

The feasibility of parameterizing four-state equilibria using relaxation dispersion measurements

Pilong Li · Ilídio R. S. Martins · Michael K. Rosen

Received: 5 May 2011 / Accepted: 7 July 2011
© Springer Science+Business Media B.V. 2011

Abstract Coupled equilibria play important roles in controlling information flow in biochemical systems, including allosteric molecules and multidomain proteins. In the simplest case, two equilibria are coupled to produce four interconverting states. In this study, we assessed the feasibility of determining the degree of coupling between two equilibria in a four-state system via relaxation dispersion measurements. A major bottleneck in this effort is the lack of efficient approaches to data analysis. To this end, we designed a strategy to efficiently evaluate the smoothness of the target function surface (TFS). Using this approach, we found that the TFS is very rough when fitting benchmark CPMG data to all adjustable variables of the four-state equilibria. After constraining a portion of the adjustable variables, which can often be achieved through independent biochemical manipulation of the system, the smoothness of TFS improves dramatically, although it is still insufficient to pinpoint the solution. The four-state equilibria can be finally solved with further incorporation of independent chemical shift information that is readily available. We also used Monte Carlo simulations to evaluate how well each adjustable parameter can be determined in a large kinetic and thermodynamic parameter space and how much improvement can be achieved in defining the parameters through additional measurements. The results

show that in favorable conditions the combination of relaxation dispersion and biochemical manipulation allow the four-state equilibrium to be parameterized, and thus coupling strength between two processes to be determined.

Keywords Allostery · Allosteric coupling strength · Relaxation dispersion · Curve fitting · Four-state equilibrium

Introduction

Allostery is a crucial mode of regulation found in most, if not all, biological pathways (Gunasekaran et al. 2004; Kalodimos 2011; Tzeng and Kalodimos 2011). It refers to the responsiveness of protein activities to perturbations at remote sites on the same molecule or complex. For example, catalytic activities of enzymes or ligand binding activities of receptors are modulated by effector molecules upon association at distal sites (Buck et al. 2004; Fenton 2008; Goodey and Benkovic 2008; Gunasekaran et al. 2004; Kalodimos 2011; Leung and Rosen 2005; Monod et al. 1965; Tzeng and Kalodimos 2011). The magnitude of activity change upon effector binding is defined as allosteric coupling strength (Monod et al. 1965). Non-unitary coupling strengths indicate the presence of allostery. It is pivotal to determine allosteric coupling strengths in order to quantitatively understand the role of allostery in different biological pathways (Buck et al. 2004; Leung and Rosen 2005; Prehoda et al. 2000; Yu et al. 2010), to guide allosteric drug discovery (Acuner Ozbabacan et al. 2010; Kar et al. 2010; Peterson et al. 2004), and to assist design and engineering of new bio-inspired systems (Dueber et al. 2003; Leung et al. 2008; Lim 2002; Stratton and Loh 2011; Vallee-Belisle and Plaxco 2010; Wright et al. 2007).

P. Li (✉) · I. R. S. Martins · M. K. Rosen
Department of Biochemistry and Howard Hughes Medical
Institute, University of Texas Southwestern Medical Center,
5323 Harry Hines Boulevard, Dallas, TX 75390-8816, USA
e-mail: pilongli@mail.med.upenn.edu

I. R. S. Martins
Departamento de Bioquímica, Faculdade de Ciências e
Tecnologia da Universidade de Coimbra, Coimbra 3001-401,
Portugal

In a simplified yet representative view, allosteric proteins in their apo form possess intrinsic two-state equilibria (Fig. 1a), often between a low activity and a high activity state (Kern and Zuiderweg 2003; Monod et al. 1965). In the presence of effectors, allosteric systems transition to a four-state equilibrium representation (Bosco et al. 2010; Li et al. 2008) (Fig. 1b). In the ideality of full saturation by effector, the systems collapse into the right edge of the four-state thermodynamic box (Fig. 1c)—i.e., a new two-state equilibrium with different populations of species than in the free state. This model is consistent with data indicating that even in the absence of effectors, many allosteric proteins appreciably sample the active state, and that effectors shift the equilibrium from inactive to active (Kern et al. 2005; Leung and Rosen 2005; Loria et al. 2008). Of course inhibitory effectors can also shift systems from active to

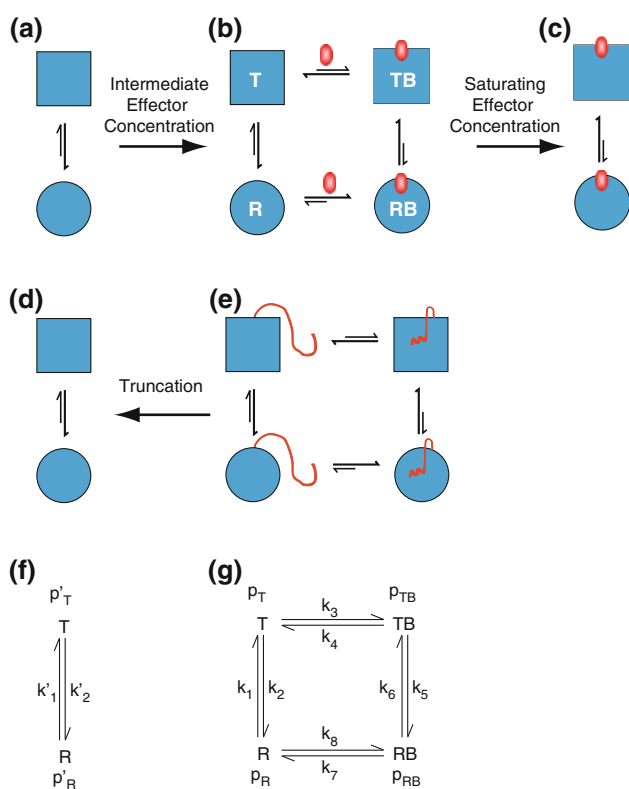


Fig. 1 Models of coupled equilibria. A minimal representation of allosteric systems requires two structural elements: a protein containing an allosteric site (in blue) and an effector (in red). In the absence of effector, allosteric proteins display intrinsic conformational equilibria between at least two states (a, d, f): T and R. In this framework, allostery arises from energetic coupling between the intrinsic and effector-binding equilibria that can be represented by a four-state equilibrium (b, e, g) between T, R, T-bound (TB), and R-bound (RB). In bimolecular systems (upper panels), coupling occurs in partially saturated conditions (b) and is overridden in the apo (a) and fully saturated cases (c). In unimolecular multi-domain systems, truncation of the effector module can be used to isolate the conformational equilibrium intrinsic to the allosteric module (d). Both bi- and uni-molecular systems can be similarly parameterized using a set of population (p) and kinetic constant (k) variables (f, g)

inactive via coupled equilibria favoring the inactive state (Yu et al. 2010). In this thermodynamic framework, the allosteric and binding equilibria are coupled, by a magnitude defined as the ratio of the equilibrium constants for the two-state equilibria of the apo-protein (Fig. 1a) and of the effector-saturated protein (Fig. 1c). Alternatively, coupling magnitude is the ratio of binding equilibrium constants of the effector for the inactive and active states of the protein (Leung and Rosen 2005; Monod et al. 1965).

A thermodynamically analogous situation can also exist in unimolecular systems. For example, in a recent study, we observed 2 μ s-ms timescale dynamic processes coexisting in the autoinhibited Dbl homology domain (DH) of the Rho/Rac1 guanine nucleotide exchange factor, Vav1, (This protein is referred to as helix-DH hereafter) (Li et al. 2008). One process is intrinsic to the DH domain, analogous to the intrinsic allosteric equilibrium depicted in Fig. 1a (Fig. 1d). The other process involves binding of an inhibitory helix, which is N-terminally adjacent to the DH domain, to the DH active site. The inhibitory helix is thermodynamically analogous to the allosteric effector, but acts in cis rather than in trans. Therefore the second process is analogous to the effector binding process in Fig. 1b (Fig. 1e). The thermodynamics and kinetics of the intrinsic process in the DH domain change upon perturbations to the helix-DH binding process, suggesting that the two equilibria are allosterically coupled (Li et al. 2008). A four-state model is required to depict the energetic landscape of the helix-DH protein (Fig. 1e).

NMR has proven to be a powerful means of quantitatively characterizing regulatory equilibria in macromolecules. The μ s-ms timescales of transitions between states and relatively large populations (>0.5%) of high-energy states often seen in proteins produce significant effects on transverse relaxation. These can be analyzed by Carr-Purcell-Meiboom-Gill (CPMG)-based measurements (Carr and Purcell 1954; Meiboom and Gill 1958) of relaxation dispersion to yield quantitative information on thermodynamic and kinetic aspects of allostery (Baldwin and Kay 2009; Kern and Zuiderweg 2003; Mittermaier and Kay 2009; Palmer et al. 2001). Such measurements are now routinely used to study two-state equilibria involving transitions on μ s-ms time scales (Baldwin and Kay 2009; Kovrigin et al. 2006; Mittermaier and Kay 2009; Palmer et al. 2001). However, in complex systems involving two coupled processes, the magnitude of coupling requires analysis of the four-state system directly (Fig. 1b, e). In recent studies, relaxation dispersion measurements have been used to characterize three-state systems (Eisenmesser et al. 2002; Grey et al. 2003; Korzhnev et al. 2004, 2005, 2006; Neudecker et al. 2006; Sugase et al. 2007; Tolkmachev et al. 2003). Solving three-state equilibria using CPMG data requires either relaxation dispersion measurements of a

large number of different coherences (Korzhnev et al. 2005; Neudecker et al. 2006), or orthogonal perturbations, such as temperature variations (Grey et al. 2003; Korzhnev et al. 2004, 2006) and ligand titrations (Eisenmesser et al. 2002; Sugase et al. 2007; Tolkatchev et al. 2003). However, solving four-state equilibria using relaxation dispersion remains largely uncharted territory.

