.49	0.50

Difference amplitudes ΔF, $\sigma_{\Delta F}$ and weight for a representative experiment at a 500 μs time delay and for mock data in the presence of different amount of noise. Values are shown for time point 14 in the middle of the reaction.

found in this data set. We then assigned the σ-values to these 10 bins. For each mock $|\mathbf{F}|$, one of the bins was chosen based on the magnitude of the structure amplitude and a σ-value was picked randomly from the set of σ-values present in this bin. A multiple of the experimental σ-value was used as the width of a Gaussian distribution of error values. An error value was picked randomly from this distribution using the Box-Muller algorithm (Box and Muller, 1958) and added to $|\mathbf{F}(\mathbf{h},t)|$. If the experimental data set did not contain an entry for a particular $\mathbf{h}$, this reflection was discarded throughout the mock data sets. If the picked error value was negative, the structure amplitude could also become negative. In this case the amplitude was reset to 0.5 of the error value. This procedure creates ''erroneous'' or ''noisy'' time-dependent data sets of amplitudes, $|\mathbf{F}^{\#}(\mathbf{h},t)|$, with standard deviation $\sigma(|\mathbf{F}^{\#}(\mathbf{h},t)|)$, with completeness identical to the experimental data sets (Tables 1 and 2). Dark-state structure amplitudes $|\mathbf{F}^{\#}{}_{\mathrm{I0}}(\mathbf{h})|$ with standard deviation $\sigma(|\mathbf{F}^{\#}{}_{\mathrm{I0}}(\mathbf{h})|)$ were derived from $|\mathbf{F}_{\mathrm{I0}}(\mathbf{h})|$ in an identical manner. In all cases, resolution was limited from 15.0 to 1.9 Å. Time-dependent difference structure amplitudes were then calculated from a ''noisy'' time-dependent data set and the ''noisy'' dark-state data set. Excitation by an intense laser pulse may generate strain in the crystal, from which streaky reflections result; and the

TABLE 1 Scaling of the experimental and the mock PYP data sets generated with kinetic mechanism S1

		R_{scale}					
Resolution [Å]	Completeness (shell) %	Experiment dark/dark	Experiment light/dark	1 s/0.5 s	2 s/1 s	5 s/3 s	10 s/5 s
15.0–3.8	98.1	0.028	0.047	0.006	0.018	0.046	0.059
3.8–3.04	98.4	0.040	0.059	0.008	0.027	0.068	0.076
3.04–2.62	97.8	0.057	0.060	0.017	0.048	0.119	0.128
2.62–2.39	97.6	0.065	0.060	0.020	0.066	0.164	0.165
2.39–2.22	97.3	0.072	0.078	0.021	0.076	0.172	0.198
2.22–2.09	96.6	0.085	0.087	0.027	0.079	0.184	0.259
2.09–1.98	94.9	0.093	0.094	0.035	0.085	0.210	0.326
1.98–1.90	91.1	0.112	0.115	0.043	0.098	0.240	0.390

Completeness and values of R_{scale} for time point 21 in presence of different amounts of noise in the structure amplitudes (see text), as a function of resolution. The experimental R_{scale} was determined by scaling dark, experimental Laue data sets collected from two different PYP crystals; and by scaling a time-dependent, experimental light Laue data set and an experimental dark data set collected from the same crystal.

$$R_{\mathrm{scale}} = \frac{\sum_{\mathrm{hkl}} \left(|\mathbf{F}(t)|^2 - |\mathbf{F}_{\mathrm{I0}}|^2 \right)}{\sum_{\mathrm{hkl}} |\mathbf{F}_{\mathrm{I0}}|^2}$$

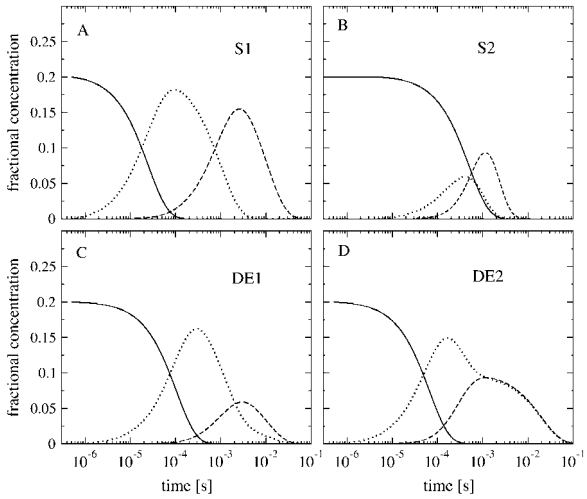

FIGURE 4 Time-dependent concentrations of intermediates calculated from the four different chemical kinetic mechanisms used in the simulations. Reaction coefficients are those given in Tables 3 and 4. Solid line: I1; dotted line: I2; dashed line: I3. (*A*) Irreversible sequential mechanism S1 in which peak concentrations of intermediates are well separated in time. (*B*) Irreversible sequential mechanism S2 in which concentrations of I2 are low and remain below those of I1 at almost all times. (*C*) Mechanism DE1 in which the second intermediate is in a side path and the second transient of I2 is negligible. (*D*) Mechanism DE2 in which there is a pronounced second transient of I2, and the concentrations of I3 and I2 are comparable at longer times.

integrated intensities of reflections having a lower peak photon count are more difficult to measure accurately. Therefore, we assigned higher noise values to the time-dependent amplitudes and lower values to the dark structure amplitudes. Each data set of difference structure amplitudes is labeled according to the amount of noise present in the time-dependent amplitudes and the dark-state amplitudes, respectively. For example, a 2 s/1 s data set corresponds to error values from Gaussians with width 2 σ for $|\mathbf{F}^{\#}(\mathbf{h},t)|$ and 1 σ for $|\mathbf{F}_{\mathrm{I0}}{}^{\#}(\mathbf{h})|$. The crystallographic statistics for the experimental reference Laue data sets and the simulated data sets with various levels of noise are shown in Tables 1 and 2.

Each time-dependent data set and the dark data set were scaled together in 20 resolution bins using XMerge (McRee, 1999). Weighted difference amplitudes $|\Delta\mathbf{F}^{\mathrm{w}}(\mathbf{h},t)|$ were calculated from the ''noisy'' differences $|\Delta\mathbf{F}^{\#}(\mathbf{h},t)|$ according to Eq (4):

$$\begin{aligned}|\Delta\mathbf{F}^{\mathrm{w}}(\mathbf{h},t)| &= w(\mathbf{h},t)\,(|\mathbf{F}^{\#}(\mathbf{h},t)| - |\mathbf{F}^{\#}_{\mathrm{I0}}(\mathbf{h})|)\\ &= w(\mathbf{h},t)\cdot|\Delta\mathbf{F}^{\#}(\mathbf{h},t)|,\end{aligned} \tag{4}$$

in which the weighting factor $w(\mathbf{h},t)$ was calculated according to Ren et al. (2001):

$$w(\mathbf{h},t) = \frac{1}{1+\dfrac{|\Delta\mathbf{F}^{\#}(\mathbf{h},t)|^2}{\langle|\Delta\mathbf{F}^{\#}(\mathbf{h},t)|^2\rangle}+\dfrac{\sigma_{\Delta\mathrm{F}}(\mathbf{h},t)^2}{\langle\sigma_{\Delta\mathrm{F}}(\mathbf{h},t)^2\rangle}}, \tag{5}$$

$\sigma_{\Delta\mathrm{F}}(\mathbf{h},t)$ is the standard deviation of the difference amplitude and was calculated as:

$$\sigma_{\Delta\mathrm{F}}(\mathbf{h},t)^2 = \sigma^2(|\mathbf{F}^{\#}(\mathbf{h},t)|) + \sigma^2(|\mathbf{F}^{\#}_{\mathrm{I0}}(\mathbf{h})|). \tag{6}$$

For each time point, time-dependent difference electron density maps $\Delta\rho(t)$ were calculated using $|\Delta\mathbf{F}^{\mathrm{w}}(\mathbf{h},t)|$ and phases calculated from the dark-state PYP atomic structure I0 on a 0.75 Å × 0.75 Å × 0.65 Å grid to a resolution of 1.9 Å. Pure, time-independent, noise- and error-free intermediate difference maps, $\Delta\rho_{\mathrm{Rj}}$ (j = 1..3), were also calculated as a reference (see also Fig. 2) by using vector subtraction of structure factors of the dark structure $\mathbf{F}_{\mathrm{I0}}(\mathbf{h})$ from those of the three intermediates $\mathbf{F}_{\mathrm{Ij}}(\mathbf{h})$.

APPLICATION OF SVD TO MOCK DATA

For each time-dependent difference map, one entire PYP molecule plus a margin of 1 Å was masked out. The difference electron density at the grid points within this mask at all time points constitutes the data matrix ***A*** and was subjected to SVD (Fig. 1, *Objective SVD*). Typically, the mask contains ~86,000 grid points, but larger masks containing 200,000 grid points could also be successfully used. To execute the necessary steps for an SVD of our time-dependent difference maps, we developed a program, Singular Value Decomposition for Time-Resolved Crystallography, written in Fortran 77. A data matrix having 21 column vectors with 86,000 grid points each can be generated from difference maps and decomposed in ~0.5 min. As an example, Fig. 5 shows the first six lSVs obtained from SVD of the 2 s/1 s data of the mechanism S1 with three intermediate states. Signal should be present in the three most significant lSVs only. In panels lSV1–lSV3, the signal is very well defined and clusters where the intermediate atomic structures differ from the dark state. However, due to the noise, some signal has spread to the fourth lSV: there is some difference density on the tail of the chromophore and at other locations, such as on the phenolate oxygen of the chromophore ring. lSV5 and lSV6 contain only noise; the electron density features are scattered randomly in space and do not cluster. As an additional test, difference map SV5-21 was reconstructed from the 17 least significant singular vectors. Here, too, difference electron density features are scattered and do not occupy chemically sensible positions. In contrast, if the difference electron density maps were reconstituted from the first four singular vectors, strong spatially contiguous signal can be observed. SV1-4 in Fig. 5 shows the resultant difference map for time point 15 of the time course. For comparison, the corresponding mock difference map TP15 is also shown. The agreement between SV1-4 and TP15 is excellent.

