New software for the singular value decomposition of time-resolved crystallographic data

Yi Zhao and Marius Schmidt


Many research topics in condensed matter research, materials science and the life sciences make use of crystallographic methods to study crystalline and non-crystalline matter with neutrons, X-rays and electrons. Articles published in the *Journal of Applied Crystallography* focus on these methods and their use in identifying structural and diffusion-controlled phase transformations, structure–property relationships, structural changes of defects, interfaces and surfaces, etc. Developments of instrumentation and crystallographic apparatus, theory and interpretation, numerical analysis and other related subjects are also covered. The journal is the primary place where crystallographic computer program information is published.

Crystallography Journals Online is available from journals.iucr.org
New software for the singular value decomposition of time-resolved crystallographic data

Yi Zhao and Marius Schmidt*

Physics Department, University of Wisconsin–Milwaukee, 1900 E. Kenwood Boulevard, Milwaukee, Wisconsin 53211, USA. Correspondence e-mail: m-schmidt@uwm.edu

Singular value decomposition (SVD) has been successfully used in the analysis of time-resolved crystallographic data. A new software package for Linux-based operating systems, called SVD4TX, has been developed. In contrast to an earlier version of the SVD program, written in Fortran, the new program provides a GTK+-based graphical user interface for easy user guidance through the entire SVD process. New features, such as an improved, more stable routine to determine a compatible kinetic mechanism, are implemented in the SVD4TX program package so that it provides almost all the necessary tools for a semi-automatic and effective SVD-based analysis of time-resolved crystallographic data in one program package.

1. Introduction

Time-resolved crystallography (Moffat, 1989) is a unique technique for determining the structures of intermediates in biomolecular and chemical reactions. It genuinely unifies structure determination with chemical kinetics. A full description of a time-resolved crystallographic experiment would go beyond the scope of this article. The reader is referred to Schmidt (2008) and Schmidt, Ihee et al. (2005) for detailed reviews. In short, macromolecular time-resolved crystallography was pioneered in the last two decades of the previous century and came to maturity only recently (Ren et al., 1999).

Experiments are of the pump-probe type. A reaction is initiated in a crystal usually by an intense, ultra-short laser flash (the pump) and probed after a time delay \( \Delta t \) by an intense X-ray pulse (the probe) of typically 100 ps duration. Since it is impossible to rotate a crystal substantially during 100 ps, the reflection intensities are collected using still exposures. For this reason, a bandwidth of X-ray radiation is required to collect the integral reflection intensities. Hence, the Laue method is used. The crystal is reoriented to cover the entire reciprocal space (Rajagopal, Kostov & Moffat, 2004). At the end of the experiment, complete data sets of Laue (structure factor) amplitudes at each time delay plus their equivalent reference amplitudes collected from the same crystal in the dark are available. The time delays are preferentially arranged equidistantly in log-time to account for the exponential behavior of chemical kinetics. The data sets are scaled together and time-dependent data sets of weighted difference structure factor amplitudes are obtained (see below).

Using phases calculated from a very precise reference structure, difference structure factors are obtained, from which a time-series of difference electron density maps is calculated. The analysis of these four-dimensional data proceeds on the level of difference maps, since difference electron densities and concentrations of the intermediates are linearly dependent on each other (Henderson & Moffat, 1971).

With more and more powerful X-ray sources emerging, time-resolved crystallography will be routinely applied to characterize biologically and pharmaceutically interesting reactions in biomolecules. The simultaneous determination of the structure of reaction intermediates and the kinetics is critical to understanding the mechanism and the pathways by which a protein performs its biological reaction (Schlichting et al., 1990; Bolduc et al., 1995; Ren et al., 2001; Schmidt et al., 2004; Schmidt, Nienhaus et al., 2005; Rajagopal et al., 2005; Ihee et al., 2005; Key et al., 2007; Knapp et al., 2006). Direct structural information on these intermediates is difficult to obtain because they have short lifetimes and tend to overlap heavily in time (Hajdu et al., 2000; Moffat, 2001; Ren et al., 2001; Schlichting, 2003). The other difficulty is the low signal-to-noise ratio of time-resolved crystallographic data, caused by a low fraction of reacting molecules in the crystal. Signal may be difficult to differentiate from noise by simple visual inspection of the electron density maps. For these reasons it was largely unknown how to extract the structure information and separate it from the kinetics (Moffat, 2001). This problem was solved by applying singular value decomposition (SVD; Golub & Reinsch, 1970), a powerful technique for dealing with sets of equations or matrices, to crystallographic data (Schmidt et al., 2003). SVD may act as a noise filter, and it determines the main components found in a time-resolved crystallographic experiment.

