

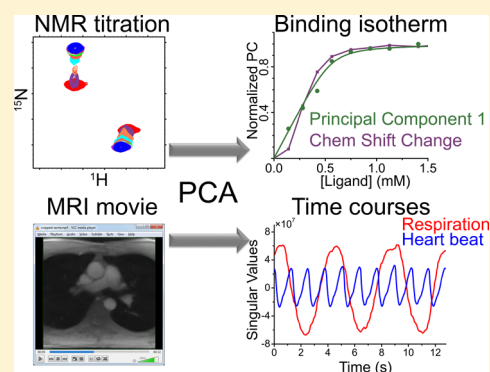
Binding Isotherms and Time Courses Readily from Magnetic Resonance

Jia Xu and Steven R. Van Doren*

Department of Biochemistry, University of Missouri, 117 Schweitzer Hall, Columbia, Missouri 65211, United States

S Supporting Information

ABSTRACT: Evidence is presented that binding isotherms, simple or biphasic, can be extracted directly from noninterpreted, complex 2D NMR spectra using principal component analysis (PCA) to reveal the largest trend(s) across the series. This approach renders peak picking unnecessary for tracking population changes. In 1:1 binding, the first principal component captures the binding isotherm from NMR-detected titrations in fast, slow, and even intermediate and mixed exchange regimes, as illustrated for phospholigand associations with proteins. Although the sigmoidal shifts and line broadening of intermediate exchange distorts binding isotherms constructed conventionally, applying PCA directly to these spectra along with Pareto scaling overcomes the distortion. Applying PCA to time-domain NMR data also yields binding isotherms from titrations in fast or slow exchange. The algorithm readily extracts from magnetic resonance imaging movie time courses such as breathing and heart rate in chest imaging. Similarly, two-step binding processes detected by NMR are easily captured by principal components 1 and 2. PCA obviates the customary focus on specific peaks or regions of images. Applying it directly to a series of complex data will easily delineate binding isotherms, equilibrium shifts, and time courses of reactions or fluctuations.



Affinity measurements are essential in understanding molecular recognition and in assessing drug discovery. Time courses of chemical and biological transformations are of wide interest. A theme shared in monitoring either equilibria or kinetics is to describe the shifts in population, the central interest of this Article. We propose to marshal a classic method of chemometrics to follow such shifts more generally.

In the case of ligand associations, a preferred spectral approach has been heteronuclear NMR, due to its information on binding site and suitability over a range of affinities.^{1–5} Typically, the ligand-binding equilibrium is monitored by shifts of NMR peaks.^{1,2,4} Arriving at affinities, however, has meant traveling through slow bottlenecks of spectral peak picking to obtain binding isotherms, usually assignment of the peaks, and global fitting of a binding isotherm consistent with the shifts of multiple peaks of the protein or macromolecule.⁶ Despite the advantages of this approach and rapidity of modern collection of spectra,^{7,8} the time invested in interpreting these spectra is a barrier to wider and faster applications. Below, we propose an improved strategy that bypasses the selection of favorable peaks in spectra and favorable features in images for analysis.

The stepwise population changes due to ligand binding in a titration are usually accompanied by changes in NMR peaks that depend on the exchange regime, i.e., the time scale of chemical exchange relative to the chemical shift differences between free and bound states. Behaviors of fast, slow, and intermediate exchange regimes are depicted in Figure S1. Peak shifts in the fast exchange regime are favored for modeling binding isotherms.^{4,9} In the slow exchange regime, peaks

representing the free state can disappear and reappear elsewhere in the bound state, complicating peak assignments. In intermediate exchange, the nonlinearity of chemical shift changes from titrations can corrupt binding isotherms with sigmoidal distortion, resulting in skewed and unreliable fits of the association⁴ (Figure S1).

Principal component analysis (PCA) reduces the dimensionality of data to reveal a simpler set of shared features or patterns. It is efficient, robust, and widely applied in chemometrics, analytical spectroscopy, and imaging.^{10,11} PCA is often implemented using singular value decomposition (SVD). The approach has only occasionally been applied to reactions monitored by 2D NMR spectra.^{12–16} These included resolution of time-dependent¹² or pH-dependent components (using CS-PCA).¹³ PCA filtered noise out of spectra to improve global fits of binding.¹⁵ SVD of peak heights from in-cell NMR spectra of proteins associating suggested the binding site.¹⁶ The SVD of these NMR studies was applied to peak pick lists,^{13–16} rather than to the stack of 2D NMR spectra “unfolded” into a stack of vectors, which avoided peak lists and worked well on sparse 2D NMR spectra.¹² In NMR-detected titrations, the applicability of PCA is regarded at this writing as limited to the fast exchange regime.^{14,17,18} The need for wide applicability to complex scenarios such as binding of multiple

Received: May 17, 2016

Accepted: July 26, 2016

Published: July 26, 2016

ligands, mixtures of chemical exchange regimes, and changing linewidths was articulated.¹⁴ The work herein responds to this need.