Fig. 6 shows how the singular values and the rSVs behave as noise is progressively added. On the left side, the magnitude of the singular values (*circles*) is shown together

with the autocorrelation of rSVs (*squares*). On the right side the first few rSVs are plotted. Panels *A* and *B* show results from noise-free data. Only three significant singular values can be detected unambiguously, as expected. The same conclusion holds for the amplitudes of the rSV (Fig. 6 *B*). The high autocorrelation of rSV4 to rSV7 shows that they still contain smoothly varying amplitudes. However, the amplitudes of both the rSVs and lSVs are negligible compared to the true signal. At the 1 s/0.5 s noise level (Fig. 6, *C* and *D*), the system is still well-behaved. No significant amplitudes except those of the first three rSVs can be observed. When the noise is further increased to the 2 s/1 s level (Fig. 6, *E* and *F*), some signal can be observed in an additional, fourth lSV, and a fourth rSV begins to rise. When the noise is further increased, there might be a significant fifth singular value (Fig. 6 *G*). However, inspection of the lSVs after application of the rotation method confirms that only four singular values and vectors are needed in the reconstitution. At high noise, the autocorrelation becomes a sensible additional criterion to select the number of significant basis vectors. At the highest noise level of 10 s/5 s (Fig. 6, *I* and *J*), the fourth lSV contributes even more to the signal and a fifth significant lSV emerges. However, reconstruction of maps with more than five basis sets is still unnecessary; the map reconstructed with all the least significant vector sets, SV6-21, does not show any significant signal (data not shown).

At noise levels at and higher than 5 s/3 s, the identification of significant singular values is hindered by an offset. The ratio of the magnitude of the significant singular values to this offset becomes increasingly unfavorable. Fig. 6 clearly shows this offset at the 5 s/3 s and 10 s/5 s noise levels (Fig. 6, *G* and *I*): the rSVs begin to deviate strongly from zero at the end of the reaction. Less significant rSVs become larger in amplitude and vary rapidly in time; however, the significant rSVs still vary smoothly. As shown by Henry and Hofrichter (1992) the 5 s/3 s and higher noise levels may be too high for a successful application of SVD, because the offset becomes comparable to the amplitude of the next significant rSVs. Nevertheless, even at the 10 s/5 s noise level, a reconstruction of the time-dependent difference maps with a sensible selection of a set of singular values and vectors is feasible because it is possible to identify the signal easily by inspecting the lSVs.

The number of significant singular values and vectors can be selected by examining the amplitude of the singular values and the autocorrelation of the rSVs and by assessing the quality of the lSVs. The first two criteria must be used if no other evidence exists to identify signal in the lSVs. The third criterion is usually not available for spectroscopic data. For example, the shape of an lSV difference absorption spectrum might obey general constraints but it also may vary within large boundaries. Therefore, a difference absorption spectrum observed in the lSVs may or may not contribute significantly to the signal. In marked contrast, the quality of the lSVs can readily be assessed in crystallographic data. Signal tends to cluster on or near to atoms of the chromophore or of the active site. Thus, the decision on which singular values and vectors are significant is facilitated.

SVD AS A NOISE FILTER: PHASE IMPROVEMENT

The two difference maps shown in Fig. 7 demonstrate one of the most important properties of the SVD, the ability to separate signal from noise. These difference maps were calculated from kinetic mechanism S1 on the highest, 10 s/5 s level of noise. Fig. 7 *A* shows the original mock difference map; the map in Fig. 7 *B* was reconstituted from the first five lSVs derived from SVD. The dark-state atomic structure and the structure of I3 are shown as a guide to the eye. In the lower left corner, one can easily identify the authentic difference electron density on and close to the chromophore. In Fig. 7 *A*, numerous additional electron density features arising largely from noise are distributed randomly in space. In contrast, Fig. 7 *B* is nearly free of these features but preserves all of the signal on the chromophore. SVD is able both to reduce the noise in difference maps and to fully preserve the signal.

There are three sources of noise in a difference Fourier map: noise in the difference structure amplitudes, errors in the native (dark) phase, and the phase error introduced by the difference approximation itself (Henderson and Moffat, 1971). The first two are random but the third is systematic. Fig. 8 *A* illustrates the effect of the difference Fourier approximation on attempts to measure the true difference structure factor $\Delta\mathbf{F}^{\mathrm{T}}$. $\Delta\mathbf{F}^{\mathrm{T}}$ is derived from the vector sum of the dark-state and intermediate-state structure factors weighted by their concentrations, and hence is time dependent. The difference Fourier approximation changes both the phase φ^{T} and amplitude $|\Delta\mathbf{F}^{\mathrm{T}}|$ of the true difference structure factor to φ^{DA} and $|\Delta\mathbf{F}|$, respectively. To examine the effects of noise due to the difference Fourier approximation alone, we examined difference maps from noise-free data generated from kinetic mechanism S1. Difference structure factors were generated by the difference approximation. The mean absolute phase difference $\langle|\varphi^{\mathrm{T}} - \varphi^{\mathrm{DA}}|\rangle$ was found to be 45° for acentric and around 43° for all reflections in all data sets, respectively. The values of $\varphi^{\mathrm{T}} - \varphi^{\mathrm{DA}}$ are uniformly distributed between −90° and 90° (Fig. 9 *A*), which explains the above result. All noise-free difference maps can therefore be considered as conventional Fourier maps calculated with the relatively small phase error of 45° for acentric reflections and essentially correct phases for centric reflections. Fig. 9 *B* shows the effect on the phase difference $\varphi^{\mathrm{T}} - \varphi^{\mathrm{DA}}$ if noise is added. Obviously, the box function in Fig. 9 *A* is convoluted with a Gaussian; the phase differences are distributed to larger values and the mean absolute phase error is increased to 56° (Fig. 9 *B*) and 66° (Fig. 9 *C*). When the signal-to-noise ratio is very low, for example in the final time points, the mean absolute phase

difference approaches 90°, as expected for a pair of randomly varying phases on the unit circle.

If noise introduced by the difference approximation were independent of time, SVD would be the right tool to minimize this noise and recover the magnitude and the true phase of the difference structure amplitude. To test this idea, we first analyzed difference maps calculated from 0 s/0 s data by SVD. SVD was applied to the time-dependent difference maps and the maps were reconstituted with the first three basis vectors. The mean phase difference $\langle|\varphi^{T} - \varphi^{DA}|\rangle$ decreased by ~5° to $\langle|\varphi^{T} - \varphi^{SVD}|\rangle$; phases φ^{SVD} were derived from a Fourier back-transformation of the reconstituted maps. Subjecting the maps to SVD thus has recovered part of the phase information necessary to correct for the difference approximation. We were able to recover more phase information and lower the average phase difference further by applying a phase recombination scheme (Fig. 10) employed by Ren et al. (2001). Here, additional information is supplied by using the dark-state structure factor $\mathbf{F}_{I0}$ and the time-dependent structure amplitude $|\boldsymbol{F}_{L}|$. The value of $\Delta\mathbf{F}^{SVD}$, whose amplitude and phase are calculated from a Fourier back-transformation of the reconstituted maps, is added to $\mathbf{F}_{I0}$. The resultant vector (*dotted line* in Fig. 10) determines the phase φ^{F_L} of the time-dependent structure amplitude $|\mathbf{F}_{L}|$. Due to the phase error in $\Delta\mathbf{F}^{SVD}$, the triangle characterized by $\mathbf{F}_{L}$, $\mathbf{F}_{I0}$, and $\mathbf{F}_{I0} + \Delta\mathbf{F}^{SVD}$ does not close. We adjust $\Delta\mathbf{F}^{SVD}$ so that the resultant $\Delta\mathbf{F}^{*}$ points to $\mathbf{F}_{L}$ with phase φ^{*}. New time-dependent difference maps can be calculated from the difference structure factors $\Delta\mathbf{F}^{*}$. If time-independent noise features still persist, the phase error can be reduced further by iteration. Fig. 11, *A* and *F*, shows the result after a second cycle: the phase improvements remain rather small. We conclude that a partial correction of error introduced by the difference approximation is completed within one SVD iteration cycle. Additional cycles do not contribute to a significant further reduction of this error.

We proceed further by adding realistic noise to the data. A time course of difference maps was generated from kinetic mechanism S1 with various levels of noise. From Fig. 8 *B*, one can see that $\Delta\mathbf{F}^{T}$ will change to $\Delta\mathbf{F}^{\#}$ as noise is added. In Fig. 11, the mean absolute phase difference $\langle|\varphi^{T} - \varphi^{DA}|\rangle$ is shown as $\langle|\varphi^{T} - \varphi_{cyclej}|\rangle$. If noise is added, all values of this phase difference lie significantly above the value of 43° observed in the noise-free data. Further, the values vary in a time-dependent fashion. For example, $\langle|\varphi^{T} - \varphi_{cyclej=0}|\rangle$ observed in the 10 s/5 s data set (Fig. 11 *E*) is 78° at the first time point, rises to 86° at the seventh, decreases to 74° at the 15th, and finally reaches 90° in the last time points. This time dependence can be understood in light of the underlying structural changes. When an intermediate characterized by a small structural difference from I0 is populated, the overall signal-to-noise ratio in the difference structure amplitudes decreases and a larger phase error results. In kinetic mechanism S, intermediate I1 is followed by I2, which decays to I3. The structure of I2 deviates from I0 only at and close to the chromophore, whereas the structures of I1 and I3 differ much more extensively from I0. Therefore, at time points 5–10, when I2 is most populated, the phase of the difference structure factors is more erroneous, but, in the last time points 19–21, the structural perturbation vanishes completely. Consequently, the mean phase error approaches 90° as expected from completely randomized phases.

