Title
Modeling of Hidden Structures Using Sparse Chemical Shift Data from NMR Relaxation Dispersion

Permalink
https://escholarship.org/uc/item/1cs3j635

Journal
Biophysical Journal, 120(2)

ISSN
0006-3495

Authors

Fenwick, R Bryn
Oyen, David
van den Bedem, Henry
et al.

Publication Date
2021

DOI
10.1016/j.bpj.2020.11.2267

Peer reviewed

Modeling of Hidden Structures Using Sparse Chemical Shift Data from NMR Relaxation Dispersion

R. Bryn Fenwick,[1,*] David Oyen,[1] Henry van den Bedem,[2] H. Jane Dyson,[1] and Peter E. Wright[1,*]

[1]Department of Integrative Structural and Computational Biology and Skaggs Institute of Chemical Biology, The Scripps Research Institute, La Jolla, California and [2]SLAC National Accelerator Laboratory, Stanford University, Menlo Park, California, and Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California

ABSTRACT   NMR relaxation dispersion measurements report on conformational changes occurring on the μs-ms timescale. Chemical shift information derived from relaxation dispersion can be used to generate structural models of weakly populated alternative conformational states. Current methods to obtain such models rely on determining the signs of chemical shift changes between the conformational states, which are difficult to obtain in many situations. Here, we use a "sample and select" method to generate relevant structural models of alternative conformations of the C-terminal-associated region of *Escherichia coli* dihydrofolate reductase (DHFR), using only unsigned chemical shift changes for backbone amides and carbonyls (^{1}H, ^{15}N, and ^{13}C$'$). We find that CS-Rosetta sampling with unsigned chemical shift changes generates a diversity of structures that are sufficient to characterize a minor conformational state of the C-terminal region of DHFR. The excited state differs from the ground state by a change in secondary structure, consistent with previous predictions from chemical shift hypersurfaces and validated by the x-ray structure of a partially humanized mutant of *E. coli* DHFR (N23PP/G51PEKN). The results demonstrate that the combination of fragment modeling with sparse chemical shift data can determine the structure of an alternative conformation of DHFR sampled on the μs-ms timescale. Such methods will be useful for characterizing alternative states, which can potentially be used for in silico drug screening, as well as contributing to understanding the role of minor states in biology and molecular evolution.

SIGNIFICANCE   Molecular motions on the μs-ms timescale are frequently employed by enzymes to accelerate progress through their catalytic cycles. These motions can be detected by NMR relaxation dispersion experiments and analyzed to give information on the rates of exchange between the ground state and various excited-state conformations, as well as on the populations of these excited states and the chemical shift differences between ground and excited states. This chemical shift information can be used to give direct insights into the structure of the excited state. In this study, we use sparse chemical shift data obtained from relaxation dispersion experiments to model the structure of an excited state of the enzyme dihydrofolate reductase.

INTRODUCTION

NMR relaxation dispersion measurements are a powerful method for identifying the regions of biomolecules that undergo conformational exchange on the μs-ms timescale between the ground state and an alternative minor state (1,2).

These experiments yield insight into the rates of interconversion between the different states as well as the populations of the states involved. X-ray crystallography and NMR spectroscopy can determine the ground-state structure, but it is often difficult to obtain a structural model for the minor state because of its inherently low population (3). Methods to obtain information on minor states using chemical shift change information obtained from the analysis of relaxation dispersion data have been pioneered by the group of Lewis Kay (4–6). These methods, which incorporate the sign of the chemical shift change (+, downfield; −, upfield), often use molecular dynamics or Monte Carlo fragment replacement methods to sample conformational

space, followed by chemical shift selection to choose the most representative structural model, are often referred to as "sample and select" methods (7–9). The structures of only a few minor states have been characterized using chemical shift changes (5,6,10,11), and many of these have taken advantage of the CS-Rosetta platform (12,13).