PCA can be computed by either SVD or eigenvector decomposition of covariance, aiming at maximization of variance with minimization of correlation and redundancy (see the [Supporting Information](#) for more detail). PCA computes new orthogonal components that are linear combinations of the original experimental variables, with the first principal component (PC1) reporting the largest variance. Jolliffe asserts that PCA is often useful for data deviating from Gaussian distributions and linear relationships of observed variables to underlying components.¹⁹

Magnetic resonance imaging (MRI) of brain and diseased tissues presents opportunities for chemometrics, such as comparing and registering images spatially, temporally, and metabolically.^{20–24} Resolution of trends of change between the frames of a stack of congruent images or 2D spectra can be undertaken by three-way multiple image analysis such as “unfold”-PCA, which simplifies the 3D stack into two dimensions for standard PCA.^{12,25}

We demonstrate how to extend unfold-PCA to extract binding isotherms successfully from 2D NMR spectra of ligand titrations in slow exchange and problematic intermediate exchange by introducing preprocessing steps. Moreover, the improved approach *needs no peak picking or peak assignments*. The algorithm is even successful in deriving binding isotherms from the unprocessed free induction decays (FIDs) from titrations in fast or slow exchange. When a second binding process has been detected spectrally, PCA can also derive it as the second component of the reaction. Likewise, this enhancement of unfold-PCA is general enough to extract multiple and periodic time-varying components from MRI movies. Applying PCA directly to a series of spectra or images saves much time in handling them and in resolving the processes present.

■ EXPERIMENTAL SECTION

Preprocessing of Spectra and Images for SVD. Each spectrum or image in the series of measurements is collected and processed under identical conditions, except for the experimental variable changed (concentration, time, pH, etc.). Each 2D spectrum or image ($F1 \times F2$ points) is rearranged as a 1D vector arrayed over the experimental variable²⁵ ([Figure S2](#)). Each vector is compressed, by deleting unchanging positions, in order to expedite computational manipulations of the matrix X' . Low intensity regions of the vectorized spectra were usually filtered out prior to SVD. Alternative choices of no scaling, autoscaling, and Pareto scaling²⁶ of the rows of X' were compared. The rows were mean-centered.¹¹

Extraction of Principle Components. SVD of X' can be expressed as

$$X'_{mn} = U_{mn} S_{nn} V_{nn}^T \quad (1)$$

where U and V^T are orthogonal matrices, S is a diagonal matrix, and subscripts denote sizes of matrices. The eigenvectors of $X'^T X'$ constitute the matrix V^T containing the singular vectors of interest, such as PC1 as the first row with the largest trend ([Figure S2](#)) and PC2 as the second row with the second largest trend. PC1 may depend on time,²⁷ [ligand],¹⁵ or other conditions.¹³ The simulations of NMR spectra used for part

of the testing PCA applied directly to them are described in the [Supporting Information](#).

■ RESULTS AND DISCUSSION

PCA Capture of Time Courses. We extended the unfold-PCA strategy of converting a 3D stack of 2D NMR spectra (perturbed by the experimental variable) into a 2D array of vectors for SVD.¹² To improve performance, we inserted preprocessing steps for data compression, noise filtration, and scaling options ([Figure S2](#)). We automated these processing and calculation procedures for multiple data formats.²⁸ This algorithm avoids user selection of features in the data ([Figure S2](#)). Its ability to capture main trends is introduced using time-lapse images of a sunset or multiplying bacteria ([Figure S3](#)). The trajectory of the setting sun is marked by PC1 ([Figure S3A,B](#)). The exponential growth in bacteria is represented by PC1, despite their motility ([Figure S3C,D](#)). Applying the same PCA approach to time-lapse 2D NMR spectra captures a reaction progress curve as PC1. Changes in ^1H – ^{15}N correlation spectra have been used to track dephosphorylation or phosphorylation rates.^{29,30} PCA applied directly to time-lapse TROSY spectra of a phosphoryl transfer enzyme reveals the time course of dephosphorylation ([Figure S3E,F](#)). The kinetics derived from unsupervised PCA of entire spectra echo those obtained from global fitting of carefully selected peak height changes²⁹ but with new ease.

Fast Exchange Scenarios. PCA was demonstrated on peak pick lists of titrations with NMR peaks in the fast exchange regime, where the shifts of the peak positions are linear combinations of the basis spectra and suffice to indicate population change.^{13,14,16} However, applying PCA directly to noninterpreted spectra means that more information is considered: not only selected peak positions but also line shapes (widths, heights, volumes, etc.) throughout the spectrum. Autoscaling³² and Pareto scaling²⁶ perform acceptably when applying the improved algorithm to fast exchange ([Figure S4A,B](#)). Autoscaling is, however, more accurate and precise for fast exchange, especially with the threshold for retention of spectral points set to 3- to 7-fold the noise level ([Figure S4A,B](#)).

The list-based and improved spectrum-based implementations of PCA reproduce conventional results in obtaining binding isotherms. An example of 1:1 protein–ligand binding in the fast exchange regime with K_D set to 270 μM is shown with the simulated titration of [Figure 1A](#). Application of PCA to lists of all peaks provides an accurate binding isotherm as PC1 plotted vs [ligand]. Fitting to standard [eq S4](#) places K_D at $271 \pm 17 \mu\text{M}$ ([Figure 1B](#)). This indicates that PCA of all peak positions, whether shifted by the ligand or not, matches conventional global fitting of only the big shifts of well-resolved peaks. It is more convenient and thorough to apply the improved unfold-PCA algorithm directly to the spectra ([Figure S2](#)). The binding isotherm captured as PC1 in this way reproduces the true populations ([Figure 1B](#)). This is also illustrated for the titration of a phosphoprotein binding domain with a phosphoThr peptide in fast exchange³¹ ([Figure 1C](#)). PC1 direct from the spectra delineates the binding isotherm fitted by K_D of $36 \pm 4 \mu\text{M}$ ([Figure 1D](#)), which closely resembles the binding isotherms and K_D of $40 \pm 5 \mu\text{M}$ globally fitted previously to the shifts of multiple amide peaks.³¹ PCA of lists of the spectral peaks picked from the titration provides PC1 fitted by a similar K_D of $34 \pm 3 \mu\text{M}$ ([Figure 1D](#)).