In this study, we use computer simulation to assess the feasibility of parameterizing four-state equilibria using relaxation dispersion measurements. The basic procedure is to (1) synthesize exact CPMG data using input parameter sets under a four-state model, (2) add random noise to the exact data, (3) fit noise-incorporated data to the same model and extract the input parameters as the adjustable variables, and (4) evaluate agreement between the fitted and input values. Due to the large number of adjustable variables, and correlations between certain parameters, the target function for the curve fitting of a four-state model is likely to contain multiple local minima. Thus, we designed a strategy called target function minima mapping from

Materials and methods

Theoretical background

The evolution of single-quantum coherence in the presence of conformational equilibria under a CPMG pulse train can be treated according to (1):

$$\mathbf{M}(t) = [\exp(\mathbf{A}_-\delta) \exp(\mathbf{A}_+\delta) \exp(\mathbf{A}_+\delta) \exp(\mathbf{A}_-\delta)]^n \mathbf{M}(0), \tag{1}$$

where $t = 4n\delta$ with $2n$ the number of spin-echo trains within the constant time period, t ; $\mathbf{M}(t)$ is the magnetization vector after the constant relaxation time period; $\mathbf{M}(0)$ is the initial magnetization vector; \mathbf{A}_\pm are the evolution of magnetization matrix before (\mathbf{A}_-) and after (\mathbf{A}_+) a 180° pulse in a spin-echo train (Korzhnev et al. 2004). For a spin with four-state exchange as in Fig. 1g, $\mathbf{M}(t)$ and $\mathbf{M}(0)$ are $[\mathbf{M}_{TB}(t), \mathbf{M}_T(t), \mathbf{M}_R(t), \mathbf{M}_{RB}(t)]^T$, and $(p_{TB}, p_T, p_R, p_{RB})^T$, respectively. The corresponding evolution matrices, \mathbf{A}_\pm , are:

$$\mathbf{A}_\pm = \begin{bmatrix} -k_4 - k_5 - R_2^0 \pm i \times \Omega_{TB} & k_3 & 0 & k_6 \\ k_4 & -k_2 - k_3 - R_2^0 \pm i \times \Omega_T & k_1 & 0 \\ 0 & k_2 & -k_1 - k_8 - R_2^0 \pm i \times \Omega_R & k_7 \\ k_5 & 0 & k_8 & -k_6 - k_7 - R_2^0 \pm i \times \Omega_{RB} \end{bmatrix}, \tag{2}$$

grids of reduced dimension, coupled with logarithms of average normalized distance plotting (M2GRED/LAND) to evaluate the smoothness of the target function surface (TFS). Indeed, M2GRED/LAND results indicate that local minima are frequently found in the TFS of all adjustable variables of the four-state equilibria. The smoothness of TFS improves dramatically with constraints on a portion of the variables, which can be determined by studying relaxation dispersion of a two-state system that represents one arm of the four-state box (Fig. 1a, d). However, it is still insufficient to pinpoint the solution with these data alone. After further incorporating chemical shift information that is readily available during experiments but generally not used in curve fitting, the four-state equilibria can be finally solved. To guide actual experiments, we evaluated how much improvement different additional measurements can bring about using Monte Carlo simulation. We also evaluated how well each adjustable parameter can be determined in a large kinetic and thermodynamic parameter space to assess in which parameter space allosteric coupling strength can be reasonably quantified.

in which k_1 – k_8 are defined in Fig. 1g; i is the imaginary unit; Ω_{TB} , Ω_T , Ω_R , and Ω_{RB} represent chemical shifts of a spin at states TB, T, R, and RB, respectively; and R_2^0 is the intrinsic transverse relaxation rate constant in the absence of exchange (R_2^0 at all four states are assumed to be identical). Due to the thermodynamic constraints, knowledge of any 7 out of the 12 kinetic and thermodynamic parameters (four populations and eight rates) is sufficient to fully determine the remaining ones. By convention, k_{ex} represents the sum of forward and reverse rate constants: $k_{ex1} = k_1 + k_2$, $k_{ex2} = k_3 + k_4$, $k_{ex3} = k_5 + k_6$, and $k_{ex4} = k_7 + k_8$. Therefore, we use four kinetic (k_{ex1} – k_{ex4}) and three population (p_T , p_{TB} , and p_{RB}) parameters as the adjustable variables for the four-state model.

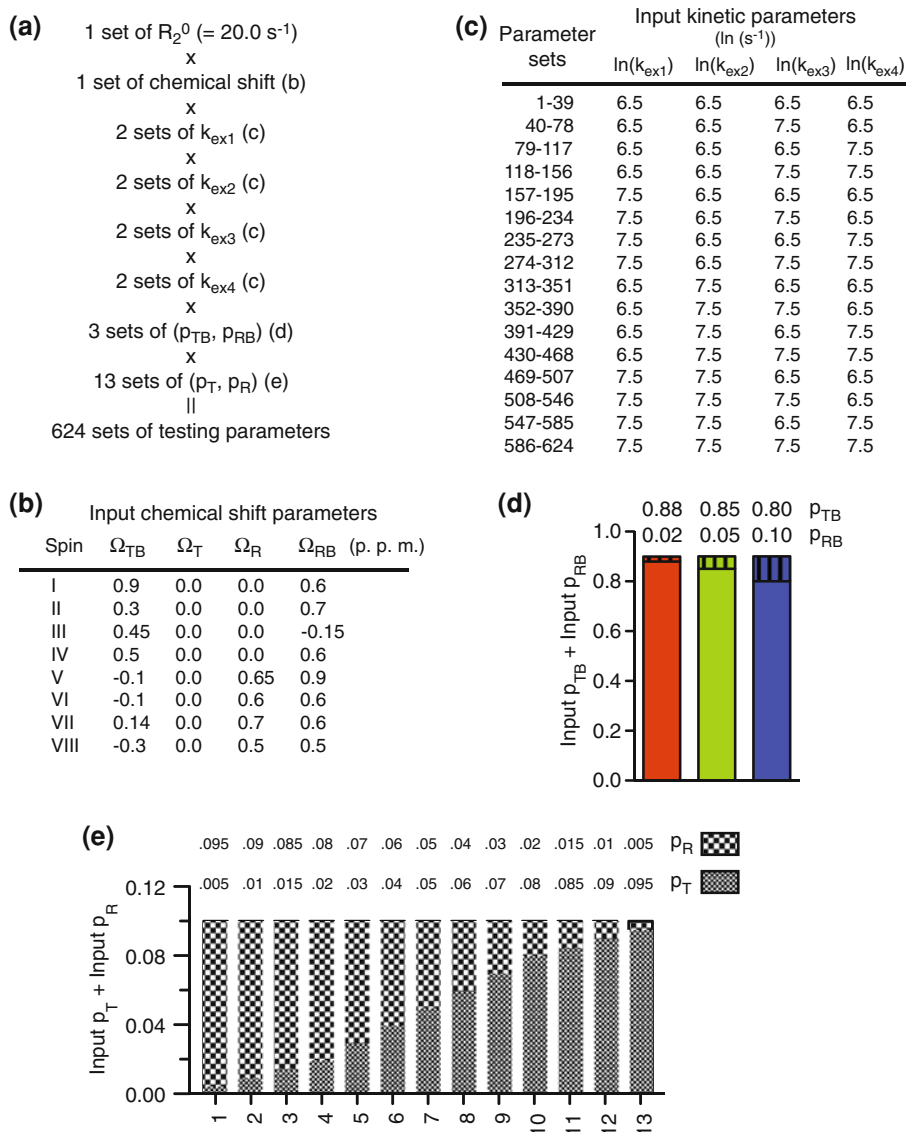
The parameter sets

We extensively studied the μ s–ms timescale dynamics of the helix-DH protein using CPMG-based relaxation dispersion measurements, and extracted the kinetic and thermodynamic parameters and chemical shifts of some spins

in the four states (Li et al. in preparation). For the current study, we constructed a large number of parameter sets based on the kinetic, thermodynamic, and methyl ^{13}C chemical shift information of helix-DH (Fig. 2). This study focuses on conformational equilibria in intermediate-to-fast exchange regimes, in which k_{ex} values are much larger than the intrinsic transverse relaxation rate constants (R_2^0) and the differences in R_2^0 of the different states can be neglected (Grey et al. 2003). Hence the R_2^0 values of all ^{13}C resonances are assumed to be identical, 20.0 s^{-1} (Fig. 2a). We note that this assumption is unlikely to hold in the analysis of ^{15}N relaxation dispersion data due to the significant chemical shift anisotropy for nitrogen nuclei. Approximate ^{13}C chemical shifts of eight methyl spins with dispersion amplitude (on 600-MHz instrument) between 10 and 40 s^{-1} were selected as the chemical shift parameter

set (Fig. 2b). Each of the four kinetic parameters takes either $e^{6.5} (\approx 665) \text{ s}^{-1}$ or $e^{7.5} (\approx 1,808) \text{ s}^{-1}$ (Fig. 2c). For the population parameters, the sum of p_{TB} and p_{RB} is 0.9 and that of p_{T} and p_{R} is 0.1, mimicking the finding that the open-closed equilibrium of the inhibitory helix is about 0.1:0.9 in helix-DH (Li et al. 2008) (Fig. 2d, e); three combinations of $(p_{\text{TB}}, p_{\text{RB}})$ are considered, in which the state TB is always the major state ($p_{\text{TB}} \geq 0.8$) (Fig. 2d); and thirteen combinations of $(p_{\text{T}}, p_{\text{R}})$, ranging from $p_{\text{T}} \gg p_{\text{R}}$ to $p_{\text{T}} \ll p_{\text{R}}$, are considered (Fig. 2e). In total, there are $624 (= 2 \times 2 \times 2 \times 2 \times 3 \times 13)$ parameter sets (Fig. 2a). The parameter set that most closely approximates the fitted variables in our analysis of helix-DH at 5°C was selected as the benchmark, in which $(p_{\text{TB}}, p_{\text{T}}, p_{\text{R}}, p_{\text{RB}}) = (0.85, 0.07, 0.03, 0.05)$, and $(k_{\text{ex}1}, k_{\text{ex}2}, k_{\text{ex}3}, k_{\text{ex}4}) = e^{(7.5, 6.5, 6.5, 6.5)} \text{ s}^{-1}$. We note that in this system the