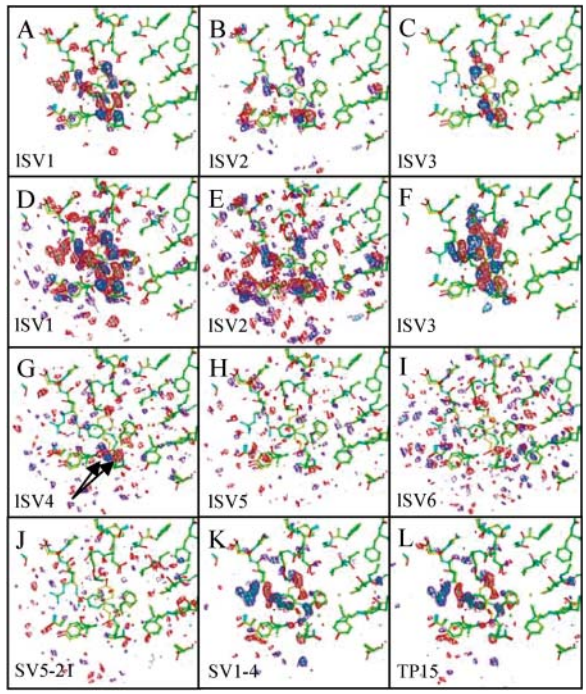

FIGURE 5 The lSVs derived from SVD analysis of mock data generated with kinetic mechanisms S1 and DE2 on the 2 s/1 s noise level. The sign of the lSVs is arbitrary; that of the reconstruction is correct. (*A*, *B*, and *C*) First three most significant lSVs, contour level red/white −3 σ/−4 σ, blue/cyan 3 σ/4 σ; (*D*, *E*, and *F*) at a lower contour level, red/white −2 σ/−3 σ; blue/cyan 2 σ/3 σ. (*G*, *H*, and *I*) Singular vectors lSV4–lSV6, contour level red/white −2 σ/−3 σ; blue/cyan 2 σ/3 σ. Arrows in lSV4 indicate signal. (*J*) Difference electron density SV5-21 at time point 15 is reconstructed from 17 least significant singular values and singular vectors lSV5–lSV21; contour level red/white −2 σ/−3 σ; blue/cyan 2 σ/3 σ. (*K*) Difference electron density SV1-4 at time point 15 is reconstructed from the four most significant singular values and singular vectors lSV1–lSV4; contour level: red/white −3 σ/−4 σ; blue/cyan 3 σ/4 σ. (*L*) Original difference electron density TP15 at time point 15; contour level: red/white −3 σ/−4 σ; blue/cyan 3 σ/4 σ.

However, SVD can contribute to the determination of the true difference structure factor $\Delta\mathbf{F}^{T}$ from $\Delta\mathbf{F}^{\#}$ even in the presence of a large amount of noise. After SVD, phase recombination can be done similar to the procedure demonstrated for the 0 s/0 s maps in Fig. 10. In addition, the noise in the time-dependent structure amplitude $|\mathbf{F}^{\#}|$ can be subjected to appropriate weighting, as outlined by Ren et al. (2001). The result are improved difference structure amplitudes $|\Delta\mathbf{F}^{*}_{cyclej}|$ and phases φ_{cyclei}, which can be used to calculate new time-dependent difference maps, and the procedure iterated. The phase improvements are in the order of 10°–15°

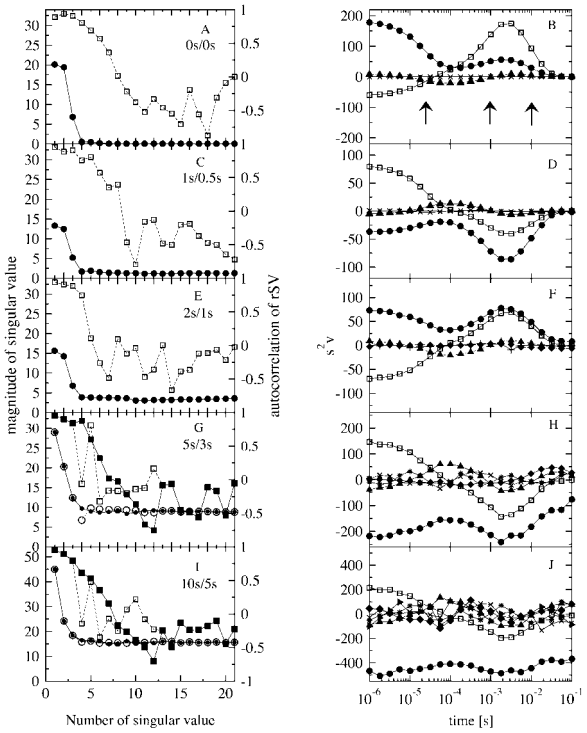

FIGURE 6 (*Left column*) Dependence of the magnitude of the singular values S (●) and the autocorrelation of the corresponding rSV (□) on the ordinal number of S, for mock data generated with kinetic mechanism S1. ○, magnitude of the singular value after rotation; ■, autocorrelation of the corresponding rSV after rotation. (*Right column*) Dependence of the magnitude of the corresponding rSVs on time. ●, first rSV; □, second rSV; ▲, third rSV; ◆, fourth rSV; ×, fifth rSV; ✻, sixth rSV; ▷, seventh rSV. (*A* and *B*) 0 s/0 s, no noise present. (*C* and *D*) Noise level 1 s/0.5 s; (*E* and *F*) noise level 2 s/1 s; (*G* and *H*) noise level 5 s/3 s; (*I* and *J*) noise level 10 s/5 s.

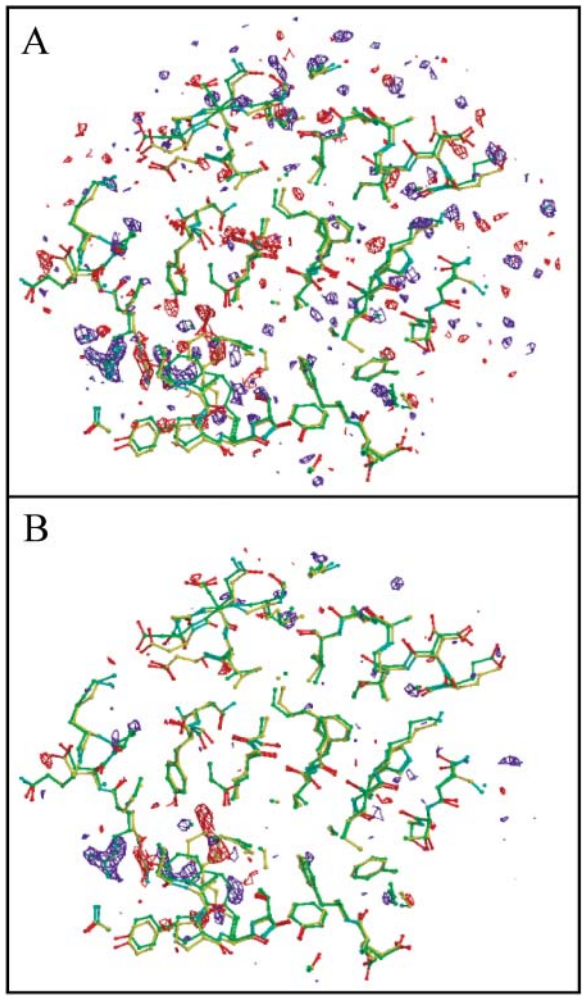

FIGURE 7 (*A*) Difference map at time point 15 on the 10 s/5 s noise level before SVD. (*B*) Same as (A) after SVD and reconstitution with the first five significant singular values and lSVs. Contour level: red/white $-3\ \sigma/-4\ \sigma$, blue/cyan $3\ \sigma/4\ \sigma$. Yellow and green atomic structures: Structure of dark state I0 and I3, respectively, as a guide to the eye.

as demonstrated for time point 15 in Fig. 11 *F*. Two to three cycles are usually sufficient for acceptable convergence, depending on the noise level.

We call this method SVD flattening, by analogy with solvent flattening (Wang, 1985). Difference electron density features originating either from phase error or from random noise in the difference structure amplitudes are partially excluded into less significant singular vectors. In contrast, those features persisting over several time points are most likely signal and retained in the maps. In SVD-flattened maps, the difference density is better defined on a higher contour level, the connectivity of the difference electron density is improved, and noise features reduced by comparison with the original maps. The necessary information is gathered automatically from the time axis in a purely analytical fashion.

The final result of the SVD procedure, with or without SVD flattening, is a set of noise-reduced (or signal-to-noise enhanced) phased difference maps, which can be used for further analysis of the mechanism (Fig. 1, *Prepare data matrix* $\boldsymbol{A}''$).

EXTRACTION OF MECHANISM FROM DIFFERENCE FOURIER MAPS: INTRODUCTION

The goals of any time-resolved spectroscopic or crystallographic experiment are the characterization of spectral or structural intermediates and the identification of their associated chemical reaction mechanism. Intermediates can be extracted from time-resolved data if their time-dependent concentrations can be described accurately, but in most instances it is not possible to extract them from complicated time traces without prior or additional knowledge (Lozier et al., 1992). In spectroscopic experiments, additional knowledge such as the nonnegativity and shape of absorption spectra, amplitude constraints, and temperature dependence may be sufficient for an unambiguous analysis of the data (Zimanyi and Lanyi, 1993; Nagle et al., 1995, Van Brederode et al., 1996). To test whether such an approach is feasible

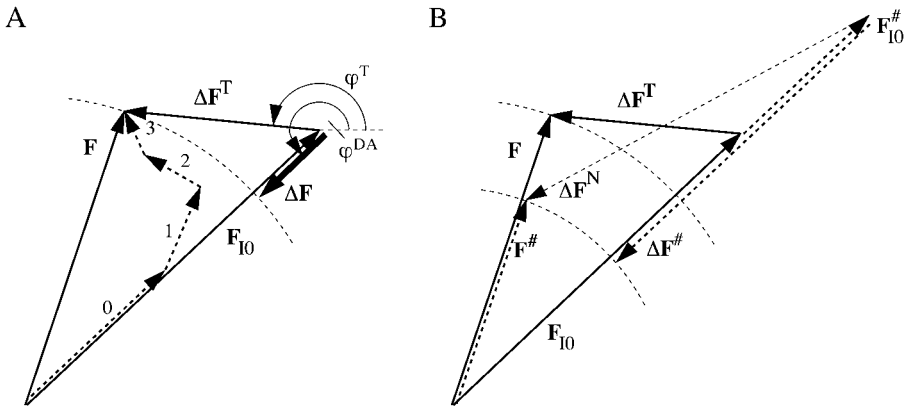

FIGURE 8 (*A*) Argand diagram for a reflection **h** to illustrate the construction of the time-dependent structure factor **F** from the dark-state structure factor $\mathbf{F}_{I0}$ and the intermediate structure factors $\mathbf{F}_{I1}$, $\mathbf{F}_{I2}$, and $\mathbf{F}_{I3}$, each weighted by their corresponding time-dependent concentrations c_{Ij}, resulting in vectors 0, 1, 2, and 3, respectively; $\Delta\boldsymbol{F}^{T}$, true difference structure factor; $\Delta\boldsymbol{F}$, difference structure factor after difference Fourier approximation; φ^{DA}, phase from difference approximation; φ^{T}, true phase of the difference structure factor. (*B*) Influence of noise: $\mathbf{F}_{I0}$ changes to $\mathbf{F}^{\#}_{I0}$ and **F** to $\mathbf{F}^{\#}$; the true difference structure factor $\Delta\mathbf{F}^{T}$ becomes the noisy difference structure factor $\Delta\mathbf{F}^{N}$ and, after the difference Fourier approximation, $\Delta\mathbf{F}^{\#}$.