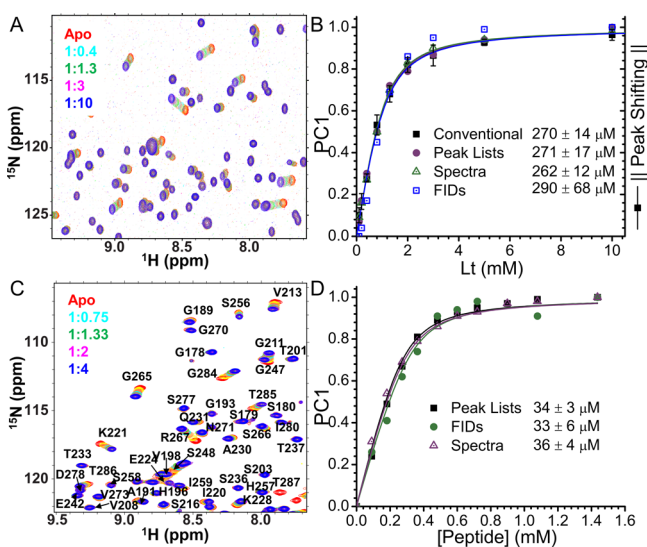


Figure 1. PC1 from SVD of titrations in fast exchange, simulated or measured, represents Langmuir binding isotherms. (A) Simulated spectral shifts in the fast exchange regime. The colors of the contours progress with ligand additions up to 10-fold excess. (B) Binding isotherms were obtained by applying SVD to the simulated spectra without peak picking (triangles), peak pick lists (circles), or the simulated raw FIDs (open squares). Black squares mark conventional, global fitting of the shifts of individual peaks. ||.|| denotes normalization of the peak shifts. (C) Superposed ^{15}N HSQC spectra of a phosphoprotein-binding FHA domain ($600\ \mu\text{M}$) titrated with a phosphopeptide from a protein kinase exhibit fast exchange behavior.³¹ (D) Binding isotherms were derived from the titration shown in (C) by applying SVD directly to the spectra (open triangles), lists of the peaks of each spectrum (squares), or FIDs (circles). The K_D of $40 \pm 5\ \mu\text{M}$ globally fitted to the peak shifts of multiple amide peaks³¹ is closest to the K_D fitted to PC1 of the spectra.

Parseval's theorem suggests that signals in time and frequency domains can be considered equivalent.³³ With this in mind, PCA of the unprocessed FIDs was also evaluated (Figure 1). PC1 derived from the array of FIDs from the simulation of fast exchange managed to obtain a binding isotherm with nearly correct affinity but larger uncertainty, i.e., K_D of $290 \pm 68\ \mu\text{M}$ (Figure 1B). This outcome is promising for PCA overcoming the high level of noise added to the simulated example (S/N of 5 at the median peak height). PCA of the sets of FIDs from the protein titration with phosphoThr peptide in fast exchange³¹ generated a binding isotherm with K_D close to the $33 \pm 6\ \mu\text{M}$ obtained by other methods (Figure 1D). The smaller uncertainties when applying PCA after Fourier transformation might reflect increased sensitivity from integration of the signals or from better signal resolution.

Slow Exchange Scenarios. Binding isotherms can be constructed conventionally in the slow exchange regime (with slower k_{off} and higher affinities) from changes of peak volumes or heights but with more difficulty and rarity. Tracking the appearance of bound state peaks is preferred⁴ but can be complicated by challenging peak assignments and peak attenuation by line broadening. PCA of the simulated titration (K_D set at $270\ \mu\text{M}$) in the slow exchange regime derives a binding isotherm as PC1 that is virtually indistinguishable (K_D of $262 \pm 9\ \mu\text{M}$) from the simulated populations (Figure 2B). SVD of the series of spectra derives robust binding isotherms from titrations in slow exchange. The fits to them are precise with all three options of scaling, provided that with autoscaling

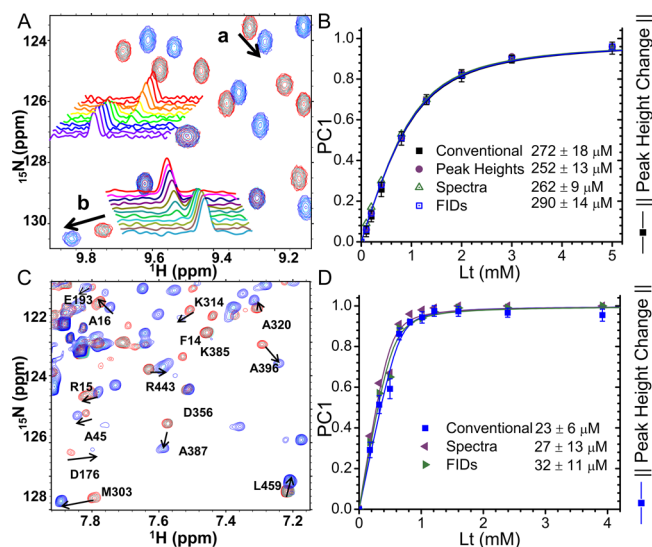


Figure 2. SVD of titrations featuring slow exchange, in simulated or measured NMR spectra, distills binding isotherms as PC1. (A) Overlay of HSQC spectra simulated with slow exchange. Protein ligand ratios of 1:0, 1:1.3, and 1:10 are represented by red, cyan, and darker blue, respectively. Insets are 1D slices of peak pairs indicated by black arrows. (B) PC1 derived from the simulated series of spectra (triangles) in panel A provides binding isotherms equivalent to plotting heights of disappearing peaks of the free state (black squares). PC1 was also calculated from peak lists (circles) or the FIDs (open squares). (C) Spectra from a slow exchange titration of an enzyme with an inhibitor. ^{15}N TROSY spectra of PMM (52 kDa, 800 MHz, 25 °C) titrated with X1P are superposed and contain amide peaks in slow exchange. PMM/X1P ratios of 1:0, 1:0.6, and 1:8 are represented by red, cyan, and blue, respectively. (D) PC1 of either the spectra or FIDs from this titration captures the binding isotherm. Standard global fitting of peak heights is shown with blue symbols for comparison.