Structure determination of excited states is vital for understanding protein function, but the characterization of structural heterogeneity may also find application in drug discovery and protein engineering (14–16). Indeed, the ability to characterize the alternative states of proteins may enable the rational design of inhibitors to interrupt the cycling of the target between ground and alternative conformational states (17). Moreover, the ability to model the alternative states of proteins that display allosteric mechanisms may enable the rational design of allosteric modulators that bind to cryptic sites revealed in the alternative state.

Here, we have determined the most probable structure of an alternative conformational state adopted by the C-terminal-associated region of *Escherichia coli* dihydrofolate reductase (DHFR). DHFR catalyzes the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate and is a model system for understanding the role of protein motions in biological function. Studies combining structural, dynamic, kinetics, and simulation data have already identified conformational changes in the active site loops and their relationship to function (18). Extensive relaxation dispersion studies of all complexes and mutant proteins of DHFR studied to date reveal the presence of a μs-ms timescale process, unrelated to the catalytic mechanism of the enzyme, in a loop that packs against the C-terminal region (18–21). In this work, we use a well-behaved mutant, N23PP/S148A, originally designed to mimic the active site sequence of human DHFR (21), to probe the structural changes associated with fluctuations in the C-terminal-associated loop. The region showing dispersion includes the loop residues 128–134 as well as the C-terminal region (residues 156–158).

Previous structure determinations of low-population excited states have relied on determination of the signs of the chemical shift changes between the ground and excited states, which can be challenging because of the need to quantify extremely small changes in cross-peak positions (22,23). The experiments required to determine the signs fall into three categories: HSQC, HSQC methods at different fields; HSQC, HMQC methods at the same field, and R1ρ measurements (22–24). To determine the chemical shift signs of backbone chemical shifts, even in model systems, multiple methods have been used to cross-validate the signs obtained. However, determination of the signs of chemical shift changes in larger proteins is difficult because the above methods rely on well-resolved peaks and excellent signal to noise, which are challenges for larger systems. Here, we avoid the problem of determining the signs of the chemical shift differences by using the absolute chemical shift changes (of ambiguous sign) to generate structures in

CS-Rosetta. To assess the utility of unsigned chemical shift differences for structure calculation we have used the Rosetta protocol developed by the Kay laboratory for a mutant of T4 lysozyme (5), except that in our case no information on the sign of the chemical shift changes was used during the Rosetta calculations. The resulting ensemble contains a set of conformations that are in excellent agreement with the experimentally determined chemical shifts.

MATERIALS AND METHODS

Sample and data collection

Uniformly ^{15}N- and ^{13}C-labeled and randomly deuterated ($\sim$68%) samples of the N23PP/S148A mutant of *E. coli* DHFR were expressed and purified as previously described, and the complex with NADP$^+$ and folate (E:folate:NADP$^+$) was prepared for NMR (20,25). Briefly, a His$_6$-SUMO-DHFR fusion protein was expressed in M9 media and nickel purified before cleavage of the affinity tag with recombinant His$_6$-ULP1. Size exclusion chromatography was used to further purify the protein. Spectra were recorded for samples in NMR buffer (50 mM phosphate (pH 6.8), containing 100 mM KCL, 1 mM EDTA, 1 mM DTT, and 0.02% NaN$_3$). NMR samples were $\sim$1.5-mM protein concentration and contained 20 mM NADP$^+$, 12 mM folic acid, and 10% D$_2$O. Acquisition of relaxation dispersion data for the backbone amides and tryptophan side-chain indoles has been reported elsewhere (20). Experiments were recorded at a temperature of 301 K at ^{1}H spectrometer frequencies of 500 and 800 MHz using Poisson gap nonuniform sampling. Carbonyl relaxation dispersion was recorded on the same sample, using previously described pulse sequences (26,27). The CPMG relaxation dispersion profiles for the C-terminal-associated region were fitted with the program GLOVE using a single global population (p_B) and rate (k_{ex}) (28).