SVD is a widely used technique (Henry & Hofrichter, 1992) for decomposing a matrix into several component matrices, exposing many of the useful and interesting properties of the original matrix (see also Hansen et al., 2006). The SVD method is based on the following theorem of linear algebra: Any \( m \times n \) matrix \( \mathbf{A} \) whose number of rows \( m \) is greater than or equal to its number of columns \( n \) can be written as the product of an \( m \times n \) column-orthogonal matrix \( \mathbf{U} \), an \( n \times n \) diagonal matrix \( \mathbf{S} \) with positive or zero elements (the singular values), and the transpose of an \( n \times n \) orthogonal matrix \( \mathbf{V} \):

\[
\mathbf{A} = \mathbf{USV}^T. \tag{1}
\]

The columns of \( \mathbf{U} \) and \( \mathbf{V} \) are called the left singular vectors and right singular vectors of \( \mathbf{A} \), respectively. The diagonal elements of \( \mathbf{S} \) are called the singular values of \( \mathbf{A} \).

In the analysis of time-resolved X-ray data, the SVD method separates time and space variables. From a series of time-dependent difference maps it determines only a few main common spatial components and their time variations, which constitute main common temporal components. The temporal components are then used to determine the number of intermediate states and relaxation times.
The relaxation times of the kinetics can be faithfully found in a global way in the right singular vectors. These relaxation times may be used to estimate a kinetic mechanism which can be fitted to the data. From such a fit, the pure difference electron density maps of the intermediates can be determined. Finally, the structures of the intermediates are modeled with the help of these difference maps.

The applicability of SVD to crystallographic data was initially demonstrated by Schmidt et al. (2003) using mock data. The SVD method was applied to calculated difference Fourier maps, simulating those to be obtained in a time-resolved crystallographic study of photoactive yellow protein. 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. The results showed that SVD allows for an unbiased differentiation between signal and noise, which laid the basis for a noise reduction procedure called SVD flattening; this ultimately led to noise-reduced and phased difference maps. These maps can be used to further identify and fit a kinetic mechanism, a prerequisite for extracting the difference electron densities and the structures of the intermediates. The SVD-based procedures were then used in multiple studies to analyze real time-resolved crystallographic data (Rajagopal, Schmidt et al., 2004; Rajagopal et al., 2005; Schmidt et al., 2005; Schmidt, Nienhaus et al., 2005; Schmidt, Ihee et al., 2005; Ihoe et al., 2005; Key et al., 2007).

For the SVD analysis of time-resolved X-ray data a number of programs needed to be used; the central program was called SVD4TX ('singular value decomposition for time-resolved X-ray data'; Schmidt et al., 2003). SVD4TX was written in Fortran and implemented the SVD algorithm and most procedures necessary for the kinetic analysis of the time-resolved crystallographic data. However, additional, stand-alone programs were used to perform basic operations such as calculation of weighted difference structure factors, Fourier transforms, quality factor analysis (Rajagopal, Schmidt et al., 2004) and phase recombination in SVD flattening. In addition, new routines had to be programmed and linked to SVD4TX for each candidate chemical kinetic mechanism to be tested; hence, substantial effort and expert knowledge was needed to perform an SVD analysis of crystallographic data.

Although SVD4TX can be used with any crystallographic data that depend on an additional parameter such as pH, temperature or dose, the program was mainly developed having time-resolved macromolecular crystallographic data in mind. Based on the existing program in Fortran, a new SVD4TX program package with a graphical user interface is described in this paper. The goals of the new program are to create a computational framework for the SVD analysis and to provide all necessary tools for the SVD analysis of time-resolved crystallographic data in one program package. Compared with the Fortran version, some new features, such as an efficient fast Fourier transform (FFT) routine, SVD flattening, quality factor analysis and a general routine to diagonalize coefficient matrices, were newly implemented; the fit routines were also improved and are more stable. Altogether, this amounts to a flexible, easy to use program for anyone who wishes to analyze crystallographic data by the singular value decomposition method. To demonstrate the ability of the new SVD4TX, we have used time series of mock data generated by the method described by Schmidt et al. (2003) with experimental noise included to simulate the experimental conditions as closely as possible.