the threshold for data inclusion is kept ≤ 7 -fold the noise level (Figure S4E,F). PC1 extracted from simulated FIDs provides a binding isotherm resembling the simulated populations, with slight deviations in points and fitted K_D of $290 \pm 14\ \mu\text{M}$ (Figure 2B). PCA was applied to the entirety of crowded ^{15}N TROSY spectra of the 52 kDa PMM enzyme titrated by its inhibitor xylose 1-phosphate (X1P), exhibiting slow exchange behavior (Figure 2C). The binding isotherm globally fitted to the increasing peak heights of several selected bound state peaks estimates K_D at $23 \pm 6\ \mu\text{M}$. (The blue curve in Figure 2D summarizes many normalized peak heights fitted.) The points of PC1 obtained directly from the spectra are fitted by K_D of $27 \pm 13\ \mu\text{M}$ and PC1 from FIDs by K_D of $32 \pm 11\ \mu\text{M}$ (Figure 2D). These PC1-derived binding isotherms match well those obtained from conventional global fitting of bound peak heights but with the advantages of minimal data handling or interpretation.

Intermediate Exchange Scenarios. Intermediate exchange is most problematic for estimating affinities due to its sigmoidal plots of NMR peak shifts⁴ vs [ligand] (Figures S1F and 3B). These nonlinear shifts can be fitted erroneously with deviations up to 2 orders of magnitude from actual.⁴ It can also be misconstrued as evidence of cooperativity.

In intermediate exchange, both line shapes and peak positions appear to be critical for capturing population change. As a simple and extreme case, NMR spectra of a titration were simulated with intermediate exchange broadening in *all* peaks in the ^1H dimension. The application of standard autoscaling³²

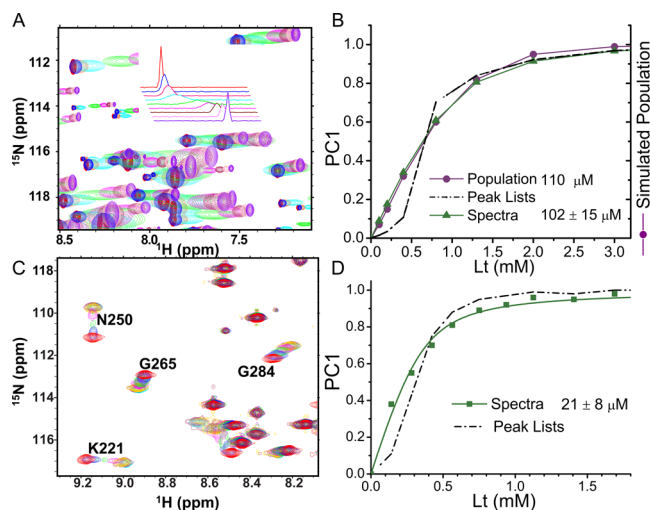


Figure 3. Suppressing the intermediate exchange distortion of binding isotherms by applying PCA directly to spectra. (A) HSQC spectra simulated to be intermediate to fast in exchange for ^1H chemical shift changes and line shapes. The inset shows slices through a shifted and broadened peak. (B) In intermediate to fast exchange, the ligand-induced peak shifts deviate sigmoidally from a 1:1 binding isotherm when applying PCA to the peak pick lists (dashed line). The lag is suppressed in PC1 (green triangles) from SVD of Pareto-scaled spectra. (C) A region of the ^{15}N HSQC spectrum of the FHA domain titrated with a phosphopeptide displays intermediate-fast exchange behavior at the peaks of four amino acids labeled. (D) PC1 of the spectra yields a binding isotherm fitted by K_D of $21 \pm 8 \mu\text{M}$, which agrees with the K_D of $20 \mu\text{M}$ measured by isothermal titration calorimetry.³¹

in the algorithm of Figure S2 falls short of the accuracy and precision needed (see purple box in Figure S4C,D). For obtaining a binding isotherm of high accuracy and precision from intermediate exchange behavior, Pareto scaling of the rows is required and improved by the threshold remaining small (Figure S4C,D). Though the shifts of all peaks are sigmoidal (Figure 3A,B), PCA of the Pareto-scaled, linearized spectra avoids any such distortion of PC1; it is best fitted by a K_D of $102 \pm 15 \mu\text{M}$ that agrees with the simulated K_D (Figure 3B). Pareto scaling with a low threshold increases the weighting of weak peaks broadened by intermediate exchange and appears to move the data closer to a Gaussian (Figure S6), the distribution optimal for PCA.¹⁹

Mixtures of Regimes. It is much more typical of titrations with NMR peaks in intermediate exchange to be accompanied by other peaks in fast or slow exchange. We simulated a titration with a mixture of all three regimes and 34% of the peaks in intermediate exchange (Figure S8A). The sigmoidal shifts of the latter are enough to cause PCA of the lists of all picked peaks to extract PC1 which is sigmoidal and unacceptable as a binding isotherm (Figure S8B). The application of PCA to these spectra instead (with Pareto scaling for accuracy) successfully captures the simulated population change as PC1 with fitted K_D within 7% of the simulated value (Figure S8B). When using only peaks in intermediate exchange from this simulation (Figure S8C), the sigmoidal distortion of PC1 from PCA of peak lists worsens, but PCA of the Pareto-scaled spectra still suppresses distortion of PC1, as is evident from fitted K_D within 13% of the actual value (Figure S8D).