Fig. 2 Parameter sets. **a** There are 624 parameter sets, in which (1) all R_2^0 s are 20.0 s^{-1} ; (2) one set of chemical shifts from eight spins is considered (**b**); (3) k_{ex} values are $e^{6.5} (\approx 665)$ or $e^{7.5} (\approx 1,808) \text{ s}^{-1}$ (**c**); (4) three sets of $(p_{\text{TB}}, p_{\text{RB}})$ are considered: (0.88, 0.02) (*red*), (0.85, 0.05) (*green*), and (0.80, 0.10) (*blue*) (**d**); and (5) 13 sets of $(p_{\text{T}}, p_{\text{R}})$ are considered (**e**)



effector is an allosteric inhibitor such that its binding shifts the system more towards the T state—i.e., $p_{TB}/p_{RB} > p_T/p_R > 1.0$. The benchmark parameter set was used to demonstrate the methodological innovations that allow us to solve the four-state equilibria of helix-DH. Monte Carlo simulation was used to evaluate the feasibility of solving four-state equilibria in all 624 parameter sets.

Generation of CPMG data

Synthetic CPMG relaxation dispersion data of four-state equilibria for a given parameter set were obtained based on the major species, TB, using:

$$R_{2,\text{eff}} = -\frac{1}{t} \times \ln \frac{M_{TB}(t)}{M_{TB}(0)}, \quad (3)$$

where t is the constant relaxation time period and $M_{TB}(t)$ and $M_{TB}(0)$ are the first elements of $\mathbf{M}(t)$ and $\mathbf{M}(0)$ for four-state equilibria, respectively. Noise-free 600- and 800-MHz single-quantum methyl ^{13}C CPMG data were generated using a program written in C modified from one kindly provided by Drs. Dmitry M. Korzhnev and Lewis E. Kay (University of Toronto). The resulting dispersion profiles $R_{2,\text{eff}}(v_{\text{CPMG}})$ per spin per field include 20 data points with v_{CPMG} varying from 50 to 1,000 Hz every 50 Hz. Each noise-free $R_{2,\text{eff}}$ data point was incorporated with a noise term randomly drawn from a Gaussian distribution of mean zero and standard deviation of $4\% \times R_{2,\text{eff}}$ to give rise to 4% noise-incorporated data. Given the availability of cryogenic probes, perdeuteration labeling, and sensitivity-enhanced NMR techniques, 4% is a practical noise level for many systems (Ishima and Torchia 2005; Lundstrom et al. 2007; Neudecker et al. 2006).

Curve fitting of synthetic CPMG data

The C-code provided by Drs. Korzhnev and Kay was also modified for global fitting of synthesized data using Levenberg–Marquardt nonlinear least-square minimization of the target function value (TFV):

$$\text{TFV}(\zeta) = \sum \frac{\left(R_{2,\text{eff}}^{\text{calc}}(\zeta) - R_{2,\text{eff}}\right)^2}{(4\% \times R_{2,\text{eff}})^2}, \quad (4)$$

where $R_{2,\text{eff}}$ are data points to be fitted, $R_{2,\text{eff}}^{\text{calc}}(\zeta)$ is calculated $R_{2,\text{eff}}$ given a set of adjustable variables, ζ represents the set of adjustable variables, and the summation is over all relaxation data points (Ishima and Torchia 2005; Korzhnev et al. 2004).

In curve fitting without constraints on parameters of the T-R edge, a Levenberg–Marquardt search was conducted from multiple initial points. The initial points were selected in order to cover the intermediate-to-fast exchange regime for

kinetic parameters and CPMG sensitive ranges for populations. These initial points were defined as follows: (1) $\ln(k_{\text{ex}})$ values varied from 5.0 to 8.0 with a step size of 0.2; p_{TB} varied from 0.5 to 0.76 with a step size of 0.02; (2) p_{RB} was constrained by p_{TB} , such that $p_{RB} = 0.8 - p_{TB}$; (3) all other adjustable parameters had only a single value: $p_T = 0.14$; $p_R = 0.06$; R_2^0 s were set at their input values; Ω_{RS} of all spins were equal to their input values; Ω_{TB} and Ω_{RB} of spins I–IV were set to be equal, and their values were arbitrarily set according to the relaxation dispersion amplitude; Ω_{TB} of spins V–VIII were equal to the corresponding Ω_T ; Ω_{RB} of spins V–VIII were equal to the corresponding Ω_R . In the process of curve fitting with constraints on parameters of the T-R edge, we did a Levenberg–Marquardt search from similar multiple initial points as above except that Ω_{RS} , k_{ex} , p_T , and p_R were no longer adjustable variables.

Monte Carlo simulation

Monte Carlo (MC) simulation was used to assess how well each adjustable parameter can be determined in the different parameter spaces (sections “[Solution determination using chemical shift constraints](#)” and “[MC simulation on all parameter sets](#)”) or with different amounts of measurements (section “[Improvement in fitted parameters by additional measurements](#)”). In each simulation, 4% random noise was added to synthesized, noise-free relaxation dispersion data. The noise-incorporated data set was subject to Levenberg–Marquardt optimization from input values. The resulting minimizer for each MC run was considered the solution for the current data set.

Results

The models of coupled equilibria

A symbolic representation of the pictorial two-state equilibria in Fig. 1a and d is shown in Fig. 1f. Following the nomenclature of the MWC model (Monod et al. 1963, 1965), the two states of the equilibrium intrinsic to the apo protein (Fig. 1a) or the biochemically isolated DH domain (Fig. 1d) are referred to as the T and R states, corresponding to the square and circle, respectively. The fractional populations for states T and R are p'_T and p'_R , respectively. p'_T and p'_R sum to one. k'_1 and k'_2 are two first-order microscopic rate constants.

A symbolic representation of the pictorial four-state equilibria from Fig. 1b and e is shown in Fig. 1g. Ligand binding to states T and R gives states TB and RB, respectively, which can also interconvert. These interactions produce a closed (as distinct from linear) four-state equilibrium (Fig. 1b). The model assumes that microscopically

the T–R transition and effector binding happen instantaneously and that diagonal transitions involving simultaneous changes in both processes are too rare to be observed (Woessner 1961). The fractional populations for states T, TB, RB, and R are p_T , p_{TB} , p_{RB} , and p_R , respectively. p_T , p_{TB} , p_{RB} , and p_R sum to one. For the unimolecular systems (Fig. 1e), k_1 – k_8 are eight first-order microscopic rate constants. For the bimolecular systems (Fig. 1b), k_3 and k_8 contain a term of free effector concentration (Palmer et al. 2001).

Design of a novel data analysis strategy

The first step in curve fitting is to establish a real-valued target function (TF) based on quantitative relationships between the data and the adjustable variables (\mathbf{x}). In the absence of noise, the true values of the adjustable variables (\mathbf{X}) is the global minimizer, a point \mathbf{x}^{**} such that

$$\text{TF}(\mathbf{x}^{**}) \leq \text{TF}(\mathbf{x})$$

for all \mathbf{x} . Random noise modulates the TF such that \mathbf{X} is no longer the global minimizer. More often, \mathbf{X} is not even a local minimizer, a point \mathbf{x}^* such that

$$\text{TF}(\mathbf{x}^*) \leq \text{TF}(\mathbf{x})$$

for all \mathbf{x} near \mathbf{x}^* . In the presence of random noise, the best one can expect from curve fitting is to find the local minimizer that is the closest to \mathbf{X} and use it to approximate \mathbf{X} . Therefore, in the second step of curve fitting, TFs are minimized to obtain as many local minimizers as possible using optimization algorithms. The next step is to determine which one, if any, of the local minimizers obtained is the solution—i.e., the minimizer closest to \mathbf{X} . In the absence of other orthogonal information, it is often assumed that the global minimizer is the solution. Whether this assumption holds depends on the roughness of the target function surface (TFS), which is dictated by the quality and quantity of the data available and by the complexity of the models. When the function is complex, there may be many local minima of comparable TFV, and the global minimizer may or may not be that closest to \mathbf{X} . Since there is no a priori information about \mathbf{X} , it is not justifiable to attribute the global minimizer, if obtainable, to the solution in scenarios where the noise-modulated TFS is rough. This is likely the case in solving the four-state equilibrium. Before attempting to solve such problems, it is necessary to probe the roughness of TFS to assess the feasibility of the fitting.

Minima mapping from grids of reduced dimensions: evaluating the feasibility of solving a four-state system

The four-state model is complicated [(1) and (2)] and the number of adjustable parameters used for fitting relaxation

dispersion data is large [(3) and (4)]. In any parameter set, there are 4 kinetic, 3 population, $3n$ chemical shift (3 for each spin, the fourth is set at zero since the chemical shift differences dictate relaxation dispersion), and $2nR_2^0$ variables, where n is the number of resonances in fast-to-intermediate exchange for which CPMG data of high quality can be obtained. The dimensionality of the system (d) increases with n . In our benchmark $n = 8$, corresponding to a total of 24 chemical shift, and $16R_2^0$ variables (Fig. 2). Since it is not practical to perform unconstrained probing of the roughness of the TFS for systems of such high dimensionality ($d = 47$), we devised a strategy to reduce the dimensionality of the search by using NMR information that is amenable to experimental perturbation.

