Biophysical *Journal*
Computational Tool

Biophysical Society

Tracking Equilibrium and Nonequilibrium Shifts in Data with TREND

Jia Xu[1] and Steven R. Van Doren[1,*]
[1]Department of Biochemistry, University of Missouri, Columbia, Missouri

ABSTRACT Principal component analysis (PCA) discovers patterns in multivariate data that include spectra, microscopy, and other biophysical measurements. Direct application of PCA to crowded spectra, images, and movies (without selecting peaks or features) was shown recently to identify their equilibrium or temporal changes. To enable the community to utilize these capabilities with a wide range of measurements, we have developed multiplatform software named TREND to Track Equilibrium and Nonequilibrium population shifts among two-dimensional Data frames. TREND can also carry this out by independent component analysis. We highlight a few examples of finding concurrent processes. TREND extracts dual phases of binding to two sites directly from the NMR spectra of the titrations. In a cardiac movie from magnetic resonance imaging, TREND resolves principal components (PCs) representing breathing and the cardiac cycle. TREND can also reconstruct the series of measurements from selected PCs, as illustrated for a biphasic, NMR-detected titration and the cardiac MRI movie. Fidelity of reconstruction of series of NMR spectra or images requires more PCs than needed to plot the largest population shifts. TREND reads spectra from many spectroscopies in the most common formats (JCAMP-DX and NMR) and multiple movie formats. The TREND package thus provides convenient tools to resolve the processes recorded by diverse biophysical methods.

INTRODUCTION

Plotting the course of biomolecular or physiological processes typically uses procedures specific to the field. In the case of spectroscopy and imaging, tracking the process can be laborious because of the steps of assigning the peaks of the spectra or features in the images, manually choosing peaks or image features subjectively judged optimal for monitoring the process of interest, and managing complications from any concurrent processes. Spectral overlap and peak broadening (e.g., from chemical exchange in NMR) can prevent correct fitting (1). A more elegant alternative to such efforts is to apply unsupervised, multivariate statistical pattern recognition such as principal component analysis (PCA). PCA has provided insight from series of measurements from diverse techniques of molecular biophysics that include magnetic resonance, vibrational, optical, and dichroic spectroscopies; x-ray scattering and diffraction; mass spectrometry; calorimetry; hydrodynamics; atomic force microscopy; electron microscopy; and imaging by fluorescence, Raman or light scattering, as well as functional magnetic resonance imaging (Table S1 in the Supporting Material and references therein). Biophysical studies have often used PCA to determine dependencies of various reactions upon time, concentration, or other conditions, e.g., in protein folding (Table S1). PCA applied directly to spectra, images, and movies appears to be a convenient and general way to determine the main trends of change among measurement frames that record many localized changes. PCA is much more accommodating of many data distributions than is often appreciated (2). It transforms many measured variables to far fewer and uncorrelated principal components (PCs) that each capture part of the trends of covariation among measured variables (2).

PCA of NMR peak lists was used to track equilibrium transitions of proteins due to pH (3) and binding of partners (4–6). Closely related singular value decomposition (SVD) of NMR peak pick lists was used to reconstruct filtered basis spectra for use in fitting biphasic ligand binding (7) or for identifying binding sites (8). The applications of PCA were recently extended directly to NMR spectra, images, and movies without choosing any peaks or features for analysis. Moreover, applying PCA directly to NMR spectra makes binding isotherms easily accessible in all chemical exchange regimes, including intermediate exchange where severe broadening and nonlinearity of peak shifts ordinarily mask the true course of molecular

Submitted June 29, 2016, and accepted for publication December 9, 2016.
*Correspondence: vandorens@missouri.edu
Editor: Jeff Peng
http://dx.doi.org/10.1016/j.bpj.2016.12.018

association (1,9). Application of SVD to series of *time*-dependent two-dimensional (2D) images or spectra extracted the dominant time course as PC1. The approach also detected multiple time-evolving processes in magnetic resonance imaging (MRI) movies as PCs. Similarly, when two sequential steps of binding were monitored by NMR, PCA detected both binding steps and the intermediate state with a single ligand bound (9). Accomplishments with SVD (PCA) have usually been limited to the laboratories that wrote task-specific code to perform the calculations, however.

Independent component analysis (ICA) can complement PCA. ICA aims instead to find *independent* components (ICs) (2). The quest of ICA for statistical independence is more demanding than PCA's aim of correlation coefficients of zero. These objectives are equivalent for Gaussian (normal) distributions. ICA can be regarded as more general than PCA and is effective for non-Gaussian data and situations where PCA fails (2). However, ICA can be very slow to compute compared with PCA, lower in convergence, and require repeated calculations. Like PCA, ICA has been used to reduce dimensionality and filter or separate data in processing signals, images (10), large biological data sets (11), and NMR spectra of mixtures (12–14).

To make these capabilities available for application to a variety of spectroscopic and imaging techniques used in biophysics, we have developed a software package named TREND (Tracking and Resolving Equilibrium and Nonequilibrium population shifts in Data). Its main means of tracking the shifts is PCA implemented with SVD. Its secondary means is an ICA algorithm, which recapitulates the PCA results we examined, provided the correct number of ICs is specified. We first sought to extract binding isotherms, equilibrium shifts, and time courses (all potentially with multiple components) from series of NMR spectra. Because of the suitability of PCA for many other kinds of series of 2D digital data frames, we utilized the Python community's support of file I/O in multiple data formats (e.g., movies and spreadsheets) and wrote code for additional spectroscopic formats, enabling wide application (e.g., JCAMP-DX, Sparky peak list). For example, we analyzed a cardiac MRI movie (15) with TREND to isolate multiple aspects of the cardiac cycle and to reconstruct movies from combinations of PCs. This software package can resolve biologically relevant reactions and processes with relative ease from many biophysical sources of complicated spectral and imaging data.

MATERIALS AND METHODS

Implementation of TREND

TREND was written in Python 2.7 and calls NumPy for linear algebra and random number generation. It implements PCA (SVD) with function calls to NumPy. We first wrote TREND for operation at the command line. We added a graphical user interface (GUI), supported by Gooey, using function calls to wxPython. Most users will prefer to use the GUI to operate TREND. TREND comprises three programs, each with both interfaces (Table 1). The executable files *trendmaingui* and *trendmain* compute the PCs or ICs across the 2D series of measurements, create temporary files used by the plotting or reconstruction programs run afterward, and plot the first three components plus benchmarks of their significance. *Trendplotgui* and *trendplot* provide optional plotting that is customizable in terms of the number and choice of normalization of the components. Optional reconstructions of the measurement series are available from *trendreconstructgui* and *trendreconstruct* (Table 1). Explanations of the flags and parameters for the command line versions are available online in the manual for TREND (https://trendmizzou.gitbooks.io/trend-manual/content/). For convenience of installation, we packaged TREND and the public domain software it depends upon using PyInstaller. Consequently, TREND does not need Python on the host system. Distributions are available for Windows 7 and later, Mac OS X 10.7 and later, and these versions of Linux: Ubuntu 14.04/Fedora 23, Ubuntu 16.04, and Red Hat 7.1/CentOS 7 (http://biochem.missouri.edu/trend).