^{15}N HSQC spectra of an FHA domain titrated with a phosphoThr peptide³¹ exhibit intermediate-fast exchange

(Figure 3C). Though numerous unaffected peaks are also present, fitting of the PC1-derived binding isotherm matches the K_D of $20 \pm 3 \mu\text{M}$ measured independently by isothermal titration calorimetry (Figure 3D). PCA is not recommended for application to FIDs with intermediate exchange broadening because of the skewing of PC1 that results (Figure S9E,F).

Applying unfold-PCA to spectra along with the preprocessing recommended herein (Figures S2 and S4) reliably defines the binding isotherms. This is much easier than seeking K_D through fitting of line shapes or competition experiments⁴ requiring prior knowledge of relative ligand affinity. Use of PCA does not change the need for [protein] to be 0.2 to 0.8 of K_D for best accuracy in fitting K_D and within 10-fold for acceptable accuracy.^{5,9} When affinities are too tight to use this range (evident as an abrupt transition), competition can then be introduced to weaken the affinity of interest into the concentration range where it can be fitted accurately.^{4,5,15}

Two-Step Binding. Next, we attempted resolution of two binding events, reactions determined to be sequential.³⁴ In the course of multiple ligand binding, mixed exchange regimes are likely to complicate previous strategies of analysis. Cogliati et al. reported a challenging mixture of exchange regimes in the two-step binding of two molecules of sodium glycochenodeoxycholate (GCDA) to bile acid binding protein³⁴ (Figure 4A).

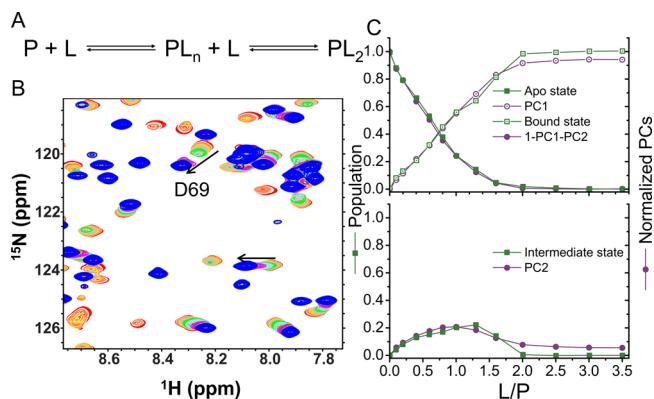


Figure 4. Principal components from SVD of spectra agree with the populations estimated earlier by line shape analysis³⁴ for a titration of two sequential binding events. (A) Scheme of the two-step binding mechanism hypothesized. (B) Chicken liver bile acid binding protein with disulfide bridge was titrated with GCDA and underwent intermediate exchange broadening, as is evident for two peaks marked with arrows in the superposed HSQC spectra.³⁴ (C) HSQC spectra of this protein titrated with GCDA, specifically ligand/protein ratios of 0, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 1.3, 1.6, 2.0, 2.5, 3.0, and 3.5, with contours ranging from red to blue. Black arrows indicate peaks in intermediate exchange.³⁴ (C) Comparison between normalized PCs (purple) and populations of the states P, PL, and PL_2 previously calculated using line shape analysis (green, adapted from Figure 3e in ref 34 with permission, copyright 2010 John Wiley & Sons).

The titrations display a mixture of fast, slow, and intermediate exchange regimes accompanying the complex binding (Figure 4B). The authors exploited line shape analysis to selected amide NMR peaks undergoing intermediate exchange broadening; see those marked with black arrows in Figure 4B.³⁴ This enabled them to estimate the proportions of the apo (P), intermediate (PL), and ligand-saturated (PL_2) states through the course of titrations³⁴ (green in Figure 4C).

The application of SVD directly to the same spectra without peak picking and with Pareto scaling results in PC1 accounting

for 61% of the variances and PC2 accounting for 12% (Table S2). PC1 approximates the disappearance of the apo state P. The quantity $1 - PC1$ (not shown) resembles but slightly exceeds the formation of the fully bound state PL2 (Figure 4C). PC2 resembles the rise and fall of the population of the singly ligated intermediate PL, once PC2 is normalized to the scale of PC1 (Figure 4C). Since the population changes of P and PL2 are highly correlated ($R = -0.93$) and hence statistically related, it is mathematically unrealistic to distinguish these two correlated components by PCA, a decorrelation technique.

When no ligand is present ($L/P = 0$) or the bile acid binding protein is saturated with the GCDA ligand (e.g., $L/P = 3.5$), PC1 and PC2 sum to 1.0 in agreement with the proportions of PL and PL2 summing to 1.0. Consequently, the sum of PC1 and PC2 is renormalized to 1.0. This implies that PL₂ should be modeled by $1 - PC1 - PC2$, which matches well the fractional concentrations of PL₂ estimated previously³⁴ (Figure 4C).

Nonlinearity and Applicability of PCA. Are the nonlinear peak shifts of the peaks in intermediate exchange (see Figures 3, 4, and S8) suitable for PCA? Neither SVD nor covariance calculations require Gaussian distributions.¹⁹ The series of NMR spectra and time-lapse images analyzed in this study *all* have a degree of the nonlinear character (non-normal distributions) exemplified more dramatically by a chaotic system (Figure S7). This may result from the spectra and images containing more components than lists of their peaks or features. It would require multiple PCs to capture most of the greater complexity to *reconstruct* the original measurements (with matrix U in eq 1). However, for this study's more modest goal of extracting the largest population shifts among the spectra or images, the nonlinearity (Figure S7) does not interfere in the largest PCs capturing the main processes. When these largest trends are abstracted from matrix V^T (eq 1), they robustly withstand nonlinearity. The central limit theorem generates an approximation of normality for most data sets, as they have the large size required by the theorem. The scaling of the data matrix of spectra appears to shift it toward a normal-like distribution (Figure S6). Thus, discovering the main trends requires far fewer PCs from matrix V^T than needed for faithful reconstruction of nonlinear spectra and images using matrix U .