2. Description of the new SVD4TX program package

The source code of the SVD4TX program is written in C for Linux-based operating systems. This new program package provides a GTK+ 2.0-based graphical user interface for easy user guidance. Fig. 1 shows the interface of SVD4TX. The goals of the SVD analysis are the determination of a chemical kinetic mechanism and the extraction of pure difference electron densities of the intermediates from the crystallographic data. The procedures to reach those goals are schematically shown in Fig. 2. In short, there are four major steps involved in the processing of the data: Step 1 involves the Fourier transformation of time-dependent structure factors to difference electron densities, from which the data matrix is constructed. The time information is then used in step 2 to reduce the noise in the difference maps by means of SVD flattening. In step 3, a number of exponentials are fitted to determine relaxation times in a global way. A candidate chemical kinetic mechanism is then selected in step 4 to extract the difference electron densities of the intermediates.

2.1. Format of input data

SVD4TX analyzes a time series of crystallographic data. In order to simplify the input process, an input file in text format must be prepared (Fig. 3), which is read by the program. Two types of data formats are now supported by the program: structure factors and difference electron density maps. At present, the FSFOUR map

![Figure 1](image1.png)

**Figure 1**
The interface of the software package SVD4TX which appears when the program is started.

![Figure 2](image2.png)

**Figure 2**
The procedures of SVD4TX. Four major steps are involved in the processing of the data to finally generate the time-independent difference maps of the intermediates: preparation of data, SVD and SVD flattening, fit of a sum of exponentials, and determination of a chemical kinetic mechanism.
format is supported for compatibility with the *XtalView*/Xfit molecular graphics suite (McRee, 1999). The location of time-dependent structure factor amplitude data and those of the corresponding initial (dark) state must be included in the input file.

The input file in Fig. 3 shows a typical time series that would be obtained from a time-resolved experiment. Four crystals were used to generate the time series from 1 μs to 100 ms. Five time-dependent data sets, \( \{ F(hkl, t) \} \), plus one reference data set (in the dark, \( \{ F^0(hkl) \} \) were measured per crystal. Each data set consists of a list of \( h, k, l \) Laue structure factor amplitudes plus their uncertainties (\( \sigma \)).

The reference data sets (ref1.hkl to ref4.hkl in this example) need to be scaled beforehand to a data set calculated from a precise initial (dark) state model. This brings them to the absolute scale. The time-dependent data sets should be scaled to their respective reference data. As a result, all data are on the absolute scale.

The file containing the calculated structure factors of the initial (dark) state including the phase has to be supplied (2phy_fc6.phs in Fig. 2). An atomic model used for mask building purposes is read in PDB file format (Protein Data Bank; Berman et al., 2000). In addition, the symmetry operators must be supplied. The program searches the keywords (see Fig. 3) to assign file paths and parameters. Difference maps can be used instead of structure factors. In this case, the location of the difference maps should be included following the keyword ‘MAP FILES’.

### 2.2. From structure factors to electron density

With the implemented FFT routine (Singleton, 1968), the time-dependent difference electron density maps are calculated from the set of time-dependent structure factor amplitudes \( \{ F(hkl, t) \} \), the corresponding structure factor amplitudes of the initial (dark) state \( \{ F^0(hkl) \} \) and the phases \( \phi^0 \) calculated from the known reference atomic model. The difference electron densities are calculated as

\[
\Delta \rho(t) = \frac{1}{V_e} \sum_{hkl} w |\Delta F(hkl, t)\| \exp(\text{i} \phi^0) \exp[-2\pi i (hx + ky + lz)],
\]