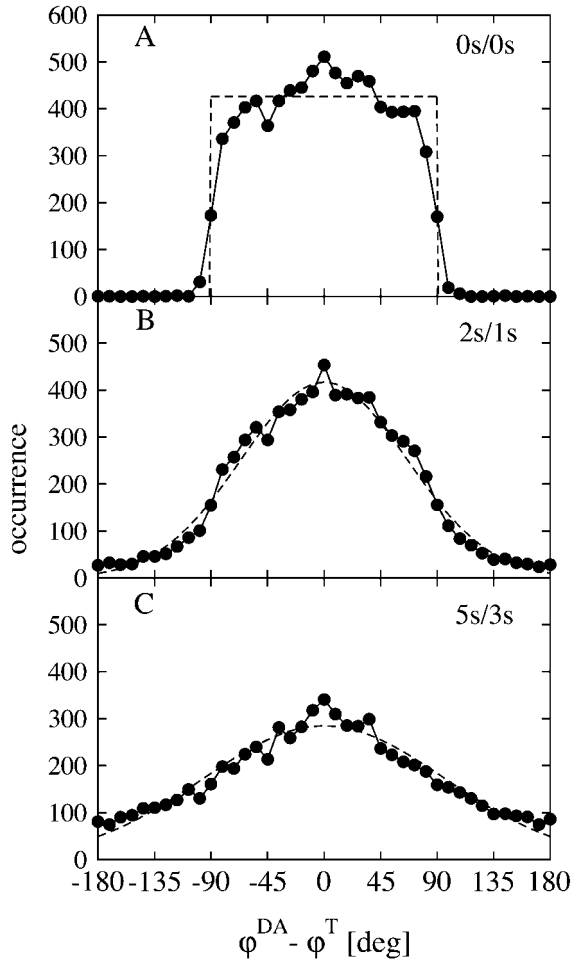

FIGURE 9 The effects of the difference Fourier approximation on the phase. (*A*) ●, phase difference between φ^{DA} generated by the difference approximation and the true phase φ^{T} in a data set of difference structure factors at time point 15 calculated from noise-free mock structure amplitudes; solid line: guide to the eye; dashed line: box function; (*B*) same as *A* with structure amplitudes on 2 s/1 s noise level, dashed line: Gaussian fit; (*C*) same as *A* with structure amplitudes on 5 s/3 s noise level, dashed line: Gaussian fit.

with time-resolved crystallographic data, we analyzed our simulated time-dependent difference electron density. The best time-dependent difference maps are represented by a noise-reduced or SVD-flattened data matrix $\boldsymbol{A}'$ or $\boldsymbol{A}''$, which can be decomposed into matrices $\boldsymbol{U}'$, $\boldsymbol{S}'$, and $\boldsymbol{V}'^{\mathbf{T}}$ (Eq. 2). The columns of $\boldsymbol{U}'$ are the lSVs, the rows of $\boldsymbol{V}'^{\mathbf{T}}$ are the rSVs, and the singular values comprise the diagonal matrix $\boldsymbol{S}'$. The N rSVs, $\mathbf{v}_n$, represent the temporal variation of the corresponding lSVs. Least-squares fitting of correctly weighted rSVs by some function is mathematically equivalent to fitting the entire data matrix using global analysis, with the advantage that the number of parameters necessary to describe the fit has decreased dramatically (Henry and Hofrichter, 1992). If a simple chemical kinetic mechanism holds (Moffat, 2001), the fit function must obey a sum of

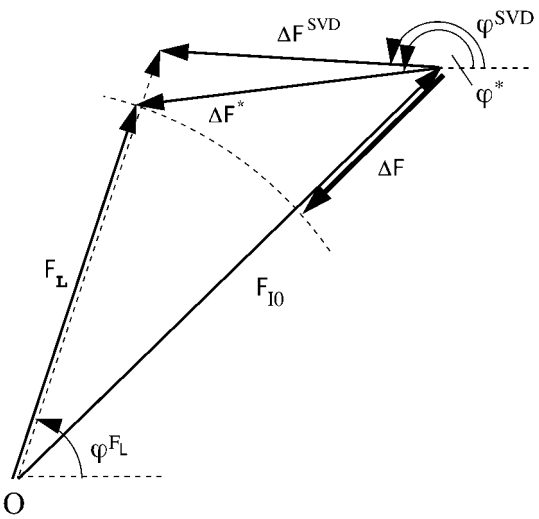

FIGURE 10 Phase combination for noise-free model data. $\mathbf{F_{I0}}$, dark state structure factor; $\mathbf{F_L}$, time-dependent structure amplitude with phase φ^{F_L}; $\Delta\mathbf{F}$, difference structure factor resulting from application of the difference approximation; $\Delta\mathbf{F}^{SVD}$, result from Fourier back-transformation of reconstituted maps with phase φ^{SVD}; $\Delta\mathbf{F}$*, improved difference structure factor with phase φ*.

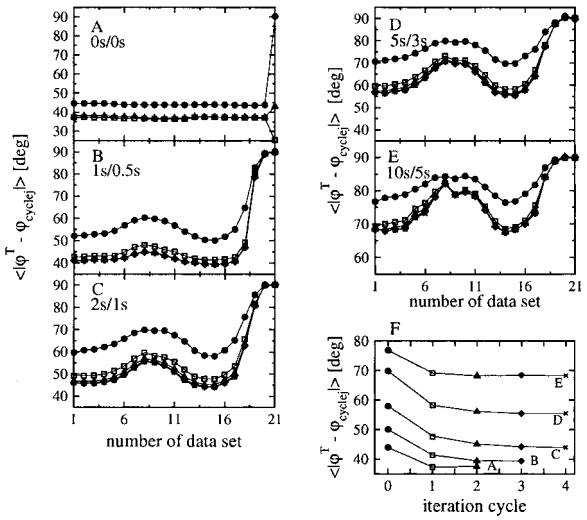

FIGURE 11 Phase improvement by SVD flattening. (*A–E*) Dependence of the mean absolute difference in phase angle between the true value φ^{T} and the current value in cycle *j*, $\varphi_{\text{cycle j}}$, on time, i.e., on the ordinal number of the data set. (*A*) 0 s/0 s, no noise present; (*B*) noise level 1 s/0.5 s; (*C*) noise level 2 s/1 s; (*D*) noise level 5 s/3 s; (*E*) noise level 10 s/5 s; (*F*) dependence of this mean absolute difference in phase angle on cycle number for time point/data set 15. ●, starting value; □, value after first cycle; ▲, value after second cycle; ◆, value after third cycle; ×, value after fourth cycle.

exponentials describing simple relaxation processes. The minimum number Q of intermediates in the mechanism then depends on the number of relaxation times observed by analysis of the rSVs. For example, three relaxation times are predicted for all kinetic mechanisms outlined in Fig. 3. The elements $v_{\text{n,i}}$ of the S significant rSVs, $\mathbf{v}_{\text{n}}$ ($n = 1..S$), are fit by a sum of exponentials whose amplitudes $D_{\text{n,q}}$ ($q = 1..Q$) plus an offset $D_{\text{n,0}}$, which differ for each rSV, and a unique set of relaxation rates, k_{q}, simultaneously used to fit all significant $\mathbf{v}_{\text{n}}$. The fit is weighted by the square of the corresponding singular value s_{n} according to Eq. 7:

$$s_{\text{n}}^2 v_{\text{n,i}} \approx v_{\text{n,i}}^{\text{fit}} = s_{\text{n}}^2 \left(D_{\text{n,0}} + \sum_{\text{q}=1}^{\text{Q}} D_{\text{n,q}} \cdot e^{-\text{k}_{\text{q}}\text{t}} \right). \qquad (7)$$

It is crucial to subsequent analysis of the time-resolved crystallographic data that the number of distinct relaxation times, common to all rSVs, can accurately be determined, as well as the amplitudes of each relaxation process for each rSV. We therefore explore how the magnitude of experimental errors influences the accuracy of such a determination.