Grid search is a commonly used method to probe and evaluate the TFS (Thisted 1988; Wen and Hsiao 2007). This approach involves setting up grids of all adjustable variables and evaluating the TFV at each grid point. Even if each parameter were to take only two different values, with the data here there are c.a. 2^{47} ($\approx 10^{14}$) grid points for the TFV calculation. Moreover, there are no intuitive ways to analyze the calculated TFV due to the high dimensionality. We designed a data analysis protocol to overcome these difficulties. First, the grid search dimension was dramatically reduced by fixing the initial values for most of the adjustable parameters and only varying the initial values of a few key adjustable parameters. Since R_2^0 variables are orthogonal to all other variables, it is easy to fit them (Ishima and Torchia 2005), and each of them was simply assigned with a reasonable initial guess. In many four-state systems, one two-state edge can be isolated by removing domains or by making point mutations that strongly skew one of the two equilibrium processes. Alternatively, both equilibria can be modulated together, generating new variants of the four-state system with different collections of rates and populations (Li et al. 2008). Often, chemical shift and relaxation dispersion analyses of these modified systems will provide qualitative estimates of the chemical shifts of each spin in the different states (Li et al. 2008; Yu et al. 2010). In our analyses here, initial values of all 24 chemical shifts were based on our chemical shift perturbation and relaxation dispersion analyses of a battery of mutant helix-DH proteins (Li et al. 2008). Initial values of all four k_{ex} s were set to be identical and varied between $e^{5.0}$ (≈ 148.4) s^{-1} and $e^{8.0}$ ($\approx 2,981$) s^{-1} , corresponding to slow and intermediate/fast exchange regimes, respectively. Initial values of p_T and p_R were given based on chemical shift (Li et al. 2008) and relaxation dispersion analyses of the helix-DH protein and its variants, and initial values for p_{TB} were varied (see “Materials and Methods” for details). These treatments reduce the dimensionality of the grid

search to 2,¹ one along p_{TB} and the other along the k_{ex} s. We note that all adjustable variables were, and must be, allowed to change during the actual optimization process; only the initial values of the adjustable variables at the start of optimization are fixed here.

A second modification was that instead of calculating the TFV at each grid point as done in standard grid search algorithms (Thisted 1988; Wen and Hsiao 2007), a minimum on the TFS was obtained from each grid point via the Levenberg–Marquardt nonlinear optimization algorithm. Theoretically, all minima of a TFS can be mapped if the optimization procedure is carried out from a sufficient number of well-designed grid points. If the number of distinct minima is high, it is not feasible to solve the four-state equilibrium unless further information is incorporated. This curve fitting strategy, modified from standard grid-search algorithms, is referred to as target function minima mapping from grids of reduced dimensions (M2GRED).

Logarithm of average normalized distance

To sufficiently map minima on the TFS, M2GRED analysis will generate a large number of minimizers: some of them are identical and others are distinct from one another. Relationships among them need to be evaluated to assess minima map of the TFS. However, these minimizers are often multi-dimensional vectors, in which there are different types of variables with utterly different absolute values. Therefore, it is usually not straightforward to assess the relationship between any pair of minimizers unless they are identical. To effectively evaluate the TFS, a quantity termed logarithm of average normalized distance (LAND) was designed to assess similarities among the multi-dimensional minimizers:

$$\begin{aligned}
 & \text{LAND}(M1 \leftrightarrow M2) \\
 & = \max \left(-1.5, \log_{10} \sqrt{\frac{\sum_{i=1}^n \frac{(M1(i)-M2(i))^2}{(|M1(i)|+|M2(i)|)^2}}{n}} \right),
 \end{aligned}
 \tag{5}$$

¹ Historically, we reduced the grid-search dimension to 2 for two purposes: (1) to reduce the computation time, and (2) to enable 3-D dimensional plotting of the minima against the two grid-searched dimensions for visualization.

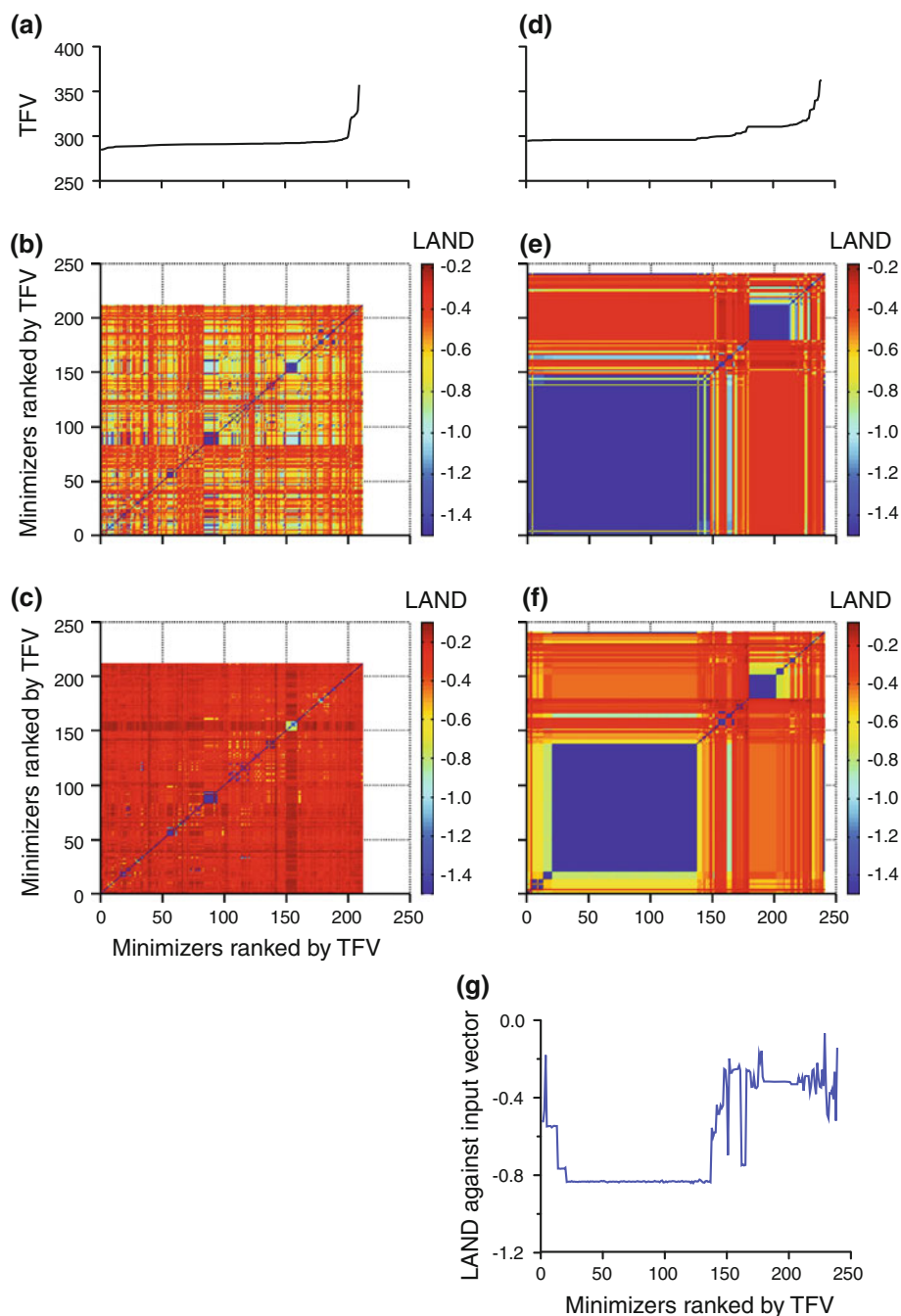
With the later advent of LAND plotting (see “[Logarithm of average normalized distance](#)” for details), the limit on grid-search dimension is not restricted to 2. In principle, initial values of any adjustable variables including p_T and p_R can vary across an appropriate range for better minima mapping of the TFS as long as computation time allows.

in which M1 and M2 are two minimizers being compared, i indices are the adjustable variables in the n dimensional vectors. The LAND is formulated based on the Euclidean distance between two n -dimensional vectors, M1 and M2, with two levels of normalization. For each element pair (M1(i), M2(i)), their squared deviation, (M1(i)-M2(i))², is divided by a normalization factor, (|M1(i)| + |M2(i)|)², which is always equal or larger than (M1(i)-M2(i))². The normalization enables co-existence of variables of different amplitudes in the vectors being compared. The sum of the normalized squared deviations is divided by the dimensionality of the vectors, n . This normalization enables comparison of vectors of different dimensionalities. The logarithm calculation is used to improve the sensitivity of the measure to smaller differences. For m minimizers under investigation, pair-wise LAND calculation results in an $m \times m$ symmetric matrix. The more similar two minimizers are, the smaller the LAND is. To avoid taking the logarithm of zero on the diagonal, all average normalized distances below or equal to $10^{-1.5}$ are set to $10^{-1.5}$, corresponding to a LAND value of -1.5 . The choice of this cut-off is arbitrary. Nevertheless, a LAND value of -1.5 roughly corresponds to average 6.3% ($= 2 \times 10^{-1.5}$) deviation per adjustable parameter. Such deviation is smaller than experimental deviation levels in many biochemical and biophysical applications. We note that LAND is designed to evaluate similarities between minimizers and does not take into account correlations between adjustable variables.