where \( x, y, z \) are the components of the position vector in the coordinate system of the unit cell (fractional coordinates), \( h, k, l \) are the reflection indices, \( V_e \) is the volume of the unit cell and \( w \) is a weighting factor for the difference structure factor amplitude \( \Delta F(hkl, t) \). The weighting of difference electron densities reduces the influence of outliers and inaccurately determined reflection intensities. The weighting factor is calculated according to Ren et al. (2001) as

\[
w = \frac{1}{1 + \Delta F^2 / |\Delta F|^2 + \sigma_{\Delta F}^2 / |\Delta F|^2},
\]

where \( \sigma_{\Delta F} \) is the variance of \( \Delta F \), and \( |\Delta F|^2 \) and \( |\Delta F|^2 \) are the mean values of \( \Delta F^2 \) and \( \sigma_{\Delta F}^2 \). For the calculation of electron density maps a few parameters such as the resolution limit, the grid numbers and the direction of the plane should be given in a dialog window (Fig. 4). To refine this process, SVD4TX divides the data into a few resolution shells (typically 5–10) and the weighting factors are calculated in the individual resolution shell. In this case, \( |\Delta F|^2 \) and \( |\Delta F|^2 \) are the mean values of \( \Delta F^2 \) and \( \sigma_{\Delta F}^2 \) in the individual resolution shells. The maximum resolution found in the data is suggested as the resolution limit in the dialog window, which can be changed manually. The suggested grid numbers can also be changed manually. The difference maps are then saved in FSFOUR format.

The structure factor amplitudes and phases will also be used later for the newly implemented procedure of SVD flattening. If, similar to the former version of SVD4TX, already existing time-dependent difference maps are imported to the program, the process of FFT may be omitted. This option is useful for a quick check. However, then, SVD flattening is not possible since it needs structure factors and \( \sigma \) values for phase recombination. It is, therefore, strongly recommended that the time-dependent Laue amplitudes with calculated phases are used for the entire SVD analysis.

### 2.3. Singular value decomposition

The SVD procedure separates time and space variables. The left singular vectors are difference maps; they are the main spatial components of the experimental, time-dependent difference maps. Each right singular vector contains the temporal variation of the corresponding left singular map, whereas the singular values weight the corresponding left singular map, whereas the singular values weight

![Figure 3](image.png)

**Figure 3**

An example of an input file. The keywords are described. The file paths of the structure factor data (or difference maps for a quick check), a PDB file to calculate a mask and the symmetry operators are needed.

![Figure 4](image.png)

**Figure 4**

Dialog for FFT options. The maximum resolution found in the data is suggested as resolution limit, which can be manually changed. The suggested number of grid points is based on Shannon sampling (Fung et al., 2009), implying voxels with edge lengths roughly half the maximum resolution limit.
the contribution of the components to the experimental difference maps.

To perform an SVD analysis, data matrix $A$ is prepared, where the difference maps from a time series are entered one by one and in temporal order as column vectors of the matrix. Since a unit cell in macromolecular crystallography can easily consist of half a million grid points, the number of grid points should be reduced to make the calculation faster. If crystallographic symmetry is present, only the asymmetric unit needs to be included. What is more, a mask can be calculated by the program using the PDB file to include only the volume occupied by protein atoms. A further reduction in the number of considered grid points is possible if those that do not contain significant difference electron densities throughout the time course are disregarded. For this purpose, the program allows the option of selecting grid points only if the positive difference electron density is above, or the negative difference electron density is below, a chosen $\sigma$ level for at least one time point; thus, the number of useful grid points can be reduced to around $10^3$–$10^5$ for moderately sized unit cells.

The SVD4TX package provides a dialog window in which one can determine SVD options such as an additional margin to the mask and select the $\sigma$ level to prepare matrix $A$, which will then decompose to the left and right singular vectors and the singular values.

Not all of the singular vectors contain signal. The number of significant singular vectors is related to the number of intermediates in the reaction (Henry & Hofrichter, 1992; Schmidt, Ihee et al., 2005; Schmidt, 2008). However, the structure information available to crystallography adds powerful methods to further guide this discrimination. First, the left singular vectors are difference maps and can be displayed on a graphics screen together with suitable atomic models, for example, that of the reference state. Signal is faithfully detected (Schmidt et al., 2003). Secondly, the difference electron density in regions of large signal can be used to automatically determine a quality factor (Rajagopal, Schmidt et al., 2004; see below), and the discrimination can be based on this. In addition, the autocorrelation of the right singular vectors can be employed for an estimate. The method of rotation (Henry & Hofrichter, 1992; Schmidt et al., 2003) is implemented in the new SVD4TX to re-collect the signal from the less significant singular vectors into the more significant. Nevertheless, whichever method is used, a set of significant singular values and vectors is the result. As an example, Fig. 5 shows a set of significant right singular vectors.