Structure calculations and analysis

The chemical shift differences $\delta\varpi N$, $\delta\varpi H$, and $\delta\varpi C'(i-1)$ (where ϖ indicates that the value is in parts per million (ppm)) between ground and excited states derived for the residues that undergo relaxation dispersion were included in CS-Rosetta calculations using the default method available in Rosetta (5,29). Structure calculations removed and remodeled residues 126–135 and 155–159 from the structure PDB: 4KJL (30). CS-Rosetta was used in the standard configuration, except for the parameters used for fragment selection based on the agreement with experimental chemical shifts, which were modified to mirror the parameters shown previously to give improved results (31). For all calculations, DHFR and homologous structures were excluded from the fragment selection process. CS-Rosetta results were analyzed using the Rosetta energy to select 4000 structures from the total 20,000 structures generated. The structures were then clustered using dihedral angle principal component analysis (PCA) within the program GROMACS to determine the projections of the structures on the two first principal components (32,33). For the cluster of structures in the excited state, a set of 174 high-energy structures that had poor packing of the Y128 and W133 side chains, where the rings were exposed to the solvent and eliminated from further analysis based on energy were used (see Results). Chemical shifts for the residues in the C-terminal region were calculated for each structure in the cluster using SPARTA+ (34); the chemical shift differences ($\Delta\varpi$) were calculated from the structures as follows:

$$\Delta\varpi = \delta_{excited} - \delta_{ground}$$

where δ_{excited} is the chemical shift of the excited state calculated from a Rosetta-derived protein structure using SPARTA+ and δ_{ground} is the

experimentally determined chemical shift of the ground state. The comparisons in this work are to the absolute values, $|\Delta\varpi|$ and $|\delta\varpi|$, which do not have signs and have units of ppm.

Chemical shift difference weighting

To be concise when discussing the fit of chemical shift differences from three different nuclei (^{1}H, ^{13}C, and ^{15}N), we combined the three chemical shift differences into a single value (normalized root mean-square errors (nRMSE), see below). To accomplish this, scale factors (w_x) were used to ensure that nitrogen shifts do not bias the perceived fit because of their larger magnitudes. We used the change in chemical shift dispersion between folded and unfolded states in a set of well-characterized proteins to determine the scale factors. The chemical shift differences were normalized by the relative change in chemical shift dispersion observed in representative protein structures upon unfolding (35). The chemical shift dispersions (ppm) for folded proteins are as follows: HN, 4.13; N, 30.9; and C', 9.37 ppm; and for unfolded proteins HN, 1.04; N, 20.1; and C', 5.12 ppm. The chemical shift normalization weights were determined as follows:

$$w_X = \frac{s_N^{folded} - s_N^{unfolded}}{s_X^{folded} - s_X^{unfolded}}$$

where s_X^{folded} and $s_X^{unfolded}$ are the chemical shift dispersions for the nucleus X of interest (e.g., N, C, H), and s_N^{folded} and $s_N^{unfolded}$ are the chemical shift dispersions of amide nitrogen. The final weights were 1.0 for amide nitrogen, 3.5 for amide proton, and 2.5 for carbonyl carbon chemical shifts. We use two nRMSEs here. The first is nRMSE$_{excited}$, which is the RMS of $|\delta\varpi|$ − $|\Delta\varpi|$, the second is nRMSE$_{ground}$, which is the RMS of $|\Delta\varpi|$. The RMSE values were only calculated using residues and nuclei that displayed chemical shift changes.

Selection of ensembles with csFit

Ensembles of structures were selected to improve agreement with the chemical shifts. The optimization was set up with a threshold constraint, which resulted in a mixed-integer quadratic program (MIQP) (36). The approach followed that of similar work to select a parsimonious, population-weighted set of structural representations that are consistent with experimental data using a quadratic program (37,38),

$$\text{minimize } ||Ew - u||_2^2$$

$$\text{subject to } ti - w \le 0, \ w - i \le 0, \ \sum w = 1, \ \sum i \le n$$

where E is the normalized chemical shift differences ($\Delta\varpi$) calculated from each structure in the ensemble, w is the weight of each structure in the population of states, and u are the normalized chemical shift differences derived from the relaxation dispersion experiments ($\delta\varpi$). Subject to the constraints of a minimal population weight t and a maximal number of population members n (cardinality constraint), the indicator array i facilitates the restraints. The cardinality constraint comes into play only when the population threshold $t < 1/n$. For reasons of computational efficiency, the set of 402 conformations (excluding packing outliers) from the excited-state cluster was selected for optimization with the signs of 23 ambiguous chemical shift differences. In our hands, the full enumeration of 2^{23} sign combinations took ~3 days. Calculations were performed using the CVXPY modeling language combined with the community edition of the CPLEX optimization library version 12.10 (39).