Conversion of a stack of 2D measurements into a matrix for analysis

A wide variety of 2D measurements can be read and analyzed by *trendmaingui* and *trendmain*. This includes images or movie frames comprising pixels, one-dimensional (1D) and 2D spectra from many spectroscopies, lists of peak positions and heights, and unprocessed NMR spectroscopic data in the time domain (free induction decays, FIDs; Fig. 1). The program reads NMR spectra in NMRpipe, Sparky, and Bruker Topspin formats (as well as FIDs in NMRpipe and Topspin formats) (Table 2) using code from Nmrglue (16). *Trendmaingui* and *trendmain* also read Agilent (Varian) VNMRJ format and JCAMP-DX formats of Bruker, Agilent, and Jeol spectrometers. To analyze the measurements from many other kinds of spectroscopy and biophysical measurements (Table S1), the program reads the most common JCAMP-DX formats, as well as spreadsheet and text formats commonly written by instruments (Table 2). *Trendmaingui* and *trendmain* read NMR peak lists either in the format of Sparky peak lists (17) or plain text files, before converting them into column vectors (3,7). Movies are read in multiple formats (i.e., avi, mov, mp4, ogv, webm) by the MoviePy module into three-dimensional (3D) arrays with color layers. *Trendmaingui* and *trendmain* convert the movie frames to gray-scale (8-bit depth) and rearrange them into 2D matrices (Fig. 2). Time-lapse series of PNG images are read, using the Scipy module of Python, and handled similarly.

In the case of NMR data, spectra very recently emerged as probably the preferred format for application of PCA (9). In the examples below, NMR

TABLE 1 Executable File Components of TREND

Executable File	Roles	Interface
trendmain.exe	preprocess and compute PCs or ICs	CLI
trendmaingui.exe[a]	preprocess and compute PCs or ICs	GUI
trendplot.exe	plot selected PCs or ICs with choice of normalization	CLI
trendplotgui.exe[a]	plot selected PCs or ICs with choice of normalization	GUI
trendreconstruct.exe	reconstruct spectra, images, or movies from PCs	CLI
trendreconstructgui.exe[a]	reconstruct spectra, images, or movies from PCs	GUI

CLI is command-line interface; and GUI is graphical user interface.
[a]Executable files with GUI are trendmaingui.app, trendplotgui.app, and trendreconstructgui.app for OS X or macOS platforms.

spectra (collected with a uniform set of parameters) were processed with NMRPipe (18) and converted to the UCSF format of Sparky (17,19). NMR spectra in UCSF format were read by *trendmaingui* or *trendmain* for conversion into 2D matrices (Fig. 2). Unprocessed NMR data in the time domain (FIDs) can also be read, processed, and the solvent signal subtracted. (Analysis of time domain data is justified by Parseval's theorem regarding the equivalency of signals in the time and frequency domains (20)).

Preprocessing

Regardless of original data format, columns from each 2D measurement read are positioned end-to-end into a single 1D vector for convenience (9) (Fig. 2). These 1D columns are arrayed over the experimental variable (concentration, pH, time, etc.) into the data matrix $\boldsymbol{X}$, which has F1 × F2 points in the column dimension and n points per row for the n experimental conditions. To expedite manipulations of this matrix $\boldsymbol{X}$ and facilitate calculations on a modest laptop computer, each vector is compressed by deleting unchanging positions, resulting in matrix $\boldsymbol{X}'$ (Fig. 2). For SVD of spectra, the user is encouraged to use a threshold that is three- to sevenfold the noise level to filter out low intensity regions of the spectra, which compresses matrix $\boldsymbol{X}'$ further. However, it is better to use a lower threshold where intermediate exchange broadening significantly weakens NMR peaks.

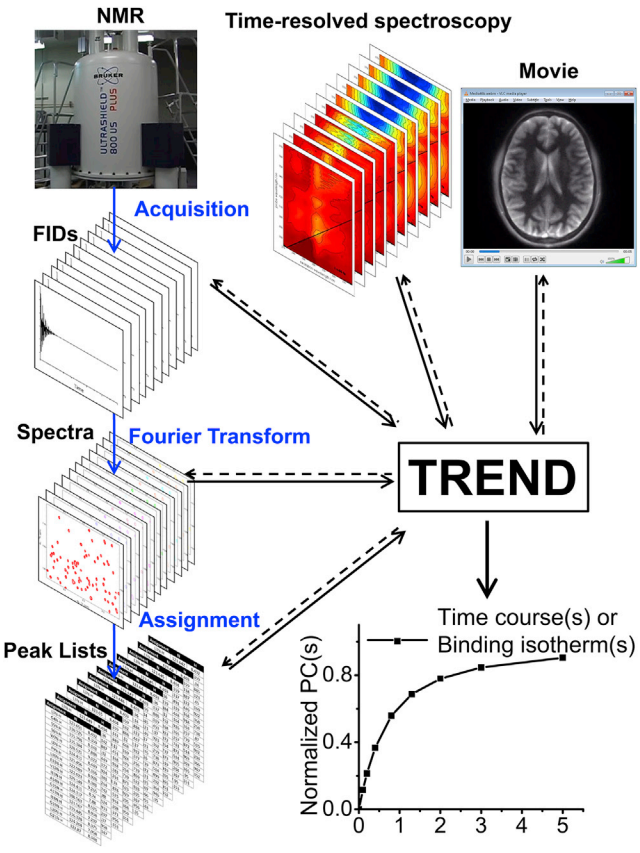

FIGURE 1 Workflows to TREND extraction of the main components of change across a series of measurements. TREND reads a series of spectra, peak lists, FIDs, images, or a movie and extracts the dominant trends from them. Spectral or electronic data series in general-purpose JCAMP-DX, text, or spreadsheet formats, NMR formats, and video formats are readable. The dashed lines signify reconstruction of measured data from the principal components chosen. To see this figure in color, go online.

As required by PCA and ICA algorithms, the rows of compressed matrix $\boldsymbol{X}'$ are centered and then optionally scaled. Scaling enlarges weaker signals relative to large signals. The options for scaling methods include *autoscaling*, *Pareto scaling*, or *no scaling* (21). *No scaling* appears acceptable in most titrations, but *autoscaling* generally enhances fits to the binding isotherms. *Autoscaling* obviates the systematic scaling of ^{15}N NMR peak shifts down by several-fold relative to ^{1}H shifts that were used in PCA of lists in (3). *Autoscaling* also generalizes to ^{1}H-^{13}C correlation spectra. *Pareto scaling* is recommended for NMR titrations with substantial intermediate exchange broadening (9). *Range*, *vast*, and *level* scaling (21) are also implemented in *trendmain* but do not work well with NMR spectra. *No scaling* has been used for MRI movies. Column centering and scaling are not necessary in our experience, but are available in *trendmaingui* and *trendmain* as they are sometimes used for PCA (22) and ICA (11). The descriptions of the data scaling and centering methods (21) are listed in Table S2.