M2GRED/LAND analysis guide curve fitting of relaxation dispersion data

The availability of the novel data analysis strategy allows effective evaluation of the smoothness of the TFS when different amounts of orthogonal information are incorporated in curve fitting. M2GRED analysis was first performed on the benchmark relaxation dispersion data for all adjustable variables of the four-state model and 210 minimizers were obtained. Their TFVs range from 284.8 to 357.5 (Fig. 3a). Most of the minima possess distinct TFVs, indicating that the TFS possesses numerous local minima. A LAND matrix of a 7-dimensional vector containing 4 k_{ex} and 3 population values is plotted in Fig. 3b, and a LAND matrix of a 24-dimensional vector containing 24 chemical shifts is shown in Fig. 3c. LAND values decrease from red to blue (see color bars in Fig. 3). A blue cluster in the matrix indicates that the minimizers in it are similar to one another. As shown in Fig. 3b, c, there are very few blue clusters in either matrix, confirming that most of the grid points lead to distinct minima—i.e., the TFS is very rough. Since previous work on three-state systems showed that

Fig. 3 Curve fitting results of benchmark CPMG data. Fitting of the benchmark CPMG data to all adjustable variables of the four-state model (**a–c**), and all adjustable variables except those on the T-R edge (**d–g**). **a** TFV values of the 210 minimizers. **b** and **c** Logarithms of average normalized distances (LAND) plots of 7-dimensional kinetic/thermodynamic parameter vectors (**b**) and of 24-dimensional chemical shift vectors (**c**) of all minimizers in curve fitting without constraints from the T-R edge. **d** TFVs of the 240 minimizers obtained from curve fitting with constraints on the adjustable variables on the T-R edge. **e** and **f** LAND plots of 5-dimensional kinetic/thermodynamic parameter vectors (**e**) and of 16-dimensional chemical shift vectors (**f**) of all minimizers in curve fitting with constraints from the two-state protein. **g** LAND values of the 21-dimension vectors (5 kinetic/thermodynamic and 16 chemical shift parameters) of all minimizers against the input vector. The rankings of minimizers in all panels are based on ascending order of their TFVs. All panels share the same *x* axis. **d**, **e**, and **f** share the *y* axis with **a**, **b**, and **c**, respectively



simple data were not sufficient to define parameters (Korzhev et al. 2004, 2005, 2006; Neudecker et al. 2006), this level of roughness is expected after fitting the benchmark relaxation dispersion data to all adjustable parameters of the four-state model.

Clearly, orthogonal information is required to solve the benchmark four-state equilibrium. In many allosteric four-state systems, at least one edge of the four can be isolated through biochemical manipulations, to reduce the system to two states (Fig. 1a, d). For example, in helix-DH, the

intrinsic dynamics in the DH domain can be isolated through analysis of the isolated DH protein lacking the inhibitory helix (Li et al., in preparation). This then enables the adjustable parameters of the isolatable edge to be determined independently and used as constraints in the four-state curve fitting. To mimic this in the simulations here, the kinetic parameter (k_{ex1}), the population ratio (p_T/p_R), and the chemical shift differences ($|\Omega_T - \Omega_R|$) on the T-R edge were fixed in further curve fitting. As a result, the numbers of kinetic, population, and chemical shift adjustable variables

are reduced from 4, 3, and 24 to 3 (k_{ex2} , k_{ex3} , k_{ex4}), 2 (p_{TB} , p_{RB}), and 16, respectively.

M2GRED analysis was then carried out on the benchmark CPMG data set, and 240 minimizers were obtained for LAND analysis. TFVs of the minimizers are shown in Fig. 3d. The LAND matrix of a 5-dimensional kinetic/population vector and that of a 16-dimensional chemical shift vector are shown in Fig. 3e and f, respectively. In the former matrix, there is one major cluster of low TFV (blue; minimizers 1–150 in Fig. 3e), meaning that the majority of the minimizers have similar fitted kinetic and thermodynamic variables. In the chemical shift LAND plot, the low TFV cluster (minimizers 1–150) diverges into a large one and some minor ones (Fig. 3f), suggesting that chemical shifts suffer from greater degeneracy compared with kinetic and thermodynamic parameters.

To assess the relationship between the minimizers and the solution—i.e., the one closest to the true solution—the LAND values between the 21-dimensional vector (containing 3 kinetic, 2 thermodynamic, 16 chemical shift variables) of all 240 minimizers and that of the input vector were calculated (Fig. 3g). Minimizers in the large blue cluster on the chemical shift plot (minimizers 21–137) are the closest to the input vector (the basin on the curve in Fig. 3g), indicating that it is, by definition, the solution.

Solution determination using chemical shift constraints

It is important to note that in the chemical shift LAND matrix there are 20 minimizers that have TFVs (between 294.9 and 295.6) slightly lower than that of the solution (295.9) (Fig. 3d–f). Thus, the TFV alone is still insufficient to determine the complete solution for the four-state equilibria, even with constraints on adjustable variables on one edge. Since foreknowledge of the solution would be unavailable in experimental applications, further information is needed to pinpoint the solution. To this end, independent chemical shift information is available from HSQC measurements for each spin in both the two-state systems and the four-state systems. In principle, this information could be used as constraints in curve fitting. But this would require knowledge of the relative sign of the chemical shifts of a spin in all four states. While the relative sign can be determined for two-state systems (Skrynnikov et al. 2002), it is challenging to determine for four-state systems. Therefore, we use the chemical shift information to identify the solution from the resulting minimizers instead. For each spin, one can independently determine the difference between its chemical shift (CSD) in the two-state system and in the four-state system. If a minimizer is the solution, its fitted population and chemical shifts should recapitulate the experimentally measured CSD. To utilize this information, the CSD was calculated

for each spin using the input populations and chemical shifts according to (6):

$$CSD = \left| \frac{(p_{TB} \times \Omega_{TB} + p_T \times \Omega_T + p_R \times \Omega_R + p_{RB} \times \Omega_{RB}) - \frac{p_T \times \Omega_T + p_R \times \Omega_R}{p_T + p_R}}{p_T + p_R} \right|, \quad (6)$$

to mimic data available from HSQC spectra (CSD^{HSQC}). CSD was also calculated for each spin using the fitted populations and chemical shift variables in all minimizers based on the same equation (CSD^{CPMG}).

If a minimizer approximates the solution closely, the CSD^{CPMG} value for each spin determined from it will predict CSD^{HSQC} well—i.e., both the slope and the correlation coefficient (r^2) of the linear correlation for all spins in a minimizer will approach unity. We determined the CSD^{CPMG} and CSD^{HSQC} values for the eight spins in each of the 240 minimizers and calculated the slopes and r^2 of the resulting 240 linear correlations (Fig. 4a). A total of 117 minimizers have r^2 values greater than 0.94 (vertical dotted line) (0.947–0.952) and slopes close to unity (horizontal dotted line) (0.945) (inset of Fig. 4a). The 117 minimizers coincide exactly with the large blue cluster of the chemical shift plot—i.e., the solution (Fig. 3f). The plot of CSD^{HSQC} versus CSD^{CPMG} of the solution is shown in Fig. 4b. There are also 20 minimizers that have TFVs slightly lower than that of the solution (Fig. 3f). The first four of these minimizers have either r^2 less than 0.8 or slope less than 0.6 and are therefore out of the axis limits of Fig. 4a. Minimizers 5–20 have r^2 between 0.857 and 0.866, and slope between 0.908 and 0.918 (red dotted circle in Fig. 4a), and are clearly distinguishable from the cluster corresponding to the solution. These results collectively suggest that CSD analysis in conjunction with the TFV is sufficient to identify the solution from all minimizers obtained for the benchmark parameter set.

The kinetic and thermodynamic parameters in the solution are shown together with their input values in Table 1. Among the five adjustable kinetic and thermodynamic variables, k_{ex3} and p_{TB} are accurately (close to their input values) and precisely (small uncertainties) determined; k_{ex2} is determined with less accuracy, while k_{ex4} and p_{RB} are determined with less precision. Monte Carlo (MC) simulation was used to assess which adjustable parameters can be faithfully fitted and which cannot in the presence of random noise. In each simulation, 4% random noise was incorporated into the noise-free benchmark data set. Ideally M2GRED/LAND analysis should be employed to find the solution for each data set. However, this is extremely time-consuming for a large number of simulations. Instead, the minimizer obtained from the input parameter vector as the initial guess was assumed to be the solution. A total of 1,000 MC simulations were carried out.

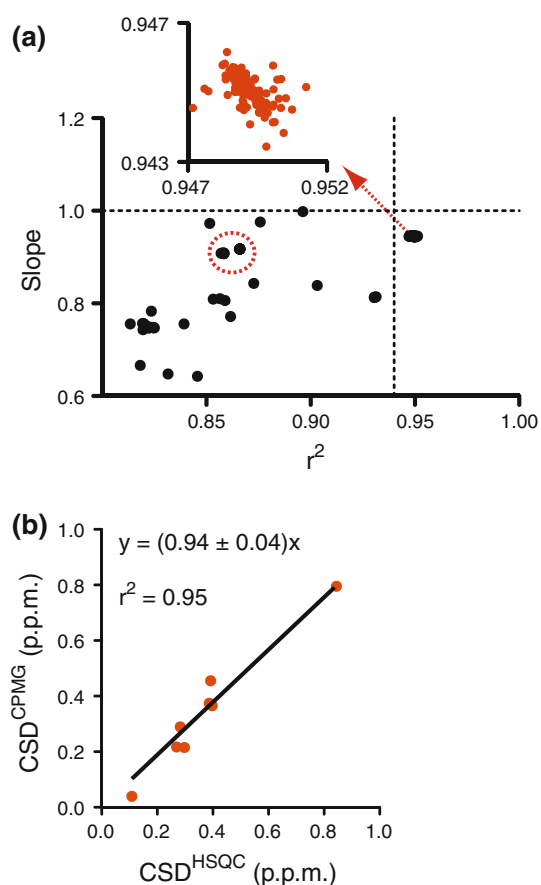


Fig. 4 Determining the solution using chemical shift differences. **a** Plot of the slopes and the correlation coefficients (r^2) of the linear correlations of chemical shift perturbations calculated from the minimizers (CSD^{CPMG}) with chemical shift perturbations obtained from the input parameters (CSD^{HSQC}). A total of 56 points, including minimizers 1–4 in Fig. 3g, have either r^2 less than 0.8 or slope less than 0.6 and are therefore out of the axis limits and not shown. The data points for minimizers 5–20 in Fig. 3g are within the dotted red circle. The dotted vertical line marks 0.94 on the x axis, and the dotted horizontal line marks the unit on the y axis. There are 117 data points, corresponding to minimizers 21–137 in Fig. 3g, with r^2 larger than 0.94, shown in the inset of **a**. **b** The plot of CSD^{CPMG} from the solution (minimizers 21–137) versus CSD^{HSQC}

The means and standard deviations of fitted kinetic/thermodynamic variables are shown in Table 1. Among them, $k_{\text{ex}3}$ and p_{TB} are accurately and precisely determined as they contribute significantly to relaxation dispersion, $k_{\text{ex}2}$ and p_{RB} are reasonably defined, and $k_{\text{ex}4}$ is poorly determined presumably due to its marginal contribution to relaxation owing to the low-populated R-RB edge ($p_{\text{R}} = 0.03$, $p_{\text{RB}} = 0.05$). Nevertheless, the MC simulation results are statistically indistinguishable from the solution (Table 1).