RESULTS

To determine the structure of the hidden excited state of the C-terminal region of DHFR observed by NMR relaxation dispersion, we used CS-Rosetta (12,13). For fragment selection before structure calculation, we used the chemical shift differences ($\delta\varpi$N, $\delta\varpi$H, $\delta\varpi$C') obtained from CPMG experiments on the E:folate:NADP$^+$ complex of the N23PP/S148A mutant of *E. coli* DHFR to characterize the structure of the excited state sampled by the C-terminal region (Table 1). The alternative C-terminal conformational state is present at a population of 3.3% and exchanges with the ground state at a rate (k_{ex}) of 630 s^{-1} (20). The N23PP mutation introduces a pair of prolines that change the residue numbering for this mutant compared with the wild-type sequence; to aid in comparison with previous work on DHFR, we have used the wild-type residue numbering throughout this work.

CS-Rosetta structure calculations

The potential difficulty of structure calculations using sparse and ambiguous chemical shift data is that it can be difficult to obtain correct structures. However, we show here that, in cases in which the difference in conformation between the ground and excited states is relatively small, it may be possible to determine the excited-state structure using ambiguous chemical shift restraints with the program Rosetta. The calculations were performed following the protocol of the Kay group for the mutant of T4 lysozyme, except that ambiguous (unsigned) chemical shifts were used to make the initial fragment selections (5). For the structure calculations performed here, all DHFR structures were removed from the fragment sets to avoid bias. To test the method, three synthetic control calculations were performed to generate the ground-state structure of the C-terminal-associated region of *E. coli* DHFR. In the first control, the structure was modeled with fragments selected purely based on the amino acid sequence, with no guidance from chemical shift information. The second control used the ground-state (native) chemical shifts of the N23PP/

TABLE 1 Chemical Shift Differences (ppm) for the C-Terminal Region of N23PP/S148A DHFR

Residue	$\delta\varpi$N*	$\delta\varpi$H*	$\delta\varpi$C' (i−1)[a]
Gln 108	0.3	0.1	0.1
Tyr 128	4.7	0.5	0.4
Glu 129	2.4	0.7	2.1
Asp 131	2.4	1.4	1.3
Trp 133	1.9	0.3	0.4
Glu 134	1.3	0.0	2.1
Leu 156	0.4	0.1	0.3
Glu 157	0.9	0.0	0.3
Arg 158	2.2	0.1	0.2

[a]Absolute values of the chemical shift differences ($|\delta\varpi|$) derived from fitting the relaxation dispersion profiles (20). The signs are unknown.

S148A mutant DHFR in the E:folate:NADP$^+$ complex. The third control targeted the ground-state structure by using the unsigned excited-state chemical shift differences from Table 1 randomly assigned above and below the ground-state chemical shift. For example, if the ground-state proton chemical shift is 8.3 ppm and the unsigned excited-state chemical shift difference is 0.4 ppm, then the native-like chemical shift was randomly assigned to either 7.9 or 8.7 ppm with a Δ of 0.4 ppm. Finally, the fourth structure generation used the ground-state chemical shifts of the E:folate:NADP$^+$ complex of N23PP/S148A DHFR together with experimentally determined changes in the chemical shift derived from the relaxation dispersion experiments (Table 1) to calculate the structure of the alternative state sampled by the C-terminal-associated region. For each of the control simulations and structure generation of the excited state from experimental data, 20,000 structures were calculated, analyzed, and compared with the known structures of *E. coli* DHFR.