Calculating principal components via SVD

The compressed, preprocessed matrix $\boldsymbol{X}'$ has m points per column and n points or experimental conditions in each row, with $m > n$. $\boldsymbol{X}'$ can be decomposed into three matrices as follows:

$$\boldsymbol{X}'_{mn} = \boldsymbol{U}_{mn}\boldsymbol{S}_{nn}\boldsymbol{V}^{\boldsymbol{T}}_{nn}, \tag{1}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}^{\boldsymbol{T}}$ are orthogonal matrices and $\boldsymbol{S}$ is a diagonal matrix that contains the square roots of eigenvalues for vectors in $\boldsymbol{U}$ or $\boldsymbol{V}$ in descending order. To obtain the trends of change across the measurements, we are interested in $\boldsymbol{V}^{\boldsymbol{T}}$, whose rows span $\boldsymbol{X}'$ and are called the right singular vectors. The $\boldsymbol{V}^{\boldsymbol{T}}$ matrix has row vectors $\boldsymbol{V}^{\boldsymbol{T}}_{\boldsymbol{nn}} = (\boldsymbol{V}^{\boldsymbol{T}}_1, \boldsymbol{V}^{\boldsymbol{T}}_2, \boldsymbol{V}^{\boldsymbol{T}}_3 \ldots \boldsymbol{V}^{\boldsymbol{T}}_{\boldsymbol{n}})$. Importantly, the first row in the $\boldsymbol{V}^{\boldsymbol{T}}$ matrix is PC1 and the second row PC2, i.e., the two largest trends of change among the series of spectra or images measured. To obtain these PCs that record the relationships among columns in $\boldsymbol{X}'$ (Fig. 2), it suffices to calculate $\boldsymbol{V}^{\boldsymbol{T}}$. The rows of $\boldsymbol{V}^{\boldsymbol{T}}$ are orthonormal eigenvectors of the symmetric matrix $\boldsymbol{X}'^{\boldsymbol{T}}\boldsymbol{X}'$ (Fig. 2). (The normalized form of $\boldsymbol{X}'^{\boldsymbol{T}}\boldsymbol{X}'$ is equivalent to the covariance matrix, the alternative algorithm for computing PCA (2).) The normalized PC1 values from the first row of $\boldsymbol{V}^{\boldsymbol{T}}$ indicate the fractional population of the main change at each measurement in the series of measurements. When obtained from a typical titration of ligand binding, PC1 represents the binding isotherm; a dissociation constant may be fitted to it (9).

Reconstruction of spectra, images, or movies by PCA

The reconstructed data set $\boldsymbol{X}_{\boldsymbol{reconst}}$, with size of $m \times n$, can be calculated as follows:

$$\begin{aligned}\boldsymbol{X}_{\boldsymbol{reconst}} = {} & \boldsymbol{U}_a\left(\boldsymbol{S}_a\boldsymbol{V}^{\boldsymbol{T}}_a\right) + \boldsymbol{U}_a\left(\boldsymbol{S}_b\boldsymbol{V}^{\boldsymbol{T}}_b\right) + \boldsymbol{U}_c\left(\boldsymbol{S}_c\boldsymbol{V}^{\boldsymbol{T}}_c\right) \\ & + \boldsymbol{U}_d\left(\boldsymbol{S}_d\boldsymbol{V}^{\boldsymbol{T}}_d\right) + \boldsymbol{U}_e\left(\boldsymbol{S}_e\boldsymbol{V}^{\boldsymbol{T}}_e\right)\ldots,\end{aligned} \tag{2}$$

where a, b, c, d, e … refer to the index of PCs generated by *trendmain* or *trendmaingui* to use in the reconstruction by *trendreconstruct* or *trendreconstructgui*. (Note a, b, c, d, e … can be nonconsecutive integers. To enable this, the "reconst" box should be selected in *trendmaingui*. When using *trendmain*, the –reconst flag should be included.) The $\boldsymbol{U}$ matrix is used for the reconstruction. It can be rewritten as column vectors: $\boldsymbol{U}_{\boldsymbol{mn}} = (\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3 \ldots \boldsymbol{U}_{\boldsymbol{n}})$, which lie in the column space of $\boldsymbol{X}'$. $\boldsymbol{U}$ can be calculated similarly to $\boldsymbol{V}^{\boldsymbol{T}}$, by solving eigenvectors of the matrix $\boldsymbol{X}'\boldsymbol{X}'^{\boldsymbol{T}}$. To recover the original 2D data series, the preprocessing steps of centering, scaling, and compression (filtering) can be reversed as described in the manual for TREND. The user can choose to reconstruct the centered and scaled matrix, matrix $\boldsymbol{X}'$, or matrix $\boldsymbol{X}$ in the format of the original data (Fig. 2).

TABLE 2 File Formats Read and Reconstructed by TREND[a]

Choice in Trendmaingui	Format	Reconstruction Support	Comment
NMR Data Formats			
fid	NMRPipe FID	yes	
ft2	NMRPipe Ft2	yes	
ucsf	Sparky UCSF	yes	
brukerfid	Bruker Topspin FID	yes	fid, ser in/1/pdata/subfolder[b]
brukerft2	Bruker Topspin spectra	yes	1r, 2rr files
agilentfid	VnmrJ, OpenVnmrJ FID	yes	fid
agilentspectra	VnmrJ, OpenVnmrJ spectra	no	Phasefile[a]
sparkylist	Sparky peak list	yes	duplicate peaks not allowed
JCAMP-DX (Joint Committee on Atomic and Molecular Physical data—Data Exchange format)			
jcamp	JCAMP-DX	no	Only supports X..(Y+Y) and (XY..XY)[c]
Text File Formats			
txt	floating point	yes	for series of text files
complextxt	complex numbers	yes	for series of text files
singletxt	complex or floating point	yes	for single .TXT file containing entire series
Spreadsheet Formats			
csv	comma-separated floating point	no	for series of .CSV files
complexcsv	comma-separated complex numbers	no	for series of .CSV files
singlecsv	comma-separated complex or floating point	no	for single .CSV file containing entire series
excel	Excel format	no	for series of Excel files
singleexcel	Excel format with tabs	no	for single file with single or multiple tabs
Images and Movies			
png	images in PNG format	yes	For series of .PNG files
movie	common video formats	yes	.ogv, .mp4, .mpeg, .avi, .mov, .webm

[a]See the online TREND manual (https://trendmizzou.gitbooks.io/trend-manual/content/).
[b]Currently the processed spectra must be saved by setting processed directory to 1.
[c]JCAMP-DX is a general format for exchanging and archiving data from many instruments, including but not limited to infrared (IR), Raman, ultraviolet-visible (UV-Vis), fluorescence, NMR, and electron paramagnetic resonance (EPR). The data stored in JCAMP-DX files can be spectral plots, contours, or peak tables. TREND supports the most common JCAMP-DX formats. The digital data in JCAMP-DX can be AFFN (ASCII FREE FORMAT NUMERIC) form or ASDF (ASCII SQUEEZED DIFFERENCE FORM). TREND supports decoding compressed data, including PAC, SQZ, DIF, SQZDUP, and DIFDUP. Two most common tabular data forms, (X++(Y..Y)) and (XY..XY) are supported. TREND reads a series of JCAMP-DX files, or a single JCAMP-DX file with one or multiple blocks. TREND supports NTUPLE format (introduced by JCAMP-DX 5.0), which is designed for multidimensional techniques with data sets with multiple variables. For example, JCAMP-DX NMR uses NTUPLE to show mixed real/imaginary FID data sets. See format details in http://www.jcamp-dx.org/, https://badc.nerc.ac.uk/help/formats/jcamp_dx/, and http://wwwchem.uwimona.edu.jm:1104/spectra/testdata/index.html.

ICA calculations

ICA is available in TREND and implemented using scikit-learn (http://scikit-learn.org/stable/modules/decomposition.html#ica). Despite the potential generality of ICA, two limitations need to be respected. Since the magnitudes of ICs cannot be determined, their contributions cannot be ranked. ICA is also prone to local minima during optimization, requiring comparisons of repeated calculations (10,11). TREND implements the FastICA algorithm for computational efficiency. FastICA preprocesses data by PCA to reduce dimensions and avoid overlearning (23–25). (Overlearning is an underdetermined situation that interferes in obtaining parameters and introduces artifacts to ICs (24,25)).