Using only significant singular values and vectors to reconstruct the data matrix $A$ and ignoring the insignificant singular vectors containing only noise and no signal, the signal-to-noise ratio in the reconstructed difference maps is improved. The data matrix $A$ can then be approximated in a least-squares sense by matrix $A'$:

$$U'SV^T = A' \approx A. \quad (4)$$

This new, noise-reduced data matrix $A'$ constitutes the basis for the SVD flattening.

2.4. SVD flattening

SVD flattening (Schmidt et al., 2004; Schmidt, Ihee et al., 2005; Schmidt, 2008), a new procedure to use time information for noise reduction, is now implemented in the SVD4TX program package. From the noise-reduced difference maps, which are reconstructed with only the significant singular values and vectors, difference structure factors with amplitude $\Delta F_{SVD}$ and phase $\phi_{SVD}^n$ are obtained by an inverse Fourier transform. These structure factors are combined with the structure factors of the dark state by a phase recombination scheme to obtain improved structure factors. The procedure of phase recombination is shown schematically in Fig. 6. First, we divide the data into a small number of resolution shells, each of which contains roughly the same number of data. The following procedure is performed for each resolution shell. The phase of time-dependent, measured Laue amplitudes, $\phi_n$, is approximately equal to the phase of the vector sum of $\Delta F_{SVD}$ and the calculated structure factor of the dark state $F_{C}$ (see Fig. 6). The amplitude of the calculated structure factor $F_{C}$ is usually not equal to that measured from the crystal in the dark, $F_{C}$, because of crystal-to-crystal differences. To correct for this and to allow for the information of the precise reference (dark state) model to enter the analysis, we calculated singular values and vectors are shown in the figure with solid spheres, squares and triangles, respectively. The fitted exponential curves are shown as solid lines.

**Figure 5** Right singular vectors from the SVD analysis. Their amplitudes are weighted by the square of their corresponding singular value. The first three (significant) right singular vectors are shown in the figure with solid spheres, squares and triangles, respectively. The fitted exponential curves are shown as solid lines.

**Figure 6** Argand diagram for phase recombination in SVD flattening. $F_{C}$: reference structure factor (amplitude and phase) calculated from a precise reference (dark state) model. $F_{C}$: observed reference (dark) structure factor amplitude having the same phase as $F_{C}$; $\Delta F_{SVD}$: difference between structure factor amplitude measured as a reference and that calculated from the reference model; allows to correct for crystal-to-crystal differences and to include information from the precise dark state model. $\Delta F_{SVD}$, $\phi_{SVD}^n$: difference structure factor amplitude and phase obtained by Fourier inversion of the noise-reduced time-dependent difference map reconstructed by using only the most significant singular values and vectors. $F_{C}$: measured time-dependent (Laue) structure factor amplitude; its direction (phase) is given by $F_{C} + \Delta F_{SVD}$; $\Delta F_{C-O}$ is added to $F_{C}$ to bring it to the scale of the calculated reference structure factors, $\sigma_{F/C}$: uncertainty (noise) of $F_{C}$ determined by the data reduction software. If the noise is small relative to the overall noise, $\Delta F_{SVD}$ with phase $\phi_{SVD}^n$ will point to the tip of $F_{C}$, otherwise it will point to the tip of $\Delta F_{SVD}$. Here, an intermediate situation is pictured. New, noise-improved difference electron density maps are calculated from $\Delta F_{SVD}$. 