In summary, when combined with relaxation dispersion data acquired at two fields, CSD information and constraints of variables on one edge are sufficient to determine the solution of the benchmark four-state equilibrium in the

Table 1 Summary of the Solution

	Input ^a	Solution ^b	MC simulation ^c
$k_{\text{ex}2}^{\text{d}}$	6.5	6.0 ± 0.3	6.3 ± 0.5
$k_{\text{ex}3}^{\text{d}}$	6.5	6.7 ± 0.2	6.4 ± 0.5
$k_{\text{ex}4}^{\text{d}}$	6.5	6.8 ± 0.9	6.5 ± 1.8
p_{TB}	0.85	0.84 ± 0.01	0.84 ± 0.02
p_{RB}	0.05	0.043 ± 0.017	0.059 ± 0.023

^a Values used for data generation

^b Values and their uncertainties of the solution

^c Means and SD of the fitted parameters from MC simulation

^d Values for $k_{\text{ex}2}$ – $k_{\text{ex}4}$ are natural logarithms of values in s^{-1}

presence of experimentally reasonable noise. We note that chemical shift analysis for slow exchange scenarios can help identify solutions for chemical shifts but not for population values since chemical shifts do not provide thermodynamic information in this exchange regime. This information can in principle be extracted from relative peak volumes. However, due to line broadening and/or population bias, it is often difficult to rely on these measurements for extraction of thermodynamic information. Further analyses are necessary to assess the value of independent chemical shift information in identifying the solution of a four-state equilibrium in slow exchange.

Improvement in fitted parameters by additional measurements

It is instructive to assess how different types of additional data affect the quality of the fitting, as a guide to experimental design. We compared the benefits of repeating the measurements at 600 and 800 MHz to those of adding 900 MHz data to the benchmark. In the former case, the adjustable parameters remain unchanged. In the latter, there are 8 additional R_0^2 values due to the introduction of a new field. The two new data sets were subjected to 1,000 MC simulations at a 4% noise level. In both cases the incorporation of additional data modestly improved both the accuracy and the precision of fitted kinetic and thermodynamic variables (data not shown). The two already well-determined kinetic parameters, $k_{\text{ex}2}$ and $k_{\text{ex}3}$, show the largest relative improvement. For example, the standard deviation of $k_{\text{ex}2}$ decreases more than 60% when 900 MHz data are incorporated. The two most poorly determined parameters, $k_{\text{ex}4}$ and p_{RB} , only show slight improvement in their precision upon both treatments. Overall, incorporation of 900-MHz data is somewhat superior to repeating measurements at the two lower fields since this information gives larger improvements of precision in 4 out of 5 kinetic/thermodynamic variables.

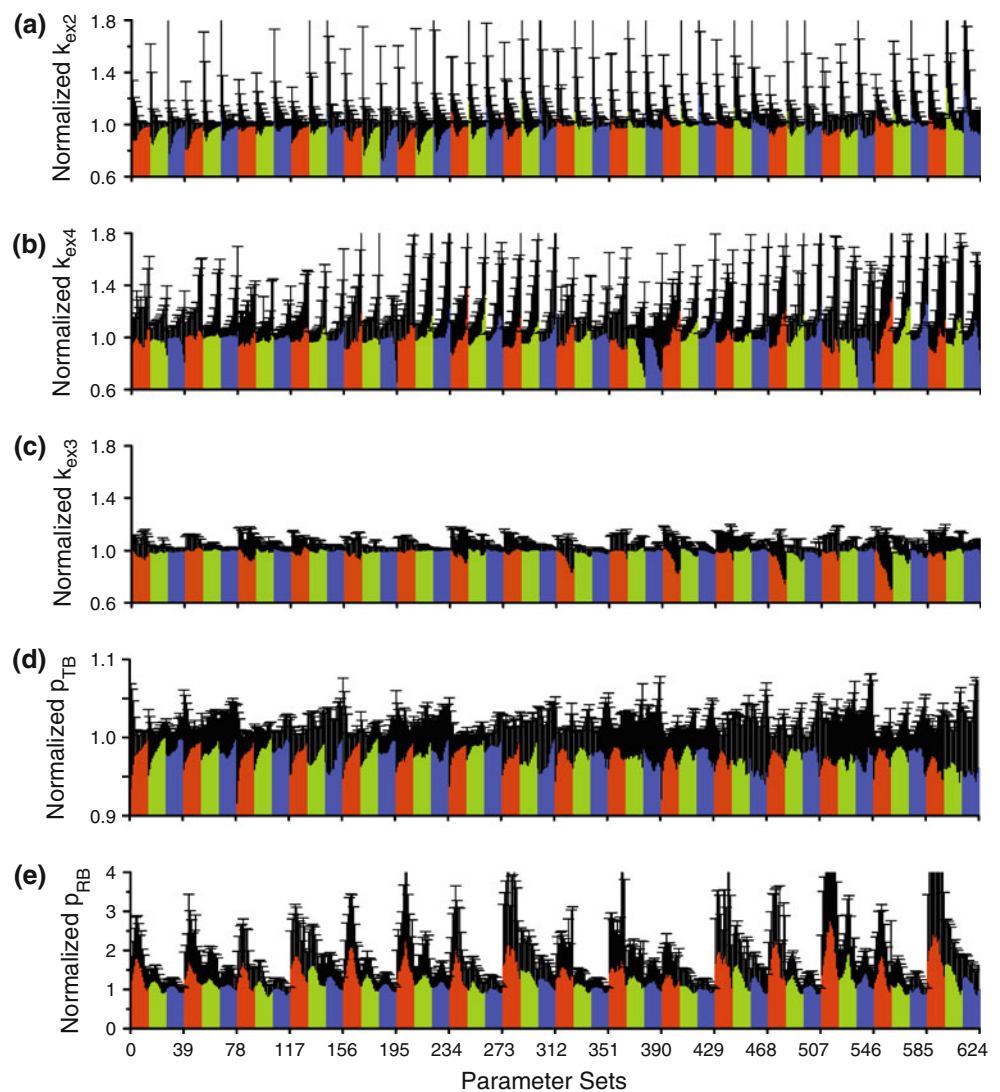
MC simulation on all parameter sets

In order to more comprehensively survey the parameter space in which it is feasible to solve a four-state equilibrium, we tested the fitting for all 624 parameter sets described in “The parameter sets” using MC simulation analysis. For each parameter set, we added 4% noise to synthetic 600- and 800-MHz CPMG data and carried out one hundred MC simulations. In each case, k_{ex1} and the p_T/p_R ratio were fixed; fitting was initiated with the starting parameter set and the resulting minimizer is assumed to be the solution based on previous observation (data not shown). The means and standard deviations of the normalized fitted kinetic and thermodynamic variables (k_{ex2} , k_{ex4} , k_{ex3} , p_{TB} , and p_{RB}) are shown in Figs. 5 and 6. Using normalized values one can readily assess the accuracy (deviation of the normalized mean from 1) and precision (size of the normalized standard deviation) of the fitting

across a large range of absolute values. For each adjustable variable, we plotted fitted values from all 624 parameter sets sequentially organized as described as follows: within each 13-column colored sector, all the input values remain constant except for input p_R (p_T) values which decrease (increase) from left to right, following the order in Fig. 2e; different colors represent different input p_{TB} (p_{RB}) values, which decrease (increase) from red to blue sectors as in Fig. 2d; each of the tri-colored (red, green, blue) super-sectors are each composed of a total of 39 columns, and correspond to a unique set of input k_{ex1} , k_{ex2} , k_{ex3} , and k_{ex4} values and are sequentially organized following the order (from top to bottom) in Fig. 2c.