Determination of relaxation times in the presence of noise

We identified the variation in the extent of reaction initiation (RI) as a major source of error in the time domain. Data are typically collected on different crystals for each time point. The laser pulse energy and the energy density (mJ/cm^2) are prone to significant fluctuations, as data collection is lengthy and data often are combined from different experimental trials. To analyze the consequences of these realistic experimental variations, we allowed the number of time points in

TABLE 3 Extraction of time-independent difference maps and reaction coefficients from time-dependent difference maps generated from mechanism S1 under a variety of experimental conditions

Extent of reaction initiation:	Constant at 20%									Variable 5%–17%	
Noise level:	1 s/0.5 s		2 s/1 s		5 s/3 s	7 s/4 s		10 s/5 s		2 s/1 s RI*	5 s/3 s RI*
Candidate mechanism:	S	DE	S	DE	S	S		S		S	S
Special conditions	None	None	None	None	None	None	SVD flattening		SVD flattening	None	RI**/SVD flattening
Criterion 1 or 2	+	+	+	+	+	−	+	−	+	+	+
Criterion 3	+	+	+	+	+	−	+	−	+	+	+
# of significant SVs	3	3	4	4	4	4	4	5	4	4	4
Rate coefficients											
k_{+1}	39,000	39,700	34,900	35,200	31,800	32,800	40,500		41,400	17,200	13,200
k_{+3}	1080	496	1240	635	2130	2080	1300		4030	3350	2300
k_{+5}/k_{-3}	86	200	71	190	21	8	50	<0	26	33	29
k_{+4}	–	467	–	476	–	–	–		–	–	–
MSD SOE 2 rates	225		185		907	1278	2536	2155	10053	64.6	838
3 rates	0.15		0.74		36.7	58.0	154	385	1033	52.6	441
4 rates								322	977	46.7	
MSD mechanism	0.15	0.16	5.96	6.04	315	972	1480	–	11249	58.4	583
Figures SOE mechan.	12 *A* 13 *A*	12 *A* 13 *B*	12 *B* 13 *C*	12 *B*	12 *C* 13 *D*	–	–	12 *D*	13 *E*	12 *F*	12 *H* 13 *F*

Mock data generated with the irreversible sequential mechanism S1 using reaction coefficients $k_{+1} = 40{,}000$, $k_{+3} = 1000$, and $k_{+5} = 100$. All data sets consisted of 21 time points. RI* denotes mock data that contained two outliers; RI** denotes data in which these outliers were corrected (see text). The right SVs were fitted by a sum of exponentials. The mean-square deviations between data and the fit are indicated in the row labeled ''MSD SOE.'' When the fit was further constrained by assuming the S or DE kinetic mechanism, the mean-square deviations are shown in the row labeled ''MSD mechanism.'' In the rows labeled ''Criterion 1 or 2'' and ''Criterion 3'', ''+'' indicates that the extracted time-independent intermediates satisfied this criterion; ''−'' indicates that they did not (see text). The units of the reaction coefficients are s^{-1}.

the data set to vary from five to 21 as the noise present in the structure amplitudes varied from the 1 s/0.5 s level to the 10 s/5 s level. In addition, the extent of RI for each time point was varied between 14% and 26% and in another case from 5% to 17%. After applying SVD to the mock data sets and preparing data matrix $\boldsymbol{A}'$ with the best-difference maps (Fig. 1, *Prepare data matrix* $\boldsymbol{A}'$), the number of significant singular vectors was chosen based on the magnitude of the singular values, on the autocorrelation of their corresponding rSVs, and on visual inspection of the lSVs, as explained above. The relaxation times were initially determined from the time courses in the rSVs, and the sum of exponentials was fitted to the rSVs, as shown in Fig. 12, *A–D*, for data on the 1 s/0.5 s–10 s/5 s noise level. Even at the 10 s/5 s level, the three predicted relaxation times are clearly observable despite the large offset at the end of the reaction. The fit is satisfactory with a mean-square deviation (MSD) of 385 (see Tables 3 and 4). If only two relaxation times were used, an MSD of 2155 was obtained; four relaxation times yielded an MSD of 322, quite similar to the one obtained from three relaxation times. We conclude that the minimal number of relaxation times and intermediates can be accurately selected even at high experimental noise.

When the RI was varied, the data points in the rSVs scatter around the fitted curve. Fig. 12, *E* and *F*, illustrate the effect when both a large variation of the RI from 5% to 17% and two time points as outliers (time point 4 with 0.1% RI and time point 11 with 40% RI) are included in simulation on the moderate 2 s/1 s noise level. Both outliers are clearly observable in the rSV (*arrows*). However, a fit with three exponentials is now difficult to distinguish from one with two exponentials; the MSD of both is comparable (see Tables 3 and 4). Because these two data points are clear outliers, we can adjust the initial data matrix $\boldsymbol{A}$ to account for the systematic error. The difference electron density at time point t_i is multiplied by a factor determined by the ratio of the fit value from the sum of exponentials and the magnitude of the element of the first rSV at this time point (Fig. 1, *Correct for extent of reaction initiation*). The result is data matrix $\boldsymbol{A}'$, whose rSV after a second cycle of SVD exhibit much smaller variations (Fig. 12 *F*). We conclude that variation in the extent of RI propagates through the SVD analysis into the rSVs. If the variation is sufficiently extreme, outliers can be identified, corrected in the original data matrix, and the SVD analysis repeated. That is, used judiciously, SVD can reduce certain sources of noise arising from systematic as well as from random error. If the random noise increases further as shown for the 5 s/3 s level, SVD flattening can be used in addition to increase the reliability of the relaxation times (Fig. 12, *G* and *H*; Tables 3 and 4).

Fitting a chemical kinetic mechanism to the rSV

The ability to accurately fit the rSVs by a sum of exponentials already limits the number of possible mechanisms that can account for the data. Only those reaction mechanisms that produce the observed number of relaxation times/rates can be used for further analysis (see Fleck, 1971). One general reaction mechanism for a reaction involving four states, I0, I1, I2, and I3, and three relaxation rates is shown in Fig. 3 *A*. All possible mechanisms for this case can be obtained from this general mechanism by setting some of the rate coefficients to zero. Two examples are given in Fig. 3, *B* and *C*. If a candidate reaction mechanism with J intermediates and rate coefficients k_r ($r = 1..R$) is chosen, the time-dependent concentration of the jth intermediate, $c_{Ij}(k_r,t)$, can be calculated at time point t_i by solving the coupled differential equations describing the mechanism, which is given in the most general form by:

$$c_{Ij}(k_r, t) = \sum_{p=1}^{J} A_{p,j}(k_r) \cdot \exp[-B_{p,j}(k_r) \cdot t], \qquad (8)$$

where the amplitudes $A_{p,j}$ and exponents $B_{p,j}$ are dependent on the set of rate coefficients k_r (Matsen and Franklin, 1950). By varying the magnitude of the rate coefficients, the concentrations are fit to the elements of the most significant rSVs using the scale factors $E_{n,j}$ for each concentration (Eq. 9):

$$s_n^2 v_{n,i} \approx v_{n,i}^{fit} = s_n^2 \cdot \sum_{j=1}^{J} E_{n,j} \cdot c_{Ij}(k_r, t). \qquad (9)$$

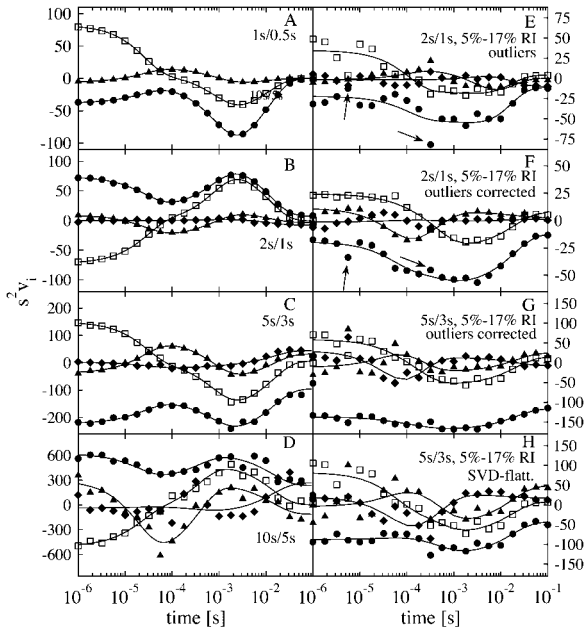

FIGURE 12 Fit of the three or four most significant rSVs with various levels of noise. ●, first rSV; □, second rSV; ▲, third rSV; ◆, fourth rSV. The solid lines are a fit to a sum of three exponentials with no mechanism constraint applied. Quantitative details of the fit are given in Tables 3 and 4. (*A*) Noise level 1 s/0.5 s; (*B*) noise level 2 s/1 s; (*C*) noise level 5 s/3 s; (*D*) noise level 10 s/5 s; (*E*) noise level 2 s/1 s and variable level of reaction initiation between 5% and 17% with two outliers (shown by the *arrows*) with 0.5% and 40% reaction initiation; (*F*) same as *E* but with outliers corrected in the original data and reanalyzed by SVD; (*G*) same as *F*, but noise on the 5 s/3 s level; (*H*) same as *G* after SVD flattening.

The time-independent difference electron density $\Delta\rho_{Ij}$ of a candidate intermediate I_j, represented by the vector $\boldsymbol{b}_{Ij}$, is then calculated according to Eq. 10 (see also Henry and Hofrichter, 1992):

$$\mathbf{b}_{Ij} = \sum_{n=1}^{S} \mathbf{u}_n \cdot s_n \cdot E_{n,j}. \quad (10)$$

Each of the finite number of candidate mechanisms that exhibit J intermediates may be subjected to this process (Fig. 1, *List all possible mechanisms*). Each candidate mechanism then yields its own set of candidate intermediates I_j, represented by $\Delta\rho_{Ij}$. How shall these candidate intermediates be assessed? If all intermediates for a particular mechanism pass the assessment, the candidate mechanism and intermediates are consistent with the data; but if one or more intermediates fail the assessment, the candidate mechanism and the intermediates are rejected. This type of assessment lies at the heart of all investigations of mechanism based on kinetic data.

After fitting the rSV with candidate mechanisms (Fig. 13, *A–F*), time-independent difference electron density was extracted and assessed using three criteria.

As a first criterion, the extracted time-independent difference maps $\Delta\rho_{Ij}$ were compared at all grid points within the mask to the corresponding accurate reference difference maps $\Delta\rho_{Rj}$ by evaluating the linear correlation coefficient between the maps (Drenth, 1994). Correlation coefficients below 0.5 indicate that the extraction failed. For example, comparing the reference difference electron density shown in Fig. 14 *A* with the difference electron density in Fig. 14 *C* gives a linear correlation coefficient of 0.3, whereas that with the difference electron density shown in Fig. 14 *D* is 0.7. Low correlation coefficients could arise due to three reasons. First, the mechanism might be incorrect. Second, the rate coefficients were not fit accurately due to large noise, which would inhibit the extraction of the time-independent maps by Eq. 10. Third, unrealistically high noise might prevent the analysis altogether (see Fig. 13 *E* as an example).