Zhao and Schmidt • SVD4TX 737

In the case of low relative noise, improvement of the phases of the SVD-flattened structure factors is relative noise is very large, then from these reactions can be determined by globally fitting a sum of exponential terms featuring common relaxation times to the significant right singular vectors:

\[ v_n(t) = \sum_{j=1}^{J} A_{nj} \exp(-t/\tau_j). \] (7)

where \( v_n \) is the \( n \)th significant right singular vector and \( J \) is the number of exponential functions found globally (Fig. 5). The pre-exponentials \( A_{nj} \) and the relaxation times \( \tau_j \) are the fit parameters.

In the program package SVD4TX, a more stable fit routine that minimizes the sum of the squares of \( m \) nonlinear functions in \( n \) variables by a modification of the Levenberg–Marquardt algorithm (Garbow et al., 1980) is employed. Initial values of relaxation times can be roughly estimated from the plot of right singular vectors and

<table>
<thead>
<tr>
<th>Quality factors from the left singular vectors from a time series of 21 time-dependent mock data sets.</th>
<th>Quality factor 1</th>
<th>Quality factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left singular vectors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.57</td>
<td>2.13</td>
</tr>
<tr>
<td>2</td>
<td>1.82</td>
<td>1.47</td>
</tr>
<tr>
<td>3</td>
<td>2.99</td>
<td>1.43</td>
</tr>
<tr>
<td>4</td>
<td>1.38</td>
<td>0.93</td>
</tr>
<tr>
<td>5</td>
<td>0.87</td>
<td>0.79</td>
</tr>
<tr>
<td>6</td>
<td>0.97</td>
<td>0.66</td>
</tr>
<tr>
<td>7</td>
<td>0.83</td>
<td>0.77</td>
</tr>
<tr>
<td>8</td>
<td>1.36</td>
<td>1.09</td>
</tr>
<tr>
<td>9</td>
<td>1.21</td>
<td>0.69</td>
</tr>
<tr>
<td>10</td>
<td>1.10</td>
<td>0.69</td>
</tr>
<tr>
<td>11</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td>12</td>
<td>0.98</td>
<td>0.87</td>
</tr>
<tr>
<td>13</td>
<td>1.17</td>
<td>0.89</td>
</tr>
<tr>
<td>14</td>
<td>1.16</td>
<td>0.62</td>
</tr>
<tr>
<td>15</td>
<td>0.89</td>
<td>0.72</td>
</tr>
<tr>
<td>16</td>
<td>0.84</td>
<td>0.82</td>
</tr>
<tr>
<td>17</td>
<td>1.14</td>
<td>1.06</td>
</tr>
<tr>
<td>18</td>
<td>1.01</td>
<td>0.95</td>
</tr>
<tr>
<td>19</td>
<td>1.00</td>
<td>0.93</td>
</tr>
<tr>
<td>20</td>
<td>0.83</td>
<td>0.75</td>
</tr>
<tr>
<td>21</td>
<td>1.30</td>
<td>0.64</td>
</tr>
</tbody>
</table>

\[ \Delta F_{c-o} \] by subtracting the observed dark state Laue amplitude |\( F_{o} | \) from |\( F_{c} | \). The addition of \( \Delta F_{c-o} \) to |\( F_{c} | \) corrects for differences that influence both the dark state Laue amplitudes and the time-dependent Laue amplitudes. The amplitude of an improved structure factor |\( F_{\text{new}} | \) is calculated (see below), which takes into account the relative noise, which is the uncertainty of |\( F_{c} | \), \( \sigma_{F_{c}} \), relative to the mean uncertainty found in that resolution shell, \( \langle \sigma_{F_{c}} \rangle \). For this a weighting factor \( p_{\text{res}} \) is calculated, which accounts for the relative noise:

\[ p_{\text{res}} = \frac{1}{1 + \sigma_{F_{c}}^2/\langle \sigma_{F_{c}} \rangle^2}. \] (5)

\( p_{\text{res}} \) is then used to calculate the amplitude |\( F_{\text{new}} | \):

\[ |F_{\text{new}}| = p_{\text{res}}(|F_{c}| + |\Delta F_{c-o}|) + (1 - p_{\text{res}})|F_{c} + \Delta F_{\text{SVD}}|. \] (6)