ICA decomposes the data matrix $\boldsymbol{X}$ as follows:

$$\boldsymbol{X} = \boldsymbol{AS}, \tag{3}$$

where $\boldsymbol{A}$ is the unknown mixing matrix that is invertible, square, and mixes the components in $\boldsymbol{X}$, and $\boldsymbol{S}$ is the matrix containing underlying independent sources. The aim of ICA is to solve for the mixing matrix $\boldsymbol{A}$ because it contains the ICs that may contain the meaningful trends sought. However, $\boldsymbol{A}$ and $\boldsymbol{S}$ both being unknown makes ICA calculations challenging (10). The equation can be rewritten as follows:

$$\boldsymbol{S} = \boldsymbol{WX} = \boldsymbol{VX_w}, \tag{4}$$

where $\boldsymbol{W}$ is the unmixing matrix that is calculated as $\boldsymbol{A}^{-1}$. To simplify and improve convergence of ICA, $\boldsymbol{X}$ is preprocessed to remove correlations and to normalize it, a process called whitening, which generates $\boldsymbol{X_w}$. FastICA implements this whitening step using PCA to calculate the whitened data matrix $\boldsymbol{X_w}$ as follows:

$$\boldsymbol{X_w} = \left(\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{E}^{\boldsymbol{T}}\right)\boldsymbol{X} \tag{5}$$

Where $\boldsymbol{E}$ is the matrix whose columns are normalized eigenvectors of the covariance matrix of $\boldsymbol{XX^T}$, and $\boldsymbol{D}$ is the diagonal matrix of the corresponding eigenvalues. The preprocessing with PCA also removes noise and reduces dimensions for ICA. The whitening simplifies the ICA problem to finding the unknown rotation matrix $\boldsymbol{V}$ that is defined as $\boldsymbol{D}^{-(1/2)}\boldsymbol{E}^{\boldsymbol{T}}$. In FastICA, $\boldsymbol{V}$ is estimated by maximizing non-Gaussian character. The equations lead to the following:

$$\boldsymbol{A} = \boldsymbol{W}^{-1} = \left(\boldsymbol{V}\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{E}^{\boldsymbol{T}}\right)^{-1} \tag{6}$$

RESULTS AND DISCUSSION

Workflows of TREND

For wide application of SVD or ICA to diverse series of 2D measurements, we wrote TREND in Python to read and analyze multiple types of data. These include diverse

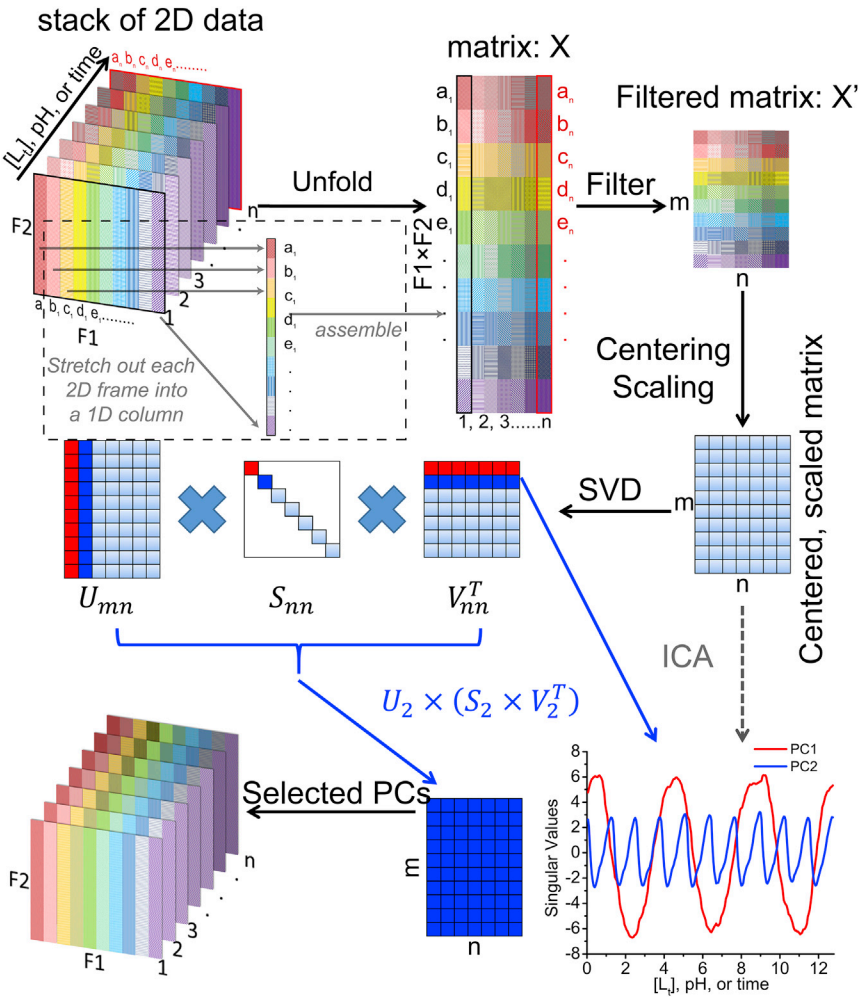

FIGURE 2 TREND implementation of PCA (SVD), ICA, and reconstruction. The algorithm reorganizes a series of 2D measurements as a series of 1D columns. The dashed box indicates the unfolding of the first 2D data (*marked by black edges*) with F2 rows and F1 columns to a long 1D column with F1 × F2 points. The data matrix $\boldsymbol{X}$ is compressed to $\boldsymbol{X}'$ and used for SVD. (No user selection of spectral peaks or image features is involved). The resulting first several right singular vectors in the decomposed matrix $\boldsymbol{V}^T$ represented the largest trend(s). ICA can be used to corroborate SVD results. Single or multiple PCs can be used to reconstruct the original data series. To see this figure in color, go online.

spectra, images, movies, or lists in text or spreadsheet formats available from many modern instruments (Fig. 1). The spectral formats include widely used JCAMP-DX standards and NMR formats. TREND can also apply PCA or ICA to a single 2D data matrix read in from a text file, spreadsheet file, or multiblock JCAMP-DX file containing multiple spectra (Table 2). The algorithm of the *trendmain* and *trendmaingui* executable files includes the following steps:

1) Convert each 2D measurement into a 1D vector arrayed by the experimental condition varied, in the data matrix $\boldsymbol{X}$.
2) Preprocess $\boldsymbol{X}$ with compression to $\boldsymbol{X}'$ and optional scaling.
3) Perform streamlined SVD or ICA to identify components (PC1, PC2, … or IC1, IC2, …) representing the major trend(s) varying with the experimental variable (Fig. 1).

The TREND package provides additional executable files for plotting the course of selected PCs or ICs or for rebuilding spectra, images, or movie from selected PCs (Table 1). For convenience, the plotting and reconstruction routines read temporary files just created by *trendmain* or *trendmaingui*; this frees the user from specifying input files, which is optional. The user may operate and customize these computations by a choice of GUI or command line arguments described in documentation for the software.