To distinguish these possibilities, two further criteria were employed based on inspection of the difference maps. Criterion 2 is the visual counterpart of the linear correlation coefficient and is usable with experimental maps where no accurate reference maps are available. The linear correlation coefficient was used here to calibrate the otherwise subjective judgment. Criterion 2 takes advantage of prior structural knowledge. Does the electron density make chemical sense? For example, is negative difference electron density observed on dark state atoms and is positive difference electron density consistent with structures commonly observed in proteins? Do positive and negative difference densities flank certain dark state atoms? If the difference electron density features assumed to be signal are weak and masked by noise, this criterion fails. Criterion 3 is successfully fulfilled if the time-independent difference electron density for an intermediate can be interpreted by one consistent atomic structure. This criterion fails if difference electron density features mix with each another in a manner that requires the presence of other atomic structures. Fig. 14 *E* shows an example. The features α, β, and γ cannot be accounted for by the structures shown in green and yellow. The map does not arise from a single structure; difference electron density remains mixed in from another intermediate. Hence, criterion 3 is not met and the mechanism can be rejected.

These structure-based criteria are the only criteria that can be applied as subjective input to a real time-resolved crystallographic experiment to allow the validity of a mechanism to be assessed (Fig. 1, *Criteria 1–3 met?*). However, it is unknown how sensitive these criteria are for the analysis of the time course of difference maps and under which circumstances this analysis might break down. To address these points, we analyzed the two candidate mechanisms S and DE from Fig. 3. Mechanism S is the correct mechanism to analyze difference maps generated by mechanism S1 and S2 but is incorrect for those maps generated from DE1 or DE2. For mechanism DE, the situation is reversed. The extracted time-independent difference electron densities $\Delta\rho_{Ij}$ were examined by criteria 1, 2, and 3. These are shown as a "+" if rated as a success or as a "−" if they failed, for the simulations in Tables 3 and 4. Examples of maps that satisfy or fail these criteria are shown in Fig. 14.

Table 3 presents the analysis of mock data generated with the simplest kinetic mechanism S1 (Figs. 3 and 4) with both the S (correct) and DE (incorrect) mechanisms, and reveals

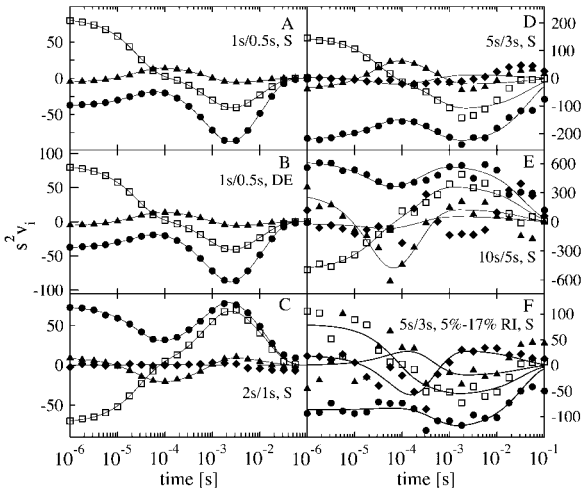

FIGURE 13 Fit of the candidate chemical kinetic mechanisms S or DE to the three or four most significant rSVs with various levels of noise. ●, first rSV; □, second rSV; ▲, third rSV; ◆, fourth rSV. The solid lines are a fit to a sum of three exponentials. (*A*) Noise level 1 s/0.5 s, candidate mechanism S; (*B*) noise level 1 s/0.5 s, mechanism DE; (*C*) noise level 2 s/1 s, candidate mechanism S; (*D*) noise level 5 s/3 s, candidate mechanism S; (*E*) noise level 10 s/5 s, candidate mechanism S; (*F*) noise level 5 s/3 s, variable level of reaction initiation between 5 and 17%, corrected for outliers, after SVD flattening, candidate mechanism S.

some interesting features. First, the analysis is relatively insensitive to noise. Criteria 1–3 are met by both mechanisms and the fit is satisfactory until the 5 s/3 s noise level (Fig. 13, *A–D*). Even at the highest 10 s/5 s noise level, the criteria can be met when applied to improved, SVD-flattened maps. Second, the analysis is insensitive to substantial variation in the extent of RI when experimentally realistic noise at the 2 s/1 s level is present. Criteria 1–3 are met when the RI varies from 14% to 26% or from 5% to 17%, with or without the outliers. Third, the correct number of relaxation times is identified and there is a good quantitative agreement between the rate coefficients used to generate the mock data and those extracted from the analysis. Fourth and most critically for the present purposes, the two distinct mechanisms S and DE could nevertheless not be distinguished (Fig. 13, *A* and *B*). If a fit with mechanism S successfully generated time-independent electron density maps that met criteria 1–3, so did the fit with mechanism DE.

Similar conclusions are drawn from the analysis in Table 4 of three sets of mock data generated with the more complicated kinetic mechanisms S2, DE1, and DE2 (Figs. 3 and 4). There is somewhat more sensitivity to noise in the difference amplitudes (compare the effect of noise levels 2 s/1 s and 5 s/3 s in Tables 3 and 4). The analysis is relatively insensitive to variation in the extent of RI between 14% and 26% (data not shown). If a larger variation in the extent of RI between 5% and 17% was allowed, the data generated from mechanism S2 could not be analyzed successfully, because there is no way to identify and fit the three rate coefficients for the mechanism accurately. The analysis of the data generated from mechanism DE1 was more robust against variations of the extent of RI. Only the reduction of the number of time points from 21 to 10 prevented a successful analysis. In most simulations, however, the analysis was relatively insensitive to reduction in the number of time points (not shown in Tables 3 and 4). In the case of mechanism DE2 analyzed with candidate mechanism S, admixtures of I2 can be observed in $\Delta\rho_{I3}$ (Fig. 14 *F*) and hence, criterion 3 failed. In this case, the candidate mechanism S can be discriminated from DE. Nevertheless, in most cases the two distinct mechanisms cannot be distinguished.

The inability to distinguish quite different mechanisms by these criteria is a direct consequence of the loss of absolute scale in the rSVs, a problem that is described and well known in spectroscopy (e.g., Zimanyi et al., 1999a). Fig. 15 *B* illustrates the problem. Here, the first most significant rSV (*circles*) was determined from a time course of difference maps generated from the kinetic mechanism S1. The solid line is the result of the fit by both candidate chemical kinetic mechanisms S and DE. Regardless of which mechanism is used, there is a similar goodness of fit. However, the con-

TABLE 4 Extraction of time-independent difference maps and reaction coefficients from time-dependent difference maps generated from mechanisms S2, DE1 and De2 under a variety of experimental conditions

Extent of reaction initiation:	Constant at 20% in all time points						Variable between 5% and 17%				
Number of time points:	21						10	21	10	21	
Noise level:	2 s/1 s				5 s/3 s		2 s/1 s				
Kinetic mechanism to generate data:	S2	DE1	DE2		DE2		S2		DE1		
Candidate mechanism:	S	DE	DE	S	DE		S		DE		
Special conditions	None	None	None	None	None	SVD flattening	None	None	SVD flattening	None	SVD flattening
Criterion 1 or 2	+	+	+	+	+	+	−	−	−	−	+
Criterion 3	+	+	+	−	−	+	−	−	−	−	+
# of significant SVs	4	4	4	4	4	4	4	4	4	4	4
Rate coefficients:											
k_{+1}	2570	8720	11,800	11,600	13,400	16,500	3080	1690	12,300	10,800	13,700
k_{+3}	3070	420	2910	5840	5890	2470	2370	273	2040	879	415
k_{+5}/k_{-3}	438	205	2940	30	611	1940	405	120	187	188	53
k_{+4}		210	60		82	67			423	40	544
MSD SOE 2 rates	30.1	34.4	14.1		147	664	13.3	37.4	69.4	29.4	90.1
3 rates	1.2	1.6	2.0		33.0	79.0	11.7	36.9	60.0	25.1	81.6
4 rates					32.2						81.6
MSD mechanism	13.1	8.9	7.5	7.4	147	366	16.6	39.0	72.7	28.6	87.2

Mock data generated with the mechanisms S2, DE1, and DE2. For mechanism S2, $k_{+1} = 2000$, $k_{+3} = 3000$, $k_{+5} = 900$; for mechanism DE1, $k_{+1} = 9500$, $k_{+3} = 330$, $k_{-3} = 210$, $k_{+4} = 400$; for mechanism DE2, $k_{+1} = 15{,}000$, $k_{+3} = 2000$, $k_3 = 2000$, $k_{+4} = 100$. The right SVs were fitted by a sum of exponentials. The mean-square deviations between data and the fit are indicated in the row labeled ''MSD SOE.'' When the fit was further constrained by assuming the S or DE kinetic mechanism, the mean-square deviations are shown in the row labeled ''MSD mechanism.'' In the rows labeled ''Criterion 1 or 2'' and ''Criterion 3'', ''+'' indicates that the extracted time-independent intermediates satisfied this criterion; ''−'' indicates that they did not (see text). The units of the reaction coefficients are s^{-1}.

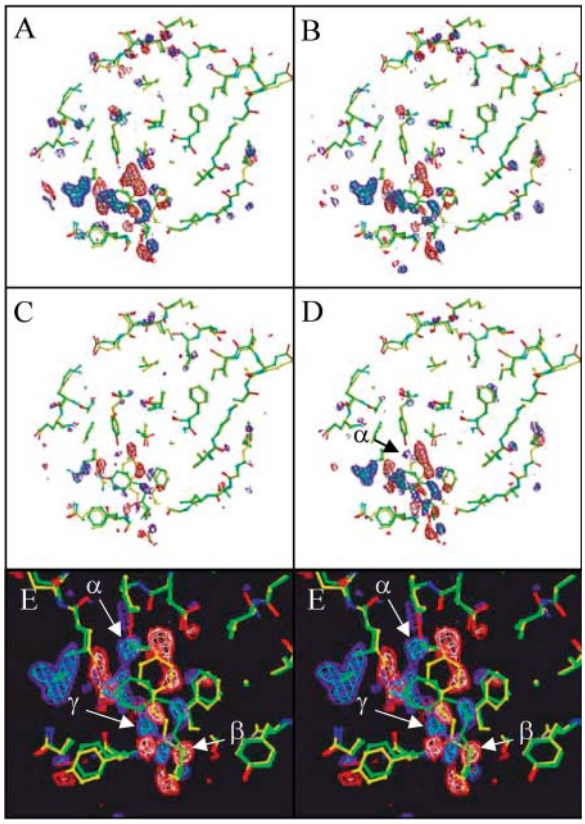

FIGURE 14 Candidate time-independent difference electron density maps $\Delta\rho_{I3}$ for intermediate I3; contour level: red/white $-3\ \sigma/-4\ \sigma$; blue/cyan $3\ \sigma/4\ \sigma$; yellow atomic structure: dark structure; green atomic structure: structure of I3. (*A*) Reference map $\Delta\rho_{I3}$ for I3 with no phase error and no noise; (*B*) $\Delta\rho_{I3}$ extracted with either mechanism S or DE from mock data generated with kinetic mechanism S1, noise level 2 s/1 s; (*C*) $\Delta\rho_{I3}$ extracted with mechanism S from mock data generated with kinetic mechanism S2, noise level 5 s/3 s; (*D*) as (*C*) after one cycle of SVD flattening; (*E*) stereo representation of $\Delta\rho_{I3}$ extracted with mechanism S from mock data generated with kinetic mechanism DE2, noise level 2 s/1 s. Features α, β, and γ indicated by the arrows cannot be accounted for by a single structure.

centrations calculated from the fitted rate coefficients differ in amplitude (*dotted* and *dashed lines*) for candidate mechanisms S and DE, respectively. Because the scale of the rSV is unknown, the factors $E_{n,j}$ (Eqs. 9 and 10) are used as linear fit parameters to adjust the concentrations. The true differences in concentration become irrelevant for the fitting problem and, hence, those mechanisms cannot be distinguished.