Characterization of the Rosetta simulations

The results of the structure calculations and controls are shown in Fig. 1. Fig. 1 *A* shows the projections of 108 published x-ray structures of DHFR on the primary dihedral angle PCA components of the residues of the C-terminal region. The majority (106 of 107) of the ground-state structures are in the same position, at −1.5 dPC 1 and −0.4 dPC 2, on the conformational coordinates. The other structure belongs to chain B in the x-ray structure of the N23PP/G51PEKN mutant of *E. coli* DHFR (4GH8) (40), at 2.0 dPC 1 and −1.2 dPC 2 (Fig. 1 *A*). The results of the three control simulations are shown in Fig. 1, *B–D*. For structures generated without chemical shifts (Fig. 1 *B*), where fragment selection was purely sequence based, we observe that there is a single major conformational state consistent with the ground state and a small, ∼1% population in the region of

2.0 and −1.2. When the ground-state chemical shifts of *E. coli* DHFR are used in the Rosetta calculation to select fragments, we observe that the conformational landscape changes (Fig. 1 *C*) and that the second state becomes increasingly populated in the region of 2.0 and −1.2; this region is very similar to the region occupied by 4GH8:B (see Fig. 1 *A*). It is important to reiterate that DHFR homologs were not present in the fragment libraries used. In the third control calculation, in which the ground-state chemical shifts are ambiguous, we again observe that the ground state is the major populated state, whereas the minor state is also populated (Fig. 1 *D*). Finally, when the structures are generated with fragments selected using ambiguous signs of the minor state chemical shifts derived from the CPMG relaxation dispersion measurements given in Table 1, we see that once again, both a ground state and the minor alternative state are observed. For the excited state cluster, 31 structures were higher in energy and had poor packing of the Y128 side chain and 143 structures with poor packing of W133 (see Fig. S1). These high-energy structures were eliminated from further analysis.

Excited-state conformational clusters

Details of the two most highly populated regions in the dihedral conformational space for the C-terminal-associated region are shown in Table 2. In all of the structure calculations, the position of the ground state is conserved at −1.36 dPC 1 and −0.35 dPC 2. The inclusion of the chemical shifts in the Rosetta calculations slightly changes the conformation of the alternative state, as indicated by a slight change in the coordinates of the center of the alternative conformational cluster (Fig. 1; Table 2). Also, we observe that the number of structures associated with the ground state and alternative states changes as the chemical shift information is varied between the different synthetic data sets. For example, when chemical shifts are not
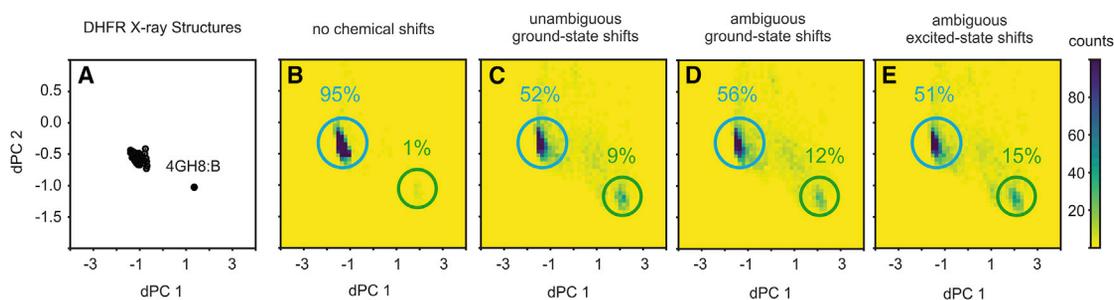
FIGURE 1 Conformational space of Rosetta structures and x-ray structures for the residues of the C-terminal region. (*A*) Projection of published x-ray structures of *E. coli* DHFR is shown. (*B*) Rosetta structures generated without chemical shifts are shown. (*C*) Rosetta structures generated using chemical shifts for the ground state are shown. (*D*) Rosetta structures generated using the ground-state chemical shift differences made ambiguous, as detailed in the Materials and Methods and Results, are shown. (*E*) Rosetta structures generated using the unsigned (ambiguous) excited-state chemical shift differences from the relaxation dispersion experiments are shown. The position of the ground state on the dihedral PCA coordinates is at −1.4 (dPC 1) and −0.4 (dPC 2); the alternative minor state is located at 2.0 (dPC 1) and −1.2 (dPC 2). The axes represent the two primary dihedral PCA components of the four simulations combined. The colors represent the binned counts for the dihedral space, with dark blue indicating >100 structures and yellow indicating the low populated regions of conformational space. Circles indicate the positions of the major (*blue*) and minor (*green*) clusters. To see this figure in color, go online.