Although we used the Python routine NumPy to implement PCA (SVD) and scikit-learn (http://scikit-learn.org/stable/modules/decomposition.html#ica) to implement ICA calculations, corresponding routines are available in R (https://mran.microsoft.com/packages/), MATLAB (The MathWorks, Natick, MA, https://www.mathworks.com/matlabcentral/fileexchange/38300-pca-and-ica-package), and the MATLAB Statistics Toolbox. Recreating the workflows and functions depicted in Figs. 1 and 2 in an R or MATLAB environment would require code to parse the file formats of interest, reduce their dimensionality (i.e., "unfold" them), preprocess for readiness for the SVD or ICA routine, and interpret or reconstruct the results in the appropriate format. TREND spares the user this effort with a package that is user-friendly for NMR and other measurements from a variety of instrumentation, including spectroscopies and imaging; see Table 2 for data

formats handled. TREND is free for academics, avoiding the cost of licensing MATLAB. TREND requires <200 MB of disk space whereas the MATLAB environment occupies 2 to 3 GB. TREND is portable and its installation lacks dependencies, other than the need for Internet access upon first usage.

We present examples of uses of TREND that illustrate 1) its performance in resolving two or more processes, which is nonroutine by conventional means, and 2) its wide applicability to trace and reconstruct concurrent, complex transformations recorded by biophysical means such as spectra or imaging.

Examples of ligand binding to two sites detected by NMR

Antecedents to TREND's direct application of PCA to spectra and images were previous PCA studies of NMR peak lists. SVD was used to filter noise out of the lists, in turn used to reconstruct clean basis spectra to resolve three pH transitions (3) or two binding events (4,7). With TREND we demonstrate a direct spectrum-driven approach to the latter examples of two biphasic associations. Fig. 3 *A* plots a two-site binding scheme, where P and L denote [protein] and [ligand], respectively. K_{D1} and K_{D2} are dissociation constants from site 1 and 2. PL_{n1} and PL_{n2} are intermediates with ligand at site 1 or 2, where $n1$ and $n2$ indicate the numbers of ligand molecules that bind cooperatively to site 1 and 2, respectively. $PL_{n1}L_{n2}$ stands for the fully bound state. Equations 3 to 6 from (7) were used to simulate populations of species from the two-site binding scheme in a series of ^{15}N HSQC spectra (Fig. 3 *B*) using methods given in Supporting Material. The curvature in the simulated shifts of several peaks (red arrows in Fig. 3 *B*) accompanies more than one mode of binding (7). PCA on the peak lists (chemical shifts) captures two smooth components, PC1 and PC2 (purple in Fig. 3 *C*), contributing 90% and 6% of the variance, respectively. The PC1 and PC2 components of the peak lists were recreated using *trendreconstruct*. PC1 captures from the curved trajectories of peak movements the main linear paths of change (Fig. S1, *A* and *B*). PC2 identifies the peak shifts orthogonal to PC1 (Fig. S1 *C*). Computing PC1 and PC2 instead directly from the simulated HSQC spectra using *trendmain* (*green in* Fig. 3 *C*)

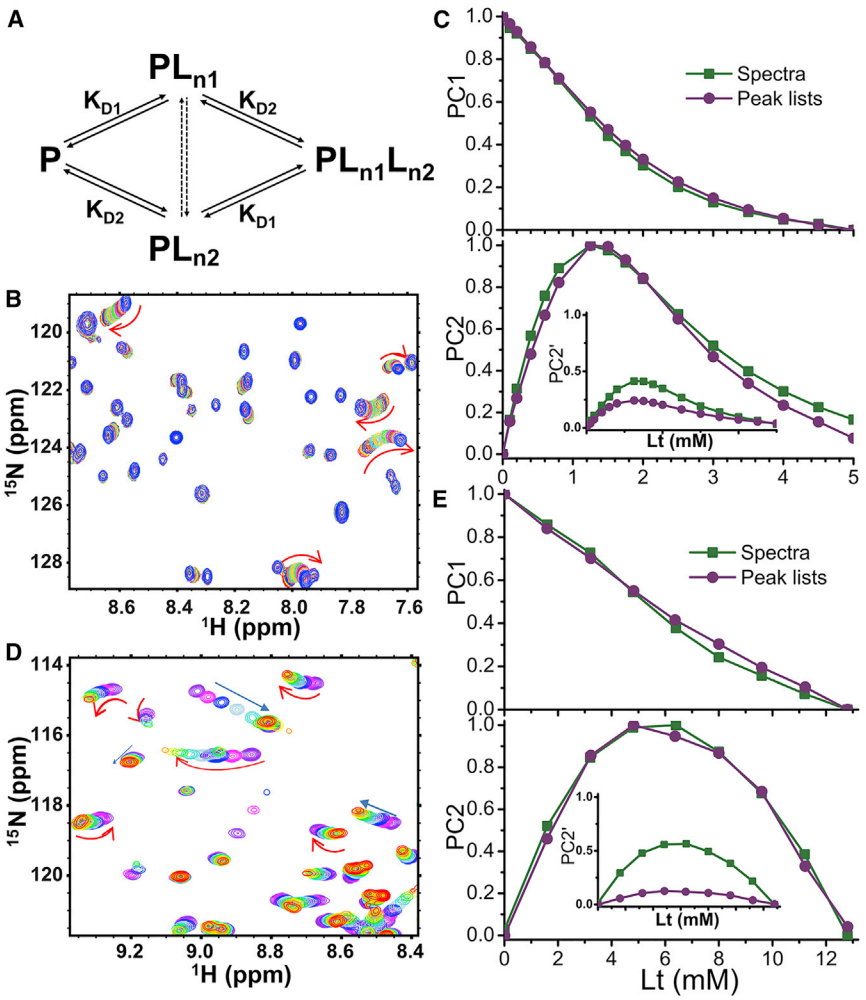

FIGURE 3 TREND identifies two components in titrations of two binding sites. (*A*) Scheme of the two-site binding model is shown. The number of ligands that bind to sites 1 and 2 are $n1$ and $n2$, respectively. (*B*) ^{15}N HSQC spectra simulated according to the two-site model, as described in Supporting Material, are plotted for ligand:protein ratios of 0, 0.1, 0.2, 0.4, 0.6, 0.8, 1.25, 1.5, 1.75, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0 with different contour colors. Both $n1$ and $n2$ equal 1 (7). In (*B*) and (*D*), linear and curved peak shifting are indicated by linear and curve arrows, respectively. (*C*) compares PCs from the spectra and peak lists from (*B*). PC1 and PC2 are normalized to the maximum amplitude of each. In the inset, each PC2 is instead normalized by the amplitude of PC1, which is symbolized by the PC2′ labeling of the ordinate. (*D*) Measured ^{15}N HSQC spectra of β-lactoglobulin titrated with ANS additions of 0, 1.6, 3.2, 4.8, 6.4, 8.0, 9.6, 11.2, and 12.8 mM (4) are plotted with different contour colors; while $n1 \approx 2$, $n2 \approx 1$ (4). (*E*) compares PCs from the spectra and peak lists of (*D*). The circles plot the PCA results from peak lists reported by (4). The plots are normalized and labeled as in (*C*). To see this figure in color, go online.