To retain the absolute scale present in the difference maps themselves, we developed a scheme we denote as ''posterior analysis'' (Fig. 1, *Posterior analysis*) that allows differences in the concentrations to be quantified. This scheme requires that, first, the structures of all ''candidate'' intermediates must be refinable from the corresponding time-independent electron density (in our simulations, the structures of the intermediates were taken as known). From the set of refined atomic structures and the dark-state structure, structure factors are calculated. The candidate mechanisms are each used to determine the time-dependent concentrations of their corresponding intermediates. From these concentrations and the structure factors, a set of candidate time-dependent difference electron density maps $\Delta\rho^{cand}(t)$ is generated following Eqs. 3 and 4. Each set of time-dependent density maps can then be compared to the original $\Delta\rho(t)$ by means of the total squared difference in the difference electron density over the entire time course. Because both $\Delta\rho^{cand}(t)$ and $\Delta\rho(t)$ and the maps in data matrix $\boldsymbol{A}$ or $\boldsymbol{A}'$ are on an absolute scale, posterior analysis resolves the ambiguity of scale. Table 5 illustrates the outcome of this approach when mock data at various noise levels are generated with mechanisms S1 and DE2, and analyzed with a candidate S or DE mechanism. The ratio of the total squared difference obtained with the incorrect candidate mechanism to the correct candidate mechanism is denoted T^*. The larger the value of T^*, the more powerful the distinction between the correct and incorrect mechanisms. T^* clearly diminishes as the noise level increases. However, the correct and incorrect candidate mechanisms can be successfully discriminated by this posterior analysis up to a noise level of 5 s/3 s (Table 5).

To summarize, three different cases must be considered. 1), Mechanisms can be distinguished readily if the $c_{Ij}(t)$ are qualitatively different as in the case of data generated with mechanism DE: candidate mechanism S simply cannot produce the observed biphasic decay of the I2 concentration (Fig. 15 *A*). 2), When the $c_{Ij}(t)$ are only quantitatively different but are qualitatively similar, as in most of the other cases (Fig. 15 *B*), the loss of the absolute scale generates a problem, which can be solved by posterior analysis. 3), When $c_{Ij}(t)$ are very similar qualitatively and quantitatively, neither SVD nor any other method can distinguish between the mechanisms.

TABLE 5 Comparison of different candidate kinetic mechanisms by posterior analysis

		Candidate mechanism		
		S	DE	
Kinetic mechanism used to generate data	Noise level	Total squared difference $[(el/Å^3)^2]$	Total squared difference $[(el/Å^3)^2]$	T^*
S1	1 s/0.5 s	3.5	18.6	5.3
	2 s/1 s	31.9	63.1	2.0
	5 s/3 s	300.1	304.5	1.0
DE2	1 s/0.5 s	56.4	14.9	3.8
	2 s/1 s	99.2	57.9	1.7
	5 s/3 s	319.3	311.9	1.0

Total squared electron density differences were determined using 10 time points and grid points in the mask that are larger than 1 σ or smaller than $-1\ \sigma$, respectively, in either map. All difference electron density values on common (absolute) scale. The column headed T^* is the ratio between the total squared difference for the incorrect and correct kinetic mechanisms.

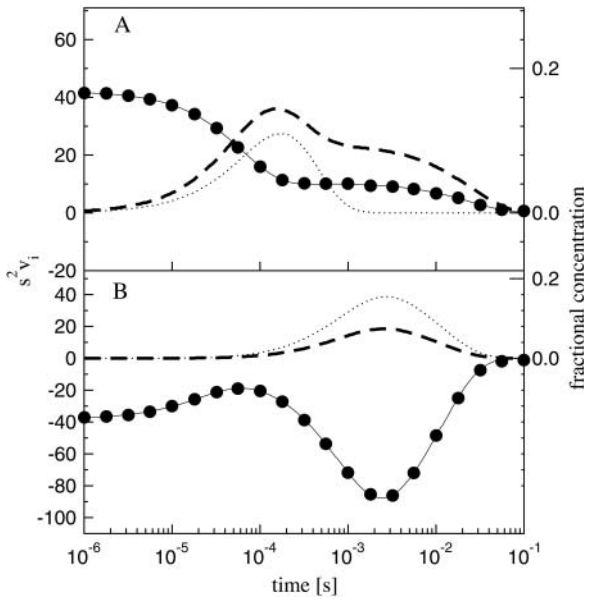

FIGURE 15 Consequences of the loss of the absolute scale in the rSVs. (*A*) Mock data generated with kinetic mechanism DE, noise level 2 s/1 s. ●: time dependence of the first rSV. Solid line: fit of the first rSV by a sum of three exponentials, for either mechanism S or DE. Dashed line: fractional concentration of intermediate I2 for candidate mechanism DE. Dotted line: same, for candidate kinetic mechanism S; (*B*) mock data generated with kinetic mechanism S1, noise level 2 s/1 s. ●: time dependence of the first rSV. Solid line: fit of the first rSV by a sum of three exponentials, for either mechanism S or DE. Dashed line: fractional concentration of intermediate I3 for candidate mechanism DE. Dotted line: same, for candidate mechanism S.

DISCUSSION

The analysis of time-resolved crystallographic data has become increasingly important because it has become technically feasible to produce a large series of these maps over several decades in time (Srajer et al., 2001; Ren et al., 2001; Anderson et al., unpublished). SVD is commonly used in the analysis of time-resolved spectroscopic data, and its application to time-resolved crystallographic data is natural, as the two types of experiments have much in common. Electron density as a function of space (and time) can be treated in a mathematically equivalent way to absorption data as a function of wavelength (and time), as both depend linearly upon the concentration of the underlying chemical species. Although SVD has recently been applied to x-ray diffraction intensities (Oka et al., 2000, 2002), these do not depend linearly on the underlying concentrations. It is not clear what the effect on the output of the SVD procedure would be. In difference maps, the data is sampled at grid points in the unit cell, whereas in spectroscopic data the sampling is over the selected wavelength range and depends on the method selected by the experimenter. Athough crystallography and spectroscopy are similar in these regards, different but complementary chemical information is provided by each method. Intermediates found in difference maps may have no spectroscopic counterpart simply because the structural changes do not affect the spectroscopic data at the wavelength range selected. Conversely, spectral intermediates associated with large absorption changes may prove to be silent in difference maps inasmuch as large electronic changes do not necessarily result in large, resolvable atomic displacements (Ng et al., 1995).

Much of the methodology that has been successfully used to analyze spectroscopic data can be directly applied to crystallographic data. In a manner analogous to the analysis of time-resolved spectroscopic data (Nagle et al., 1995), one can simplify the assumed mechanism by systematically setting rate coefficients to zero and comparing the deviation from the observed difference maps by posterior analysis. For example, with mock data generated from mechanism S1, such an approach would test all possible mechanisms present in the general reaction scheme in Fig. 3 *A*. Ideally, an irreducible set of rate coefficients such as k_{+1}, k_{+3}, and k_{+5} in this example would describe the simplest mechanism compatible with the data. By Occam's razor, all other more complicated mechanisms should be disregarded, even if they have a similar quality of fit. Candidate mechanisms will be selected not only by evaluating the deviations from observed data but also by the time-independent difference electron density, which must be chemically and structurally plausible. If some electron density is left uninterpreted by a single atomic model in the vicinity of the active center or the chromophore, there must be another structure or conformation present in the data and the mechanism should be rejected. There is essentially no counterpart in spectroscopic data to this powerful structural constraint. However, intermediates may have the same time dependence in which case they are intrinsically inseparable by SVD or indeed by any other method. The ability to assess a mechanism's validity by applying very general structural constraints is unique to time-resolved crystallography and should eventually lead to a confined set of candidate mechanisms.

One practical concern in the application of SVD lies in the selection of those grid points that will form the data matrix $\boldsymbol{A}$. It is desirable to choose only those grid points containing signal for analysis by SVD, as grid points that do not contain meaningful signal only add noise to the analysis (Henry and Hofrichter, 1992). Signal in difference maps is concentrated in a limited volume of the asymmetric unit, so the majority of grid points contain only noise and a minority contain signal (see Srajer et al., 2001). However, selection of a restricted number of grid points may introduce bias into the analysis, as grid points containing the signal of a lower population intermediate might be overlooked. On the other hand, the selection of too numerous grid points decreases the significance of the real signal. In this study, we used a mask to separate the protein plus a small margin from the bulk water in the crystal. Within such a mask, protein intermediate structures may freely evolve, although the volume analyzed is substantially smaller than the entire asymmetric unit. After

SVD, the contents of the mask can be displayed and the significance of the lSVs can be judged visually. Random noise tends to spread uniformly through space, whereas signal is consistent with difference electron density features that concentrate on top of atoms and on other chemically meaningful positions. These features tend to be connected in space and to have a well-defined shape. These criteria can be used, along with the magnitude of the SV and the autocorrelation of the rSV, to decide on the appropriate number of singular values to use in the subsequent analysis. It may also allow generation of a new, smaller mask that contains all of the ''real'' signal, as assessed by an analysis of the initial SVD output, although excluding the noise that may have been retained by the older, larger mask. This also allows for the possibility of the application of phase improvement methods such as isomorphous noise suppression (Ren et al., 2001), after SVD has separated signal from noise in a largely unbiased manner.