In curve fitting, a variable can only be well defined if it contributes significantly to the measurements. As shown in Figs. 5a and 6a, k_{ex2} is reasonably determined in the later part of each colored sector, corresponding to conditions where p_T is higher, and the T-TB edge thus makes a larger

Fig. 5 Fitted kinetic and thermodynamic variables. Fitted k_{ex2} (a), k_{ex4} (b), k_{ex3} (c), p_{TB} (d), and p_{RB} (e) are shown for all 624 parameter sets. Each parameter is normalized to its input value. The error bars are the SD of the normalized fitted parameters. In each panel, there are 16 red, 16 green, and 16 blue sectors, and the color of each sector codes for (p_{TB} , p_{RB}) as in Fig. 2d. Each of the 48 (16 red + 16 green + 16 blue) sectors contains 13 individual bars. The 13 bars sequentially correspond to 13 (p_T , p_R) in Fig. 2e. The horizontal sequence from left to right of 16 sectors of common color codes for the vertical sequence from top to bottom of the kinetic parameter combinations in Fig. 2c. There are a total of 624 individual bars in each panel, corresponding to the 624 kinetic/thermodynamic parameter sets



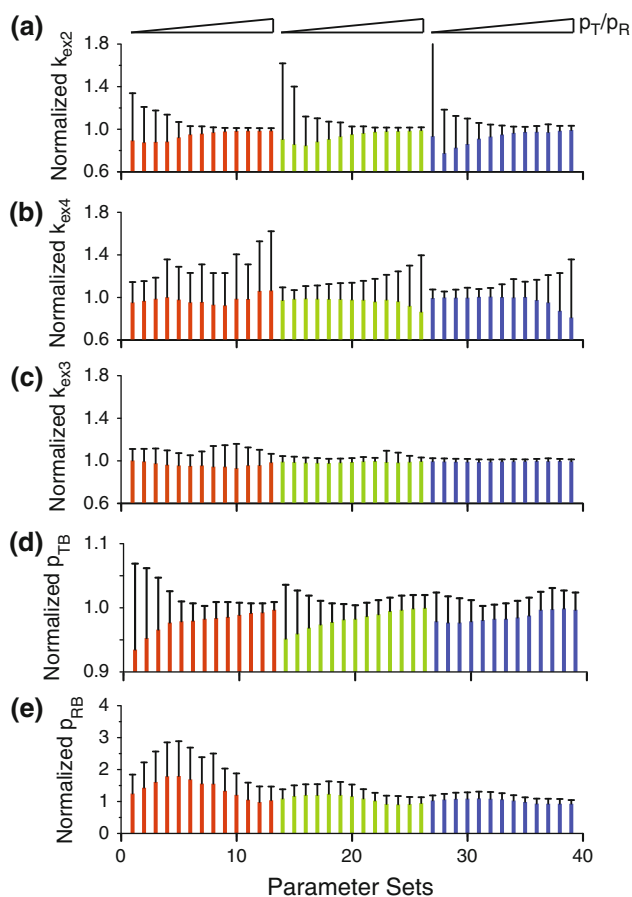


Fig. 6 Fitted kinetic and thermodynamic parameters. For clarity, fitted k_{ex2} (a), k_{ex4} (b), k_{ex3} (c), p_{TB} (d), and p_{RB} (e) of the first 39 parameter sets in Fig. 5 are shown. The three wedges above a show ascending order of the ratios between p_T to p_R in each color sector (see Fig. 2e for details)

contributions to relaxation. In contrast, Figs. 5b and 6b show that k_{ex4} is reasonably determined in the early part of each of the colored sectors. In these parameter sets, p_R is higher, resulting in larger contributions to relaxation from the R-RB edge. Owing to the opposite responses to alterations in p_T/p_R , there is a negative correlation between the certainty of k_{ex2} and k_{ex4} (compare Figs. 5a–b, 6a–b). Both the accuracy and precision of the fitted k_{ex3} are reduced when $(p_{TB}, p_{RB}) = (0.88, 0.02)$ and p_T is high as shown in 9th, 11th, 13th, and 15th red sectors (parameter sets 313–325, 391–403, 469–481, 547–559, respectively) in Fig. 5c. This is because, as p_{RB} becomes small and p_T becomes high, the T-TB edge replaces the TB-RB edge as the relaxation-dominating edge. k_{ex3} is very well determined when the TB-RB edge is less skewed—i.e., $(p_{TB}, p_{RB}) = (0.85, 0.05)$ or $(0.8, 0.1)$ (green and blue sectors in Figs. 5c and 6c)—as the edge dominates relaxation in these cases. In principle, an edge can become relaxation-dominating if its k_{ex} value switches from the fast to intermediate

exchange regime. However, there are no such examples in the current parameter sets.

As far as adjustable population variables are concerned, p_{TB} is very well determined since it is the major species and always on edges that contribute to relaxation significantly (Figs. 5d and 6d); p_{RB} is better determined if the TB-RB edge is less skewed [blue sectors of Fig. 5e, $(p_{TB}, p_{RB}) = (0.8, 0.1)$], and it tends to be over-estimated if the TB-RB edge is severely skewed [red sectors of Fig. 5e, $(p_{TB}, p_{RB}) = (0.88, 0.02)$]. Therefore, the ratio of p_{TB}/p_{RB} tends to be accurately estimated when the TB-RB edge is less skewed, and underestimated when the TB-RB edge is severely skewed.

In principle there are two ways to calculate the coupling between the equilibria in the four-state box: $(p_T/p_{TB})/(p_R/p_{RB})$ or $(p_T/p_R)/(p_{TB}/p_{RB})$. However, since the ratio, p_T/p_R , is determined in our approach through a biochemically-created two-state system, it can be quantified most accurately. This speaks to calculating coupling through the latter measure. From the analysis in the preceding paragraph, the ratio $(p_T/p_R)/(p_{TB}/p_{RB})$ can be most accurately determined when the TB-RB edge is less skewed, but will have systematic deviations if the TB-RB edge is severely skewed. Nevertheless, the systematic deviation may be especially significant for signaling proteins, in which the coupling strengths tend to be small, ranging from a several-fold to tens of folds (Hantschel et al. 2003; Lietha et al. 2007; Moarefi et al. 1997; Pluskey et al. 1995; Yohe et al. 2008; Yu et al. 2010).

Discussion

Coupled equilibria play important roles in controlling information flow in biochemical systems. The most representative class of such systems is allosteric proteins, which have been found in most biological pathways (Gunasekaran et al. 2004; Tzeng and Kalodimos 2011). Allosteric proteins in their apo form possess intrinsic dynamic equilibria (Fig. 1a), often between low activity and high activity states (Kern and Zuiderweg 2003; Monod et al. 1965). Effectors control the activities of allosteric proteins by modulating this intrinsic equilibrium upon binding to remote sites (Bosco et al. 2010; Buck et al. 2004; Leung and Rosen 2005; Li et al. 2008). Coupling between the intrinsic equilibrium and the effector binding equilibrium is the essence of regulation in allosteric proteins (Bosco et al. 2010; Buck et al. 2004; Leung and Rosen 2005; Li et al. 2008).

It is increasingly appreciated that coupled equilibria are also prevalent in complex, multi-domain proteins and play critical role in the regulation of this class of molecules (Faraldo-Gomez and Roux 2007; Levinson et al. 2006;

Masterson et al. 2008; Sondermann et al. 2004; Vogtherr et al. 2006; Yohe et al. 2008; Young et al. 2001; Yu et al. 2010). There, activity often appears to be controlled by a series of equilibria involving binding interactions between distinct modular elements. Thermodynamic coupling between these interactions enables activity (enzymatic and/or binding to other molecules) to be strongly suppressed, even though the individual equilibria may not be strongly biased. For example, we recently demonstrated that in the multidomain signaling protein Vav1, the intramolecular binding of an autoinhibitory helix into the enzymatic active site of the protein occurs with an equilibrium constant of ~ 10 . But this equilibrium is shifted tenfold further toward the bound state by additional contacts between the CH and PH domains of the molecule (Li et al. 2008; Yu et al. 2010). Thus, the helix-active site and CH–PH equilibria are coupled tenfold, and together provide strong (100-fold) suppression of enzymatic activity. It is the coupling between the two equilibria that enables one process to communicate with the other.

Deciphering coupling strengths is pivotal in fully understanding the regulation of allosteric proteins and other multi-domain proteins. Previously, we use NMR spectroscopic analysis of the different biochemical entities (mutant, truncated, or other biochemically modified proteins) to directly dissect complicated thermodynamic equilibria and quantify coupling strengths among multiple equilibria (Li et al. 2008; Yu et al. 2010). This work relied on chemical shift measurements to quantify populations of states, in order to bypass the difficulties inherent in analysis of relaxation dispersion in complex systems. Here we demonstrate that an approach based on relaxation dispersion measurements can complement these methods and extend determination of coupling strengths to many systems in signal transduction pathways. In addition to measuring populations, the relaxation dispersion approach can also provide information on rates of interconversion between states, data not available in purely chemical shift based approaches. Our analyses relied on a novel data analysis strategy, called M2GRED/LAND, to probe the smoothness of the TFS with different amounts of orthogonal information used in fitting CPMG data to the four-state model (Fig. 1g). In multi-domain systems, one edge of the four-state thermodynamic box can generally be isolated by removing domains or by making point mutations that strongly perturb certain equilibria, as demonstrated in a recent study (Li et al. 2008). The biochemically/genetically isolated edge becomes a two-state system amenable to routine relaxation dispersion analysis. Information acquired from the two-state system enables constraints on a quarter of the adjustable variables of the four-state model. M2GRED/LAND analysis suggests that incorporation of the information from the two-state

system, indeed, dramatically reduces the roughness of TFS. Further NMR spectroscopic analysis of the different biochemical entities offer orthogonal chemical shift information to facilitate solution of four-state equilibria (Li et al. 2008; Yu et al. 2010).

Acknowledgments This work is submitted in honor of the 50th birthday of Dr. Lewis Kay, who inspired us to always aim higher. We thank Drs. Dmitry M. Korzhnev and Lewis Kay for providing data analysis software. We thank Dr. Xiaolan Yao for a critical reading of the manuscript. I. M. was supported by fellowships from the Foundation for Science and Technology (Portugal) and the Luso-American Development Foundation (Portugal). This work was supported by NIH grant GM066930 to M. K. R. and by the Howard Hughes Medical Institute.