In the case of low relative noise, |\( F_{\text{new}} | \) will be |\( F_{c} | + |\Delta F_{c-o} | \). If the relative noise is very large, then |\( F_{\text{new}} | \) will be \( |F_{c} + \Delta F_{\text{SVD}}| \). In Fig. 6 an intermediate situation is sketched, which results in improved difference structure factors \( \Delta F_{\text{SVD}} \) (amplitude and phase). New difference electron density maps with improved signal-to-noise ratio are then obtained by Fourier transforming these \( \Delta F_{\text{SVD}} \). The improvement of the phases of the SVD-flattened structure factors is of the order of 10–15\(^\circ\) depending on the noise level (Schmidt et al., 2003). The SVD-flattened difference maps can then be subjected to SVD flattening again until convergence is reached. These maps are much better suited for interpretation and further analysis. Fig. 7 shows a difference map of a single time point before and after SVD flattening.

2.5. Quality factor analysis

The quality factor analysis is a new feature of the program SVD4TX. It can be used to estimate the signal-to-noise ratio of individual difference maps, and it has been applied to analyze experimental time-resolved crystallographic data by Rajagopal and co-workers (Rajagopal, Schmidt et al., 2004; Rajagopal et al., 2005) and by Ihee et al. (2005). The quality factor is evaluated using left singular vectors, which are difference maps. The quality factor is defined here as the ratio of the mean absolute difference electron density in a spherical region of diameter of 2–3 Å within a signal region to the σ value of the whole map. SVD4TX looks for two regions with the largest signals separated by at least 5 Å, integrates the absolute difference electron densities in the mentioned spherical region and calculates quality factors. A larger quality factor means more significant signal in the map. Only maps with the highest quality factors are significant and will be selected for further analysis. The quality factor analysis allows for an unbiased, automatic approach to evaluating the significance of the left singular vectors. An example of the calculated quality factors of left singular vectors is shown in Table 1. It is obvious from these values that the first three left singular vectors have significantly high quality factors and therefore high signal-to-noise ratio. In combination with the other criteria mentioned above and below, one can faithfully determine that for each example here the first three singular vectors and values are significant, and it is with these that the analysis proceeds.

2.6. Kinetic mechanisms from the SVD

SVD4TX can fit kinetic mechanisms consisting of first-order reactions to the data. As mentioned above, the right singular vectors show the temporal variation of the corresponding left singular vectors. The time courses in the right singular vectors are linear combinations of the time-dependent concentrations of intermediates. For first-order reactions, the concentrations obey exponential functions (Steinfeld et al., 1989). Therefore, the relaxation times \( \tau_j \) of these reactions can be determined by globally fitting a sum of exponential terms featuring common relaxation times to the significant right singular vectors:

\[ v_n(t) = \sum_{j=1}^{J} A_{nj} \exp(-t/\tau_j). \] (7)
coefficients are \( k \) and \( I \) end mechanism are selected. These mechanisms each employ three rates \( I_0) = 75 \text{ s} \). The fitted rate coefficients \( k \) ([equation (7)]). By varying the rate coefficients \( k \) in the coefficient matrix, and therefore the eigenvalues of the coefficient matrix, the relaxation times are determined in the previous step, the program offers a dialog window to select a candidate mechanism by checking the boxes with the corresponding rate coefficients \( k \) from a general mechanism that employs the \( J \) states, plus the initial (dark) state, as shown in Fig. 8. For the selected candidate mechanism, the program suggests the initial values of rate coefficients \( k \); no expert user knowledge will be required. With the selected candidate mechanism, a coefficient matrix, and therefore the eigenvalues of the coefficient matrix, the program determines the best fitting rate coefficients. Once the rate coefficients are obtained, the concentration profiles for each intermediate \( j \), \( I(k, t) \) for the selected candidate mechanism can be calculated. As examples, the sequential mechanism and dead end mechanism are selected. These mechanisms each employ three intermediates, \( I_1 \), \( I_2 \) and \( I_5 \), plus the initial state, \( I_0 \). The fitted rate coefficients are \( k(I_1 \rightarrow I_2) = 35000 \text{ s}^{-1}, k(I_2 \rightarrow I_3) = 1200 \text{ s}^{-1}, k(I_3 \rightarrow I_0) = 75 \text{ s}^{-1} \) for the sequential mechanism and \( k(I_1 \rightarrow I_2) = 35000 \text{ s}^{-1}, k(I_2 \rightarrow I_3) = 700 \text{ s}^{-1}, k(I_3 \rightarrow I_0) = 350 \text{ s}^{-1}, k(I_0 \rightarrow I_5) = 280 \text{ s}^{-1} \) for the dead end mechanism, respectively. From this, the concentrations of the intermediates are calculated. They are shown in Fig. 9. It should be mentioned at this point that multiple mechanisms can fit the data equally well (Schmidt et al., 2004). Especially for more complex kinetic schemes, additional information obtained, for example, from time-resolved spectroscopy (De la Mora-Rey & Wilmot, 2007) may be required to distinguish between the degenerated mechanisms. Although all intermediates that populate significantly can be extracted from the crystallographic data, a unique kinetic mechanism is needed to finally decide whether these intermediates are on or off the catalytic path.