Periodic and Multiple Components from MRI by PCA.

We tested the fitness of this SVD approach for wider applications to measurements paralleling macromolecular NMR spectra in being complex and responsive to coordinated processes, e.g., MRI movies. The SVD algorithm extracts from an MRI movie of brain fluctuations³⁵ the periodic flow of cerebral spinal fluid as PC1 (Figure 5A,B). PC1 from the full breadth of the movie frames appears similar to the reported modulation of image intensities within the box confined to the third ventricle³⁶ (Figure 5A,B). PC1 represents the 5 cycles of respiration, each with 2.5 s of inspiration and 2.5 s of expiration, similarly to the conventional plot of the localized intensities of the MRI signal³⁶ (Movie S1). PC1 being smoother than the local intensity changes may reflect the integration of more covarying data and the noise filtering that is intrinsic to PCA.

We also applied this PCA approach to an MRI movie of a chest cross-section³⁷ through the large arteries (the aorta and pulmonary trunk) and vein (superior vena cava) each connected to the heart (Figure 5C). The aorta, pulmonary trunk, and superior vena cava pulse in unison upon contraction of the heart, while chest dimensions undulate more slowly with breathing³⁷ (Movie S2). Applying unfold-PCA to the standard magnitude view of the MRI movie easily extracts four time

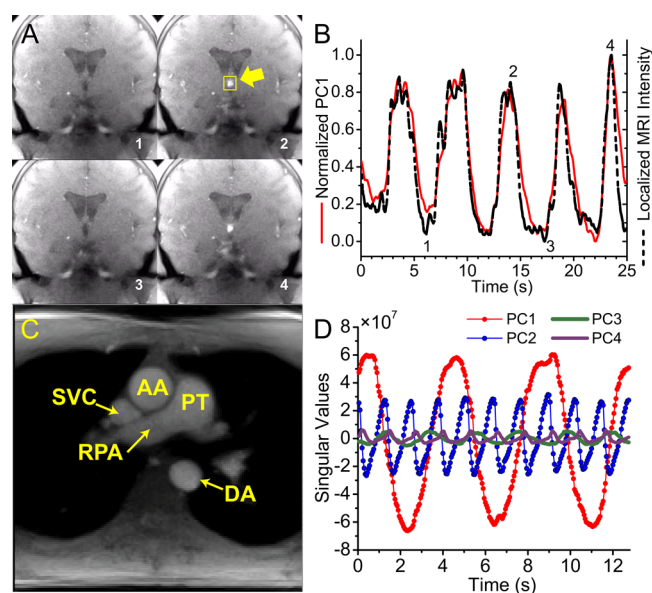


Figure 5. SVD extracts the time courses of pulsation in MRI movies of cross sections through the brain³⁵ or chest.³⁷ (A) Frames from the brain imaging (Movie S1, adapted from ref 35 with permission, copyright BiomedNMR/CC-BY-SA-3.0) feature cerebral spinal fluid flow most apparent within the box pointed out by an arrow in frame 2.^{35,36} (B) PC1 from the movie captures five cycles of breathing, plotted with the red line. Signal intensities within the boxed central region with the arrow in the third ventricle are plotted with the black dashed line. (C) A frame from the movie of ref 37 (adapted with permission, copyright 2014 John Wiley & Sons) is labeled AA for ascending aorta, DA for descending aorta, PT for pulmonary trunk, RPA for right pulmonary artery, and SVC for superior vena cava. (D) The time courses of the four PCs generated by unsupervised SVD are plotted and suggest four types of periodic fluctuations. This movie³⁷ is synchronized with plotting of its PC1 and PC2 in Movie S2.

courses as PC1 to PC4. PC1 represents breathing with three cycles of inspiration and expiration (red in Figure 5D and Movie S2). PC2 represents the pulsation of the major arteries and superior vena cava upon heart contraction for ten consecutive heart beats; the troughs mark the expansion of the vessels (blue in Figure 5D and Movie S2). The process represented by PC3 is unclear but is synchronized to breathing and repeats at exactly twice the frequency of PC1 and breathing. Movie reconstruction²⁸ using only PC3 suggests subtle fluctuations in the pulmonary trunk (not shown), which ties to the lungs. PC4 is clearly synchronized to the cardiac cycle. Movie reconstruction²⁸ reveals that PC4 affects the pulmonary trunk the most and the aorta slightly. The crests of PC4 (Figure 5D) probably represent contraction of the heart (systole) because they are narrow and immediately precede the bolus of blood that appears in the arteries (troughs in PC2). The broad troughs of PC4 probably represent the relaxation of the heart known as diastole, with its rapid filling and subsequent slower filling phases; these are evident as the steeper and more gradual slopes at the bottom of the troughs (Figure 5D). Thus, the strategy of applying PCA directly to the series of images resolves multiple concurrent processes. Two PCs are as intuitive as breathing and heart beat while another PC represents phases of the cardiac cycle.