reproduces their counterparts extracted from peak lists very well, although each component contributes much less of the variance (38% and 15%, respectively). However, when PC2 values are normalized by PC1, there is a systematic difference in amplitude of PC2 consistent with its percentage of the variances listed above (*inset in* Fig. 3 *C*). This simulated two-site binding example and a number of 1:1 ligand-binding examples (2,9) suggest that normalized PCs extracted from lists of picked peaks in the fast-exchange regime can be reproduced well by applying PCA to the series of spectra. However, PC1 and PC2 extracted by TREND from the FIDs from the simulated two-site binding example are skewed with sigmoidal deviation from the PCs obtained from either the peak lists or spectra (Fig. S2, *A* and *B*). In the investigation of a titration of β-lactoglobulin with 1-anilinoaphthalene-8-sulfonate (ANS), Konuma et al. resolved two binding components using PCA of the assigned peaks from the NMR spectra of the titration (4). They observed curved trajectories (*red arrows in* Fig. 3 *D*) and linear trajectories (*blue arrows in* Fig. 3 *D*), suggesting the presence of multiple binding sites. Fast exchange behavior supported reliable PCA of the chemical shift data in peak lists, which provided binding isotherms (4). TREND extracted the PCs from the spectra (*green in* Fig. 3 *E*) and unprocessed FIDs from the titration (*green in* Fig. S2, *A* and *B*). These PCs are compared with the previously reported binding isotherms (*purple in* Fig. 3 *E*). The binding populations of (4) are reproduced well by the normalized PC1 and PC2 derived from the spectra despite the t1-noise present (Fig. S3), and less well by PC1 and PC2 obtained from the FIDs. (The residual solvent signal was subtracted on-resonance from the FIDs using the *trendmaingui* option of a convolution difference window (26). In cases of especially poor solvent suppression, this subtraction might not be enough for reliable PCs.) When choosing the form of NMR data to analyze, application of TREND directly to spectra appears to be the most consistently accurate.

Reconstruction of the spectra of the ANS titration with *trendreconstructgui* using *only* PC1 and PC2 introduces artifacts that are ghosts of the peaks from each spectrum of the titration (not shown). The cumulative contribution ratio (reported by *trendmaingui*) saturates at eight PCs, suggesting eight is sufficient to represent the series of spectra. Using eight PCs in the reconstruction removed the ghosts of peaks and reproduced well the spectra and their biphasic trajectories of peak shifts upon additions of ANS (Fig. S3). The need for eight or more PCs is typical of the need for faithful reconstruction of series of spectra and images. Nonlinearity is typical of such series and spreads their variances across many PCs; see Fig. S7 in (9). This spreading of variances to many PCs could account for the need for many PCs for faithful reconstruction. Inspection of the reconstructed and original spectra finds both fast and fast-intermediate exchange regimes (Fig. S3). Application of PCA directly to the spectra, followed by reconstruction, accommodated this mixture of behaviors, as recently proposed (9).

ICA for confirming components

TREND supports optional use of ICA. If the number or significance of PCs obtained comes into question, ICA can be used to test the significance and validity of the PCs. It is also conceivable that ICA may be able to resolve components from some experiments that are not resolvable by PCA. ICA of peak pick *lists* from the two-site binding example of Fig. 3 *B* yields ICs equivalent to PC1 and PC2 (Fig. S4). We tested ICA with various numbers K of trial components with series of *spectra* containing N true components. When $K \leq N$, ICA derives components that are very similar to those from PCA (Fig. S5). However when $K > N$, which means trying to extract more "independent components" than true components, ICA always fails in our experience, as evident from components that are jagged and meaningless (Fig. S5, *E* and *F*). Consequently, we propose that this failure of ICA can be used to count the meaningful components. The ICA should be repeated with incrementally higher K trial components. The lowest value of K at which ICA fails implies $K - 1$ significant components (see Fig. S5 for two examples of the iterative process). The drawback of ICA validation of components is in repeating FastICA calculations $N + 1$ times for each trial number of components, preferably with three to five repetitions of each, to escape local minima. Though the process is repetitive, it requires no previous knowledge of the number of components. Deciding the PCs that are significant may be quicker by identifying the PCs that contribute the most to scree plots (the convention) and which have large autocorrelation coefficients (smoothness) (7). However, recapitulation of PCs by ICs may engender more confidence in the reproducibility of the analysis.

Cardiac MRI movie resolved into components

Real-time imaging by MRI generates complex movies that are suitable to showcase the capabilities of TREND. An MRI movie of a slice through the four chambers of the heart (15; http://www.biomednmr.mpg.de/images/stories/movies/Media18.ogv) was analyzed by TREND. A movie for each of the first four individual PCs was reconstructed using *trendreconstructgui*, aiding interpretation of the PCs. PC1 follows the time course of breathing where the trough represents inhalation (Fig. 4 *A*; Movie S1). Fig. 4 *B* plots a frame from the PC2 movie (Movie S2) where the left ventricle is relaxed and open, known as diastole. Fig. 4 *C* plots a frame from the PC2 movie where the left ventricle and heart overall are contracted in systole. The time course of PC2 follows the alternation between the crests representing diastole and narrow troughs representing systole (Fig. 4 *A*). In the crests of PC2, the phases

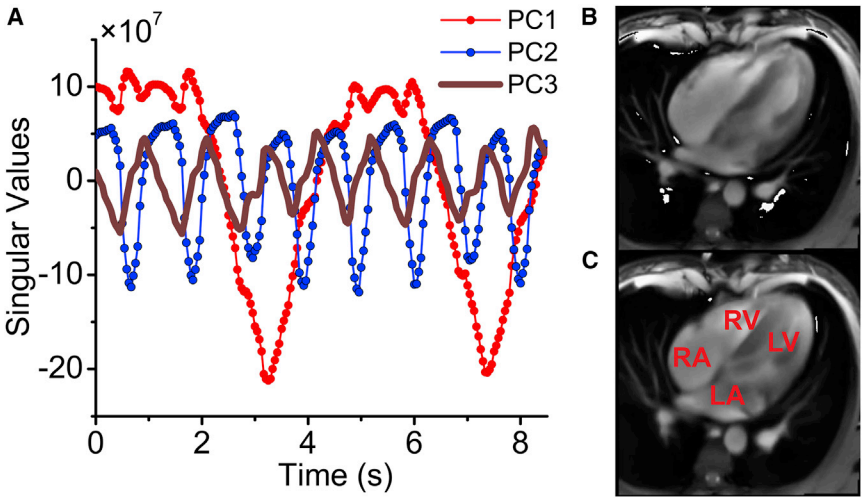

FIGURE 4 SVD captures from a cardiac MRI movie the time courses of breathing, diastole, and systole from a "four-chamber" angle of view. (*A*) PC1 represents respiration is shown. PC2 tracks the oscillation between diastole and systole. (*B*) This frame in the reconstructed PC2 movie is during diastole with the open cavities especially evident in the ventricles (Movie S2). (*C*) In this frame from the same movie, all four chambers are contracted (systole). RA is right atrium; RV is right ventricle; LV is left ventricle; and LA is left atrium. To see this figure in color, go online.

of rapid filling and subsequent slower filling of the ventricles can be observed. (An overview of the cardiac cycle is provided by Cardiovascular Physiology Concepts, Indianapolis, IN, http://www.cvphysiology.com/Heart%20Disease/HD002b.htm). The troughs of PC3 coincide with the isovolumetric contraction phase that begins systole (Fig. 4 *A*). The left ventricle and atrium walls and interiors alternate in appearance in the PC3 movie (Movie S3). Bright density between the left ventricle and atrium in the PC3 movie at the troughs in the PC3 time course suggests the closed state of the left atrioventricular (mitral) valve. Coinciding with this is detectable rotation of the right atrium and ventricle. The PC4 movie represents sudden overall rotations of the heart (Movie S4). The time courses indicate synchronization of these rotations (PC4) with both the cardiac cycle (PC2) and each inspiration of a breath (PC1); see Fig. S6 *A*. The rotations appear largest when a breath in begins and ends. These observations illustrate the ability of TREND to resolve and aid interpretation of concurrent processes.