2.7. Calculation of time-independent electron density data of intermediates

During the fit of the concentrations of intermediates to the right singular vectors, the linear scale factors \( E_{nj} \) were determined:

\[
s_j^2 \nu(t)_n^{obs} \simeq \nu(t)_n^{fit} = s_j^2 \sum_k E_{nj}(k, t),
\]

where \( \nu(t)_n^{obs} \) and \( \nu(t)_n^{fit} \) are the \( n \)th observed and fitted right singular vectors, respectively. The fit is weighted by the square of the corresponding singular value \( s_j^2 \). The scale factor \( E_{nj} \) determines the

Figure 8
Example for the selection of a candidate mechanism from a reaction employing three intermediates plus the initial (dark) state. The rate coefficients are those that would appear in a coefficient matrix describing this reaction after instantaneous initiation. Here a sequential mechanism is selected by checking the boxes with the corresponding rate coefficients.

Figure 9
Fractional concentrations of intermediates (c_{annot}) for the simplified mechanisms as a function of time \( t \); sequential (a) and dead end (b) mechanisms employing three intermediates, \( I_1, I_2 \) and \( I_5 \), plus the initial (dark) state, \( I_0 \). Solid, dashed and dotted lines: time-dependent concentrations of \( I_1, I_2 \) and \( I_5 \), respectively. Dashed-dotted line: concentration of the reference state \( I_r \).

Figure 10
The time-independent difference electron density of the third intermediate. (a) Original, noise- and error-free, phased difference electron density of intermediate 3. This intermediate was used together with two other intermediates to calculate the time-dependent mock data. (b) Time-independent difference map of intermediate 3, which has been extracted from the time course of mock data containing about four times the experimental noise. Red and white: negative difference electron density on the \( -3\sigma/4\sigma \) level; blue and cyan: positive difference electron density on the \( 3\sigma/5\sigma \) level, respectively. The atomic model of this intermediate (in green) as well as that of the reference dark state (in yellow) is overlaid to guide the eye. HC4 and Arg52: hydroxyxicamnamic acid and arginine 52 moieties were used to simulate prominent structural changes.
The software package SVD4TX with graphical user interface was developed to perform an analysis of time-resolved crystallographic data based on singular value decomposition almost instantly in a user friendly way. The new SVD4TX covers almost all the necessary tools for SVD analysis of time-resolved crystallographic data. One of the new features of SVD4TX is that the time-dependent difference electron density maps can be calculated from a set of time-dependent structure factor amplitudes obtained from a time-resolved crystallographic experiment, corresponding data set of reference amplitudes and phases obtained from the known dark state atomic model. Quality factor analysis provides the possibility to select the most significant left singular vectors by estimating the signal-to-noise ratio of individual maps. A sophisticated SVD-flattening procedure is available within the same software package. A general approach to solve a series of differential equations to calculate the concentrations of the intermediates is implemented. This approach diagonalizes the coefficient matrix that describes a kinetic mechanism of first-order reactions. A fit to the data that uses this approach is more stable and yields faithful results. The structured design of the program allows for easy implementation of the new features of this program and of the new kinetic mechanism with an arbitrarily large number of intermediates.

This work was supported in part by the College of Letters and Sciences of the University of Wisconsin–Milwaukee, USA, and by the Deutsche Forschungsgemeinschaft, SFB 533 and grant Schm 1423/2-1.

References