Tallying Meaningful Principal Components. Determining the number of meaningful PCs can become important when there are concurrent processes. Scree plots of the contributions

of PCs are widely trusted and give especially clear suggestions of the significant PCs for the peak lists and movies that we analyzed. Additional strategies of counting significant PCs were proposed (e.g., singular values and RMSD)^{15,38} but appear inconclusive in all applications of unfold-PCA to the series of spectra and images that we have examined, except to highlight the ubiquity of nonlinear behavior (Figure S7). Even for a simple titration with NMR peaks in slow to intermediate exchange, using the percentage of the variances accounted for cannot judge the adequacy of the single component (Figure S6). The criterion that a PC be *smooth* (high autocorrelation),¹⁵ however, appears more reliable for recognizing a meaningful component, when coupled with some understanding of the processes. For example, in 1:1 protein–ligand binding, the hyperbolic PC1 curve represents the binding isotherm regardless of the proportion of variance contributed by PC1. This inspection of PC1 works for the slow-intermediate exchange example (Figure S6). When more than one significant component is present, the shapes of lesser PCs need to be checked.¹⁵ In analyses of protein–ligand titrations with two reactions (see Figure 4), PC1 and PC2 are smooth and clearly larger than other PCs (Figure S10).

Limits to Applications of PCA to Spectra and Images.

We have encountered instances of deterioration or failure of the improved unfold-PCA algorithm. PCs were corrupted when spectral windows, signal averaging, management of water suppression, or gain were not uniform. This is usually overcome by applying SVD to peak pick lists. SVD of unprocessed FIDs diminished by simulated intermediate exchange failed to represent the binding isotherms of the titrations (Figure S9F). This is avoided by Fourier transformation. When SVD is applied to 1D spectra of abnormally low digital resolution, the accuracy of the binding isotherm deteriorates (Figure S6). However, PCA appears remarkably reliable in representing at least two processes from a series of 2D measurements.

Potential Applications to Digital Data. Unfold-PCA, improved by preprocessing steps described, can process many kinds of series of comparable spectra and images. It makes most sense to apply it to data that are complex but that respond to one or more concerted processes, for the purpose of finding the main trends. Macromolecular NMR and MRI provide good examples. Plotting the course of protein folding intermediates recorded by expedited NMR spectra³⁹ is another potential application. Potential applications may extend to other series of 2D measurements such as spectra, gels, and imaging of microarrays,⁴⁰ chromatographic separations,⁴¹ electrochemistry,⁴² and chemical biology signals.^{43,44}

CONCLUSIONS

The application of this PCA strategy (enhanced by preprocessing) to a series of spectra or MRI images offers convenience and wide applicability to characterizing concerted processes. Such applications will expand the accessibility of affinities, equilibria, kinetics, and time-evolving processes. This will include noninterpreted, unassigned, and overlapped features in spectra and movies, which may number two or more concurrent processes. For example, NMR studies will be enabled to elucidate binding isotherms masked by intermediate exchange and/or two or more concurrent processes.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b01918.

Further background, methods of testing, ten figures, two tables, and references (PDF)

MRI magnitude movie of cerebral spinal fluid flow, with plotting of PCs (Movie S1) (AVI)

MRI magnitude movie of ascending aorta in the chest, with plotting of PCs (Movie S2) (AVI)

AUTHOR INFORMATION

Corresponding Author

*E-mail: vandorens@missouri.edu.

Notes

The authors declare no competing financial interest. Software for performing the analyses²⁸ (free for academics) may be requested from <http://biochem.missouri.edu/trend>.

ACKNOWLEDGMENTS

We are grateful to J. Frahm and his group for real-time MRI movies and to H. Molinari and L. Ragona for spectra of titrations of chicken bile acid binding protein. We thank Y. Fulcher and L. Beamer for discussion, Beamer for PMM, and T. Mawhinney for synthesizing X1P. This work was supported by NSF grant MCB1409898. Spectrometer purchases were supported in part by NIH grants RR022341 toward the 800 MHz and GM57289 toward the 600 MHz system.

REFERENCES

- (1) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. *J. Med. Chem.* **2005**, *48*, 2518–2525.
- (2) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531–1534.
- (3) Shortridge, M. D.; Hage, D. S.; Harbison, G. S.; Powers, R. J. *Comb. Chem.* **2008**, *10*, 948–958.
- (4) Williamson, M. P. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16.
- (5) Fielding, L. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51*, 219–242.
- (6) Lowe, A. J.; Pfeffer, F. M.; Thordarson, P. *Supramol. Chem.* **2012**, *24*, 585–594.
- (7) Gal, M.; Schanda, P.; Brutscher, B.; Frydman, L. *J. Am. Chem. Soc.* **2007**, *129*, 1372–1377.
- (8) Amero, C.; Schanda, P.; Dura, M. A.; Ayala, I.; Marion, D.; Franzetti, B.; Brutscher, B.; Boisbouvier, J. *J. Am. Chem. Soc.* **2009**, *131*, 3448–3449.
- (9) Markin, C.; Spyropoulos, L. *J. Biomol. NMR* **2012**, *53*, 125–138.
- (10) Trygg, J.; Holmes, E.; Lundstedt, T. *J. Proteome Res.* **2007**, *6*, 469–479.
- (11) Adams, M. J. *Chemometrics in Analytical Spectroscopy*; Royal Society of Chemistry: Cambridge, 2004.
- (12) Jaumot, J.; Marchan, V.; Gargallo, R.; Grandas, A.; Tauler, R. *Anal. Chem.* **2004**, *76*, 7094–7101.
- (13) Sakurai, K.; Goto, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 15346–15351.
- (14) Konuma, T.; Lee, Y. H.; Goto, Y.; Sakurai, K. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 107–118.
- (15) Arai, M.; Ferreón, J. C.; Wright, P. E. *J. Am. Chem. Soc.* **2012**, *134*, 3792–3803.
- (16) Majumder, S.; DeMott, C. M.; Burz, D. S.; Shekhtman, A. *ChemBioChem* **2014**, *15*, 929–933.