References

- Acuner Ozbabacan SE, Gursoy A, Keskin O, Nussinov R (2010) Conformational ensembles, signal transduction and residue hot spots: application to drug discovery. *Curr Opin Drug Discov Devel* 13:527–537
- Baldwin AJ, Kay LE (2009) NMR spectroscopy brings invisible protein states into focus. *Nat Chem Biol* 5:808–814
- Bosco DA, Eisenmesser EZ, Clarkson MW, Wolf-Watz M, Labeikovsky W, Millet O, Kern D (2010) Dissecting the microscopic steps of the cyclophilin A enzymatic cycle on the biological HIV-1 capsid substrate by NMR. *J Mol Biol* 403:723–738
- Buck M, Xu W, Rosen MK (2004) A two-state allosteric model for autoinhibition rationalizes WASP signal integration and targeting. *J Mol Biol* 338:271–285
- Carr HY, Purcell EM (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys Rev* 94:630–638
- Dueber JE, Yeh BJ, Chak K, Lim WA (2003) Reprogramming control of an allosteric signaling switch through modular recombination. *Science* (New York, NY) 301:1904–1908
- Eisenmesser EZ, Bosco DA, Akke M, Kern D (2002) Enzyme dynamics during catalysis. *Science* 295:1520–1523
- Faraldo-Gomez JD, Roux B (2007) On the importance of a funneled energy landscape for the assembly and regulation of multidomain Src tyrosine kinases. *Proc Natl Acad Sci USA* 104:13643–13648
- Fenton AW (2008) Allostery: an illustrated definition for the ‘second secret of life’. *Trends Biochem Sci* 33:420–425
- Goodey NM, Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. *Nat Chem Biol* 4:474–482
- Grey MJ, Wang C, Palmer AG III (2003) Disulfide bond isomerization in basic pancreatic trypsin inhibitor: multisite chemical exchange quantified by CPMG relaxation dispersion and chemical shift modeling. *J Am Chem Soc* 125:14324–14335
- Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins* 57:433–443
- Hantschel O, Nagar B, Guettler S, Kretzschmar J, Dorey K, Kuriyan J, Superti-Furga G (2003) A myristoyl/phosphotyrosine switch regulates c-Abl. *Cell* 112:845–857
- Ishima R, Torchia DA (2005) Error estimation and global fitting in transverse-relaxation dispersion experiments to determine chemical-exchange parameters. *J Biomol NMR* 32:41–54
- Kalodimos CG (2011) NMR reveals novel mechanisms of protein activity regulation. *Protein Sci* 20:773–782
- Kar G, Keskin O, Gursoy A, Nussinov R (2010) Allostery and population shift in drug discovery. *Curr Opin Pharmacol* 10:715–722

- Kern D, Zuiderweg ER (2003) The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* 13:748–757
- Kern D, Eisenmesser EZ, Wolf-Watz M (2005) Enzyme dynamics during catalysis measured by NMR spectroscopy. *Methods Enzymol* 394:507–524
- Korzhev DM, Salvatella X, Vendruscolo M, Di Nardo AA, Davidson AR, Dobson CM, Kay LE (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430:586–590
- Korzhev DM, Neudecker P, Mittermaier A, Orekhov VY, Kay LE (2005) Multiple-site exchange in proteins studied with a suite of six NMR relaxation dispersion experiments: an application to the folding of a Fyn SH3 domain mutant. *J Am Chem Soc* 127:15602–15611
- Korzhev DM, Bezsonova I, Evanics F, Taulier N, Zhou Z, Bai Y, Chalikian TV, Prosser RS, Kay LE (2006) Probing the transition state ensemble of a protein folding reaction by pressure-dependent NMR relaxation dispersion. *J Am Chem Soc* 128:5262–5269
- Kovrigina EL, Kempf JG, Grey MJ, Loria JP (2006) Faithful estimation of dynamics parameters from CPMG relaxation dispersion measurements. *J Magn Reson* 180:93–104
- Leung DW, Rosen MK (2005) The nucleotide switch in Cdc42 modulates coupling between the GTPase-binding and allosteric equilibria of Wiskott-Aldrich syndrome protein. *Proc Natl Acad Sci USA* 102:5685–5690
- Leung DW, Otomo C, Chory J, Rosen MK (2008) Genetically encoded photoswitching of actin assembly through the Cdc42-WASP-Arp2/3 complex pathway. *Proc Natl Acad Sci USA* 105:12797–12802
- Levinson NM, Kuchment O, Shen K, Young MA, Koldobskiy M, Karplus M, Cole PA, Kuriyan J (2006) A Src-like inactive conformation in the abl tyrosine kinase domain. *PLoS Biol* 4:e144
- Li P, Martins IR, Amarasinghe GK, Rosen MK (2008) Internal dynamics control activation and activity of the autoinhibited Vav DH domain. *Nat Struct Mol Biol* 15:613–618
- Lietha D, Cai X, Ceccarelli DF, Li Y, Schaller MD, Eck MJ (2007) Structural basis for the autoinhibition of focal adhesion kinase. *Cell* 129:1177–1187
- Lim WA (2002) The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr Opin Struct Biol* 12:61–68
- Loria JP, Berlow RB, Watt ED (2008) Characterization of enzyme motions by solution NMR relaxation dispersion. *Acc Chem Res* 41:214–221
- Lundstrom P, Vallurupalli P, Religa TL, Dahlquist FW, Kay LE (2007) A single-quantum methyl ¹³C-relaxation dispersion experiment with improved sensitivity. *J Biomol NMR* 38:79–88
- Masterson LR, Mascioni A, Traaseth NJ, Taylor SS, Veglia G (2008) Allosteric cooperativity in protein kinase A. *Proc Natl Acad Sci USA* 105:506–511
- Meiboom S, Gill D (1958) Modified spin-echo method for measuring nuclear relaxation times. *Rev Sci Instrum* 29:688–691
- Mittermaier AK, Kay LE (2009) Observing biological dynamics at atomic resolution using NMR. *Trends Biochem Sci* 34:601–611
- Moarefi I, LaFevre-Bernt M, Sicheri F, Huse M, Lee CH, Kuriyan J, Miller WT (1997) Activation of the Src-family tyrosine kinase Hck by SH3 domain displacement. *Nature* 385:650–653
- Monod J, Changeux JP, Jacob F (1963) Allosteric proteins and cellular control systems. *J Mol Biol* 6:306–329
- Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: a plausible model. *J Mol Biol* 12:88–118
- Neudecker P, Korzhnev DM, Kay LE (2006) Assessment of the effects of increased relaxation dispersion data on the extraction of 3-site exchange parameters characterizing the unfolding of an SH3 domain. *J Biomol NMR* 34:129–135
- Palmer AG III, Kroenke CD, Loria JP (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* 339:204–238
- Peterson JR, Bickford LC, Morgan D, Kim AS, Ouerfelli O, Kirschner MW, Rosen MK (2004) Chemical inhibition of N-WASP by stabilization of a native autoinhibited conformation. *Nat Struct Mol Biol* 11:747–755
- Pluskey S, Wandless TJ, Walsh CT, Shoelson SE (1995) Potent stimulation of SH-PTP2 phosphatase activity by simultaneous occupancy of both SH2 domains. *J Biol Chem* 270:2897–2900
- Prehoda KE, Scott JA, Mullins RD, Lim WA (2000) Integration of multiple signals through cooperative regulation of the N-WASP-Arp2/3 complex. *Science* 290:801–806
- Skrynnikov NR, Dahlquist FW, Kay LE (2002) Reconstructing NMR spectra of “invisible” excited protein states using HSQC and HMQC experiments. *J Am Chem Soc* 124:12352–12360
- Sondermann H, Soisson SM, Boykevich S, Yang SS, Bar-Sagi D, Kuriyan J (2004) Structural analysis of autoinhibition in the Ras activator Son of sevenless. *Cell* 119:393–405
- Stratton MM, Loh SN (2011) Converting a protein into a switch for biosensing and functional regulation. *Protein Sci* 20:19–29
- Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447:1021–1025
- Thisted RA (1988) Elements of statistical computing. Chapman and Hall, London
- Tolkachev D, Xu P, Ni F (2003) Probing the kinetic landscape of transient peptide-protein interactions by use of peptide (¹⁵N) NMR relaxation dispersion spectroscopy: binding of an anti-thrombin peptide to human prothrombin. *J Am Chem Soc* 125:12432–12442
- Tzeng SR, Kalodimos CG (2011) Protein dynamics and allostery: an NMR view. *Curr Opin Struct Biol* 21:62–67
- Vallee-Belisle A, Plaxco KW (2010) Structure-switching biosensors: inspired by Nature. *Curr Opin Struct Biol* 20:518–526
- Vogtherr M, Saxena K, Hoelder S, Grimme S, Betz M, Schieberr U, Pescatore B, Robin M, Delarbre L, Langer T et al (2006) NMR characterization of kinase p38 dynamics in free and ligand-bound forms. *Angew Chem Int Ed Engl* 45:993–997
- Wen SH, Hsiao CK (2007) A grid-search algorithm for optimal allocation of sample size in two-stage association studies. *J Hum Genet* 52:650–658
- Woessner DE (1961) Nuclear transfer effects in nuclear magnetic resonance pulse experiments. *J Chem Phys* 35:41–47
- Wright CM, Heins RA, Ostermeier M (2007) As easy as flipping a switch? *Curr Opin Chem Biol* 11:342–346
- Yohe ME, Rossman K, Sondek J (2008) Role of the C-terminal SH3 domain and N-terminal tyrosine phosphorylation in regulation of Tim and related Dbl-family proteins. *Biochemistry* 47:6827–6839
- Young MA, Gonfloni S, Superti-Furga G, Roux B, Kuriyan J (2001) Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* 105:115–126
- Yu B, Martins IR, Li P, Amarasinghe GK, Umetani J, Fernandez-Zapico ME, Billadeau DD, Machius M, Tomchick DR, Rosen MK (2010) Structural and energetic mechanisms of cooperative autoinhibition and activation of Vav1. *Cell* 140:246–256