A movie reconstructed from all four of these PCs using *trendreconstructgui* captures the major morphological changes of the cardiac cycle (Movie S5), but is not as smooth and nuanced as the original (15; http://www.biomednmr.mpg.de/images/stories/movies/Media18.ogv). *Trendmaingui* reports autocorrelation coefficients exceeding 0.7 for the first 44 PCs, suggesting their information content. Inspection of the scree plot and the cumulative contribution plot generated by *trendmaingui* indicates that the first four PCs account for ~69% of the statistical variance across the movie, 10 PCs account for 85%, and 20 account for 93% (Fig. S6 *B*). Reconstruction of the cardiac MRI movie using the first 10 PCs imparts much increased realism to the depiction of the turbulent blood flow in the cardiac chambers and smoothness to the cardiac movement (Movie S6). Doubling the PCs to the first 20 enhances the fidelity further but more subtly (Movie S7). Omission of PC1 removes the largest background of breathing changes to the chest cavity, while preserving the cardiac cycle portrayal (Movie S8).

In reconstruction of other movies and NMR spectra, we also observed the faithfulness of the reconstruction to increase with number of PCs. Eight or more PCs may often be desirable for satisfying reconstruction of a measurement series. The scree plot and secondarily the autocorrelation coefficients appear useful for anticipating the number of PCs beneficial for reconstruction of the measurement series.

CONCLUSIONS

Direct application of PCA (or ICA) to 2D measurements using TREND will expand the accessibility of equilibrium and time-evolving processes measured by spectra and imaging. No curation, selection, assignment, or resolution of specific spectral peaks or image features is necessary using this unsupervised statistical approach. TREND can be applied "on-the-fly" on an instrument host computer during data collection to assess if the process or reaction has progressed far enough. Multiple concurrent processes, measured by biophysical techniques, have been readily resolved into principal or independent components. Movies and spectra can be reconstructed with TREND from the user's choice of principal components. These capabilities will introduce, to our knowledge, new convenience and insight to analyses of spectrally detected reactions and imaging-detected processes studied by biophysics, physiology, and other disciplines.

SUPPORTING MATERIAL

Supporting Materials and Methods, two tables, six figures, and eight movies are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(16)34321-1.

AUTHOR CONTRIBUTIONS

J.X. wrote the code. J.X. and S.R.V.D. designed the approach, performed the research, analyzed the results, and wrote the manuscript.

ACKNOWLEDGMENTS

We are grateful to K. Sakurai, T. Konuma, and Y. Goto for spectra of the ANS titration of β-lactoglobulin; J. Frahm and his group for real-time MRI movies; M. D. Stanley for setting up the TREND website; A. G. Roberts, K. Stiers, and reviewers for beta-testing; and Y. Fulcher for discussion of PCA.

The work was supported by NSF grant MCB1409898.

Access to the TREND licensing (free for academics) and software downloads is available at http://biochem.missouri.edu/trend and https://nmrbox.org/.

SUPPORTING CITATIONS

References 27–58 appear in the Supporting Material.

REFERENCES

1. Williamson, M. P. 2013. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* 73:1–16.
2. Jolliffe, I. T. 2002. Principal Component Analysis. Springer-Verlag, New York.
3. Sakurai, K., and Y. Goto. 2007. Principal component analysis of the pH-dependent conformational transitions of bovine beta-lactoglobulin monitored by heteronuclear NMR. *Proc. Natl. Acad. Sci. USA.* 104:15346–15351.
4. Konuma, T., Y. H. Lee, …, K. Sakurai. 2013. Principal component analysis of chemical shift perturbation data of a multiple-ligand-binding system for elucidation of respective binding mechanism. *Proteins.* 81:107–118.
5. Majumder, S., C. M. DeMott, …, A. Shekhtman. 2014. Using singular value decomposition to characterize protein-protein interactions by in-cell NMR spectroscopy. *ChemBioChem.* 15:929–933.
6. Cembran, A., J. Kim, …, G. Veglia. 2014. NMR mapping of protein conformational landscapes using coordinated behavior of chemical shifts upon ligand binding. *Phys. Chem. Chem. Phys.* 16:6508–6518.
7. Arai, M., J. C. Ferreon, and P. E. Wright. 2012. Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *J. Am. Chem. Soc.* 134:3792–3803.
8. Cobbert, J. D., C. DeMott, …, A. Shekhtman. 2015. Caught in action: selecting peptide aptamers against intrinsically disordered proteins in live cells. *Sci. Rep.* 5:9402.
9. Xu, J., and S. R. Van Doren. 2016. Binding isotherms and time courses readily from magnetic resonance. *Anal. Chem.* 88:8172–8178.
10. Shlens, J. "A Tutorial on Independent Component Analysis." Preprint, submitted April 11, 2014. arXiv:1404.2986.
11. Yao, F., J. Coquery, and K.-A. Lê Cao. 2012. Independent principal component analysis for biologically meaningful dimension reduction of large biological data sets. *BMC Bioinformatics.* 13:24.
12. Nuzillard, D., S. Bourg, and J. Nuzillard. 1998. Model-free analysis of mixtures by NMR using blind source separation. *J. Magn. Reson.* 133:358–363.
13. Ladroue, C., F. A. Howe, …, A. R. Tate. 2003. Independent component analysis for automated decomposition of in vivo magnetic resonance spectra. *Magn. Reson. Med.* 50:697–703.
14. Monakhova, Y. B., A. M. Tsikin, …, S. P. Mushtakova. 2014. Independent component analysis (ICA) algorithms for improved spectral deconvolution of overlapped signals in 1H NMR analysis: application to foods and related products. *Magn. Reson. Chem.* 52:231–240.
15. Zhang, S., A. A. Joseph, …, J. Frahm. 2014. Real-time magnetic resonance imaging of cardiac function and flow-recent progress. *Quant. Imaging Med. Surg.* 4:313–329.
16. Helmus, J. J., and C. P. Jaroniec. 2013. Nmrglue: an open source Python package for the analysis of multidimensional NMR data. *J. Biomol. NMR.* 55:355–367.
17. Goddard, T. D., and D. G. Kneller. 2000. SPARKY. University of California, San Francisco, San Francisco.
18. Delaglio, F., S. Grzesiek, …, A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6:277–293.
19. Lee, W., M. Tonelli, and J. L. Markley. 2015. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics.* 31:1325–1327.
20. Cavanagh, J., W. J. Fairbrother, …, N. J. Skelton. 2007. Preface to the First Edition. *In* Protein NMR Spectroscopy, 2nd. J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, and N. J. Skelton, editors. Academic Press, Burlington, VT, pp. vii–x.
21. van den Berg, R. A., H. C. Hoefsloot, …, M. J. van der Werf. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics.* 7:142.
22. Selvaratnam, R., S. Chowdhury, …, G. Melacini. 2011. Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc. Natl. Acad. Sci. USA.* 108:6133–6138.
23. Hyvärinen, A., and E. Oja. 2000. Independent component analysis: algorithms and applications. *Neural Netw.* 13:411–430.
24. Hyvärinen, A., J. Karhunen, and E. Oja. 2002. Practical considerations. *In* Independent Component Analysis. John Wiley, New York, pp. 269–271.
25. Särelä, J., and R. Vigario. 2003. Overlearning in marginal distribution-based ICA: analysis and solutions. *J. Mach. Learn. Res.* 4:1447–1469.
26. Marion, D., M. Ikura, and A. Bax. 1989. Improved solvent suppression in one-and two-dimensional NMR spectra by convolution of time-domain data. *J. Magn. Reson.* 84:425–430.
27. Gualfetti, P. J., O. Bilsel, and C. R. Matthews. 1999. The progressive development of structure and stability during the equilibrium folding of the alpha subunit of tryptophan synthase from *Escherichia coli*. *Protein Sci.* 8:1623–1635.
28. Rüther, A., M. Pfeifer, …, S. Lüdeke. 2014. Reaction monitoring using mid-infrared laser-based vibrational circular dichroism. *Chirality.* 26:490–496.
29. Kakitani, Y., R. Fujii, …, A. Angerhofer. 2006. Triplet-state conformational changes in 15-cis-spheroidene bound to the reaction center from *Rhodobacter sphaeroides* 2.4.1 as revealed by time-resolved EPR spectroscopy: strengthened hypothetical mechanism of triplet-energy dissipation. *Biochemistry.* 45:2053–2062.
30. Kim-Shapiro, D. B., S. B. King, …, S. K. Ballas. 1998. Time resolved absorption study of the reaction of hydroxyurea with sickle cell hemoglobin. *Biochim. Biophys. Acta.* 1380:64–74.
31. Isin, E. M., and F. P. Guengerich. 2007. Multiple sequential steps involved in the binding of inhibitors to cytochrome P450 3A4. *J. Biol. Chem.* 282:6863–6874.
32. Frank, G. A., M. Goomanovsky, …, G. Haran. 2010. Out-of-equilibrium conformational cycling of GroEL under saturating ATP concentrations. *Proc. Natl. Acad. Sci. USA.* 107:6270–6274.
33. Shapiro, D. B., R. M. Esquerra, …, D. S. Kliger. 1996. A study of the mechanisms of slow religation to sickle cell hemoglobin polymers following laser photolysis. *J. Mol. Biol.* 259:947–956.
34. Esquerra, R. M., R. A. Goldbeck, …, D. S. Kliger. 1998. Spectroscopic evidence for nanosecond protein relaxation after photodissociation of myoglobin-CO. *Biochemistry.* 37:17527–17536.