- (17) Furukawa, A.; Konuma, T.; Yanaka, S.; Sugase, K. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *96*, 47–57.
- (18) Selvaratnam, R.; Chowdhury, S.; VanSchouwen, B.; Melacini, G. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 6133–6138.
- (19) Jolliffe, I. T. *Principal Component Analysis*, 2nd ed.; Springer-Verlag: New York, 2002; p 19, 396.
- (20) Witjes, H.; Simonetti, A. W.; Buydens, L. *Anal. Chem.* **2001**, *73*, 548A–556 A.
- (21) Nika, V.; Babyn, P.; Zhu, H. *J. Med. Imaging (Bellingham)* **2014**, *1*, 024502.
- (22) Wu, J.; Gong, G.; Cui, Y.; Li, R. *J. Magn. Reson. Imaging* **2016**, DOI: [10.1002/jmri.25279](https://doi.org/10.1002/jmri.25279).
- (23) Kawaguchi, H.; Shimada, H.; Kodaka, F.; Suzuki, M.; Shinotoh, H.; Hirano, S.; Kershaw, J.; Inoue, Y.; Nakamura, M.; Sasai, T.; Kobayashi, M.; Suhara, T.; Ito, H. *PLoS One* **2016**, *11*, e0151191.
- (24) Huizinga, W.; Poot, D. H. J.; Guyader, J. M.; Klaassen, R.; Coolen, B. F.; van Kranenburg, M.; van Geuns, R. J. M.; Uitterdijk, A.; Polfliet, M.; Vandemeulebroucke, J.; Leemans, A.; Niessen, W. J.; Klein, S. *Med. Image Anal.* **2016**, *29*, 65–78.
- (25) Huang, J.; Wium, H.; Qvist, K. B.; Esbensen, K. H. *Chemom. Intell. Lab. Syst.* **2003**, *66*, 141–158.
- (26) van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. *BMC Genomics* **2006**, *7*, 142.
- (27) Casanovas, O.; Jaumot, M.; Paules, A. B.; Agell, N.; Bachs, O. *Oncogene* **2004**, *23*, 7537–7544.
- (28) Xu, J.; Van Doren, S. R., submitted for publication.
- (29) Xu, J.; Lee, Y.; Beamer, L. J.; Van Doren, S. R. *Biophys. J.* **2015**, *108*, 325–337.
- (30) Mayzel, M.; Rosenlow, J.; Isaksson, L.; Orekhov, V. Y. *J. Biomol. NMR* **2014**, *58*, 129–139.
- (31) Ding, Z.; Wang, H.; Liang, X.; Morris, E. R.; Gallazzi, F.; Pandit, S.; Skolnick, J.; Walker, J. C.; Van Doren, S. R. *Biochemistry* **2007**, *46*, 2684–2696.
- (32) Noda, I.; Ozaki, Y. *Two-dimensional Correlation Spectroscopy - Applications in Vibrational and Optical Spectroscopy*; Wiley: West Sussex, England, 2004.
- (33) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. J. In *Protein NMR Spectroscopy*, Second ed.; Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M., Skelton, N. J., Eds.; Academic Press: Burlington, 2007; pp vii–x.
- (34) Cogliati, C.; Ragona, L.; D'Onofrio, M.; Günther, U.; Whittaker, S.; Ludwig, C.; Tomaselli, S.; Assfalg, M.; Molinari, H. *Chem. - Eur. J.* **2010**, *16*, 11300–11310.
- (35) Dreha-Kulaczewski, S.; Joseph, A. A.; Merboldt, K. D.; Ludwig, H. C.; Gartner, J.; Frahm, J. https://commons.wikimedia.org/wiki/File:Dreha-Kulaczewski_JNeurosci_CSF_flow_Supplementary_movie1.webm, 2014.
- (36) Dreha-Kulaczewski, S.; Joseph, A. A.; Merboldt, K. D.; Ludwig, H. C.; Gartner, J.; Frahm, J. *J. Neurosci.* **2015**, *35*, 2485–2491.
- (37) Joseph, A.; Kowallick, J. T.; Merboldt, K. D.; Voit, D.; Schaetz, S.; Zhang, S.; Sohns, J. M.; Lotz, J.; Frahm, J. *J. Magn. Reson. Imaging* **2014**, *40*, 206–213.
- (38) Lee, J. M.; Yoo, C. K.; Choi, S. W.; Vanrolleghem, P. A.; Lee, I. B. *Chem. Eng. Sci.* **2004**, *59*, 223–234.
- (39) Rennella, E.; Brutscher, B. *ChemPhysChem* **2013**, *14*, 3059–3070.
- (40) Rao, A. N.; Rodesch, C. K.; Grainger, D. W. *Anal. Chem.* **2012**, *84*, 9379–9387.
- (41) Teisseyre, T. Z.; Urban, J.; Halpern-Manners, N. W.; Chambers, S. D.; Bajaj, V. S.; Svec, F.; Pines, A. *Anal. Chem.* **2011**, *83*, 6004–6010.
- (42) Britton, M. M.; Bayley, P. M.; Howlett, P. C.; Davenport, A. J.; Forsyth, M. *J. Phys. Chem. Lett.* **2013**, *4*, 3019–3023.
- (43) Mizukami, S.; Takikawa, R.; Sugihara, F.; Hori, Y.; Tochio, H.; Walchli, M.; Shirakawa, M.; Kikuchi, K. *J. Am. Chem. Soc.* **2008**, *130*, 794–795.
- (44) Zhu, X.; Chi, X.; Chen, J.; Wang, L.; Wang, X.; Chen, Z.; Gao, J. *Anal. Chem.* **2015**, *87*, 8941–8948.