The noise level in the structure amplitudes and variations of the extent of reaction initiation are major problems associated with the SVD analysis. In time-resolved spectroscopy the data exhibit signal-to-noise ratios better than 100. In time-resolved difference maps, the signal arises from the displacement of a few electrons and the noise is in the order of some fraction of an electron, with typical signal-to-noise ratios less than 10 (Srajer et al., 2001). Our mock data suggest that the SVD flattening increases this value to ~17. This may be judged from the magnitude of the true electron density feature on the phenolate oxygen of the PYP chromophore relative to the noise level, before and after SVD flattening. The limit of noise tolerated in the structure amplitudes depends on the complexity of the mechanism and lies between the 2 s/1 s level and the 5 s/3 s level for moderately complicated mechanisms, and substantially above the 5 s/3 s level for simple reaction mechanisms. This suggests strongly that experimental structure amplitudes of the accuracy currently obtained (Ren et al., 1999) are indeed adequate for SVD analysis and even sophisticated schemes such as SVD with self-modeling (Zimanyi et al., 1999a,b; Kulcsar et al., 2001) may be applied.

The analysis of this mock data gives insight into the proper design of time-resolved crystallographic experiments by illustrating the effects on the applicability of SVD of noise levels, varying extent of reaction initiation, and number of time points. Of these confounding factors, the difference in the extent of reaction initiation from time point to time point has proved to be the most challenging in earlier experiments (Ren et al., 2001; Srajer et al., 2001). In such experiments where reciprocal space is the fast variable (all of reciprocal space is collected for each time point from a single crystal), large differences in the extent of RI are observed from time point to time point because of crystal-to-crystal variation (different absorption characteristics for crystals of different size, etc.). However, in an experiment where time is the fast variable (the same slice of reciprocal space is recorded for all time points from a single crystal), such effects would be minimized. Data collected using such new methods would be very well suited to analysis by SVD, and should be of similar quality to the simulated low noise data used in this study.

The identification of mechanism and the refinement of time-independent structural intermediates from experimental time-resolved crystallographic data are likely to be possible, using SVD and the other strategies developed here.

Discussions with Vukica Šrajer, Spencer Anderson, and Hyotcherl Ihee are highly appreciated.

Supported by National Institute of Health grants GM36452 and RR07707 to K.M.

REFERENCES

Alter, O., P. O. Brown, and D. Botstein. 2000. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. USA*. 97:10101–10106.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.

Borgstahl, G. E. O., D. R. Williams, and E. D. Getzoff. 1995. 1.4 Ångstrom structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry*. 34:6278–6287.

Bourgeois, D., T. Ursby, M. Wulff, C. Pradervand, A. Legrand, W. Schildkamp, S. Laboure, V. Srajer, T. Y. Teng, M. Roth, and K. Moffat. 1996. Feasibility and realization of single pulse Laue diffraction on macromolecular crystals at ESRF. *J. Synchrotron Rad.* 3:65–74.

Box, G. E. P., and M. E. Muller. 1958. A note on the generation of random normal deviates. *Ann. Math. Stat*. 28:610–611.

Doruker, P., A. R. Atilgan, and I. Bahar. 2000. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. *Proteins*. 40:512–524.

Drenth, J. 1994. Principles of Protein X-ray Crystallography. Springer, New York.

Fleck, G. M. 1971. Chemical Reaction Mechanisms. Holt, Rinehart and Winston, New York.

Genick, U. K., G. E. Borgstahl, K. Ng, Z. Ren, C. Pradervand, P. M. Burke, V. Srajer, T. Y. Teng, W. Schildkamp, D. E. McRee, K. Moffat, and D. E. Getzoff. 1997. Structure of a photocycle intermediate by millisecond time-resolved crystallography. *Science*. 275:1471–1475.

Genick, U. K., S. M. Soltis, P. Kuhn, I. L. Canestrelli, and D. E. Getzoff. 1998. Structure at 0.85 A resolution of an early protein photocycle intermediate. *Nature*. 392:206–209.

Golub, G. H., and C. Reinsch. 1970. Handbook series linear algebra: singular value decomposition and least squares solutions. *Numer. Math.* 14:402–420.

Henderson, R., and K. Moffat. 1971. The difference Fourier technique in protein crystallography: errors and their treatment. *Acta Crystallogr. B*. 27:1414–1420.

Henry, E. R. 1997. The use of matrix methods in the modeling of spectroscopic data sets. *Biophys. J*. 72:652–673.

Henry, E. R., and J. Hofrichter. 1992. Singular value decomposition: Application to analysis of experimental data. *Meth. Enzymol*. 210:129–192.

Hoff, W. D., I. H. M. van Stokkum, H. J. van Ramesdonk, M. E. von Brederode, A. M. Brouwer, J. C. Fitch, T. E. Meyer, R. van Grondelle, and K. J. Hellingwerf. 1994. Measurement and global analysis of the absorbance changes in the photocycle of photoactive yellow protein from Ectothiorhodospira halophila. *Biophys. J*. 67:1691–1705.

Kulcsar, A., J. Saltiel, and L. Zimanyi. 2001. Dissecting the photocycle of the bacteriorhodopsin E204Q mutant from kinetic multichannel difference spectra. Extension of the method of singular vector decomposition with self-modeling to five components. *J. Am. Chem. Soc.* 123: 3332–3340.

Lozier, R. H., A. Xie, J. Hofrichter, and G. M. Clore. 1992. Reversible steps in the bacteriorhodopsin photocycle. *Proc. Natl. Acad. Sci. USA.* 89:3610–3614.

Matsen, F. A., and J. L. Franklin. 1950. A general theory of coupled sets of first order reactions. *J. Am. Chem. Soc.* 72:3337–3341.

McRee, D. E. 1999. Practical Protein Crystallography, 2nd ed. Academic Press, San Diego.

Moffat, K. 1989. Time-resolved macromolecular crystallography. *Annu. Rev. Biophys. Biophys. Chem.* 18:309–323.

Moffat, K. 2001. Time-resolved biochemical crystallography: a mechanistic perspective. *Chem. Rev.* 101:1569–1581.

Moffat, K. 2002. The frontiers of time-resolved macromolecular crystallography: movies and chirped x-ray pulses. *Faraday Discuss.* 122: 65–77.

Moffat, K., and R. Henderson. 1995. Freeze trapping of reaction intermediates. *Curr. Opin. Struct. Biol.* 5:656–663.

Nagle, J. F., L. Zimanyi, and J. K. Lanyi. 1995. Testing BR photocycle kinetics. *Biophys. J.* 68:1490–1499.

Ng, K., E. D. Getzoff, and K. Moffat. 1995. Optical studies of a bacterial photoreceptor protein, photoactive yellow protein, in single crystals. *Biochemistry.* 34:879–890.

Oka, T., N. Yagi, T. Fujisawa, H. Kamikubo, F. Tokunaga, and M. Kataoka. 2000. Time-resolved x-ray diffraction reveals multiple conformations in the M-N transition of the bacteriorhodopsin photocycle. *Proc. Natl. Acad. Sci. USA.* 97:14278–14282.

Oka, T., N. Yagi, F. Tokunaga, and M. Kataoka. 2002. Time-resolved x-ray diffraction reveals movement of F helix of D96N bacteriorhodopsin during M-MN Transition at neutral pH. *Biophys. J.* 82:2610–2616.

Perman, B., V. Srajer, Z. Ren, T. Teng, C. Pradervand, T. Ursby, D. Bourgeois, F. Schotte, M. Wulff, R. Kort, K. Hellingwerf, and K. Moffat. 1998. Energy transduction on the nanosecond time scale: Early structural events in a xanthopsin photocycle. *Science.* 279:1946–1950.

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

Ren, Z., B. Perman, V. Srajer, T. Y. Teng, C. Pradervand, D. Bourgeois, F. Schotte, T. Ursby, R. Kort, M. Wulff, and K. Moffat. 2001. Molecular movie from nanosecond to second time scales at 1.8 Å resolution displays the photocycle of photoactive yellow protein, a eubacterial blue-light receptor. *Biochemistry.* 40:13788–13801.

Romo, T. D., J. B. Clarage, D. C. Sorensen, and G. N. Phillips Jr. 1995. Automatic identification of discrete substates in proteins: singular value decomposition analysis of time-averaged crystallographic refinements. *Proteins.* 22:311–321.

Schlichting, I., and K. Chu. 2000. Trapping intermediates in the crystal: ligand binding to myoglobin. *Curr. Opin. Struct. Biol.* 10:744–752.

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

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

Ujj, L., S. Devanathan, T. E. Meyer, M. A. Cusanovich, G. Tollin, and G. H. Atkinson. 1998. New photocycle intermediates in the photoactive yellow protein from Ectothiorodospira halophila: picosecond transient absorption spectroscopy. *Biophys. J.* 75:406–412.

Van Brederode, M. E., W. D. Hoff, I. H. M. Van Stokkum, M. L. Groot, and K. Hellingwerf. 1996. Protein folding thermodynamics applied to the photocycle of the photoactive yellow protein. *Biophys. J.* 71:365–380.

Wang, B. C. 1985. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* 115:90–112.

Zimanyi, L., A. Kulcsar, J. K. Lanyi, D. F. Sears, Jr., and J. Saltiel. 1999a. Singular value decomposition with self modeling applied to determine bacteriorhodopsin intermediate spectra: analysis of simulated data. *Proc. Natl. Acad. Sci. USA.* 96:4408–4413.

Zimanyi, L., A. Kulcsar, J. K. Lanyi, D. F. Sears, Jr., and J. Saltiel. 1999b. Intermediate spectra and photocycle kinetics of the Asp96 -> asn mutant bacteriorhodopsin determined by singular value decomposition with self modeling. *Proc. Natl. Acad. Sci. USA.* 96:4414–4419.

Zimanyi, L., and J. K. Lanyi. 1993. Deriving the intermediate spectra and photocycle kinetics from time-resolved difference spectra of bacteriorhodopsin. *Biophys. J.* 64:240–251.