35. Hendler, R. W., S. K. Bose, and R. I. Shrager. 1993. Multiwavelength analysis of the kinetics of reduction of cytochrome aa3 by cytochrome c. *Biophys. J.* 65:1307–1317.

36. Chung, H. S., M. Khalil, and A. Tokmakoff. 2004. Nonlinear infrared spectroscopy of protein conformational change during thermal unfolding. *J. Phys. Chem. B.* 108:15332–15342.

37. Uy, D., and A. E. O'Neill. 2005. Principal component analysis of Raman spectra from phosphorus-poisoned automotive exhaust-gas catalysts. *J. Raman Spectrosc.* 36:988–995.

38. Zapata, A. L., M. R. Kumar, …, P. J. Farmer. 2013. A singular value decomposition approach for kinetic analysis of reactions of HNO with myoglobin. *J. Inorg. Biochem.* 118:171–178.

39. Hendriks, J., and K. J. Hellingwerf. 2009. pH dependence of the photoactive yellow protein photocycle recovery reaction reveals a new late photocycle intermediate with a deprotonated chromophore. *J. Biol. Chem.* 284:5277–5288.

40. Martínez, J. C., N. A. Chequer, …, T. Cordova. 2012. Alternative metodology for gold nanoparticles diameter characterization using PCA technique and UV-Vis spectrophotometry. *Nanosci. Nanotech.* 2:184–189.

41. Wasserman, S. R., P. G. Allen, …, N. M. Edelstein. 1999. EXAFS and principal component analysis: a new shell game. *J. Synchrotron Radiat.* 6:284–286.

42. Kalinin, S. V., B. J. Rodriguez, …, Z.-G. Ye. 2009. Spatial distribution of relaxation behavior on the surface of a ferroelectric relaxor in the ergodic phase. *Appl. Phys. Lett.* 95:142902.

43. Lichtert, S., and J. Verbeeck. 2013. Statistical consequences of applying a PCA noise filter on EELS spectrum images. *Ultramicroscopy.* 125:35–42.

44. Seo, J., Y. An, …, C. Choi. 2016. Principal component analysis of dynamic fluorescence images for diagnosis of diabetic vasculopathy. *J. Biomed. Opt.* 21:46003.

45. Cohen, A. E., and W. E. Moerner. 2007. Principal-components analysis of shape fluctuations of single DNA molecules. *Proc. Natl. Acad. Sci. USA.* 104:12622–12627.

46. Hansen, L. K., J. Larsen, …, O. B. Paulson. 1999. Generalizable patterns in neuroimaging: how many principal components? *Neuroimage.* 9:534–544.

47. Hashimoto, A., Y. Yamaguchi, …, E. Tamiya. 2015. Time-lapse Raman imaging of osteoblast differentiation. *Sci. Rep.* 5:12529.

48. Rector, D. M., R. F. Rogers, …, J. S. George. 2001. Scattered-light imaging in vivo tracks fast and slow processes of neurophysiological activation. *Neuroimage.* 14:977–994.

49. Furnival, T., R. K. Leary, and P. A. Midgley. 2016. Denoising time-resolved microscopy image sequences with singular value thresholding. *Ultramicroscopy.* Published online May 10, 2016. http://dx.doi.org/10.1016/j.ultramic.2016.05.005.

50. Kim, T. W., C. Yang, …, H. Ihee. 2016. Combined probes of x-ray scattering and optical spectroscopy reveal how global conformational change is temporally and spatially linked to local structural perturbation in photoactive yellow protein. *Phys. Chem. Chem. Phys.* 18:8911–8919.

51. Pérez, J., P. Vachette, …, D. Durand. 2001. Heat-induced unfolding of neocarzinostatin, a small all-beta protein investigated by small-angle x-ray scattering. *J. Mol. Biol.* 308:721–743.

52. Malmerberg, E., Z. Omran, …, R. Neutze. 2011. Time-resolved WAXS reveals accelerated conformational changes in iodoretinal-substituted proteorhodopsin. *Biophys. J.* 101:1345–1353.

53. Haldrup, K. 2014. Singular value decomposition as a tool for background corrections in time-resolved XFEL scattering data. *Philos. Trans. R. Soc. B.* 369:20130336.

54. Boetker, J. P., J. Rantanen, …, B. J. Boyd. 2016. Anhydrate to hydrate solid-state transformations of carbamazepine and nitrofurantoin in biorelevant media studied in situ using time-resolved synchrotron x-ray diffraction. *Eur. J. Pharm. Biopharm.* 100:119–127.

55. Oka, T., N. Yagi, …, 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.

56. Macnaughtan, D., L. B. Rogers, and G. Wernimont. 1972. Principal-component analysis applied to chromatographic data. *Anal. Chem.* 44:1421–1427.

57. Maggio, R. M., L. Cerretani, …, E. Chiavaro. 2012. Application of differential scanning calorimetry-chemometric coupled procedure to the evaluation of thermo-oxidation on extra virgin olive oil. *Food Biophys.* 7:114–123.

58. Idborg, H., P. O. Edlund, and S. P. Jacobsson. 2004. Multivariate approaches for efficient detection of potential metabolites from liquid chromatography/mass spectrometry data. *Rapid Commun. Mass Spectrom.* 18:944–954.