TABLE 2 Details of the Two Largest Clusters from Each of the Rosetta Calculations

	Native Position (dPC1/dPC2)	Alternative Position (dPC1/dPC2)	Number of Native Conformations[a]	Number of Alternative Conformations[a]
No shift	−1.36/−0.35	1.84/−1.13	3820	44
Ground-state chemical shifts	−1.36/−0.35	2.00/−1.19	2080	365
Unsigned ground-state chemical shift differences	−1.36/−0.35	2.00/−1.19	2238	482
Unsigned excited state chemical shift differences	−1.36/−0.35	2.00/−1.19	2059	586

[a]The total number of structures used for clustering was 4000, representing the lowest-energy 20% of the 20 thousand structures generated.

included in the Rosetta calculations, the majority of configurations (96%) are in the ground-state cluster, whereas only 1% are in the alternative conformation. When the ground-state chemical shifts are used, we observe a substantial increase in the occupancy of the alternative state, which increases to 9% of the total configurations, whereas the ground-state population decreases to 52%. In the calculations for which unsigned chemical shift differences were used, we still observe a large proportion of ground-state structures (56%) as well as 12% of the alternative conformational state. Finally, when structures are generated using the ambiguous $\delta\varpi$-values from the CPMG relaxation dispersion measurements (corresponding to the excited state), we observe the ground state at 51% population and the alternative state at 15%.

Agreement with chemical shifts

To validate the structure of the alternative state, predicted chemical shifts for the Rosetta structures were calculated using SPARTA+ (34) and compared with the experimental chemical shift differences derived from the relaxation dispersion data. In Table 3, the RMSE$_{excited}$ between experimental and calculated chemical shift differences are given for the top five structures of each cluster.

We have also calculated an nRMSE that weights the different nuclei by their chemical shift dispersion change observed between folded and unfolded protein structures (35). This enables the chemical shift agreement to be combined for multiple nuclei and summarized as a single value for a given structure. A comparison of the

normalized chemical shift RMSE$_{ground}$-values for the two clusters (Fig. 2) shows that the ground-state structure has a significantly worse agreement with the experimental $|\delta\varpi|$-values and that the alternative-state structures are in better agreement with the experimental data for the excited state.

The results in Fig. 2, A and B also illustrate the better agreement of the alternative state with the experimental chemical shifts compared with the native state. At the same time, Fig. 2 C shows agreement between the x-ray structure 4GH8 chain B and the experimental chemical shift differences. The major chemical shift difference outliers are the nitrogen chemical shifts for residues 131 and 128. For the amide of Y128, a hydrogen bond to water is observed in the 4GH8:B structure, whereas the amide of 131 is hydrogen-bonded to the carbonyl of Y128. The data in Fig. 2 validate the minor cluster as the correct backbone configuration for the alternative conformational state of DHFR. RMSE$_{ground}$ is the RMSE to the ground-state chemical shifts for each of the structures (Table 3). It is important to note that in Table 3, we have selected the structures that have the lowest nRMSE$_{excited}$ for both the ground- and excited-state clusters. Therefore, it is not surprising that the nRMSE$_{ground}$ is larger than the nRMSE$_{excited}$ for all the structures in this table. We observe that the nRMSE$_{ground}$ for the excited-state cluster is larger than for the structures in the native-state cluster. That we observe both the lower nRMSE$_{excited}$ for the excited-state compared with the ground-state cluster and a lower nRMSE$_{ground}$ for the ground-state cluster compared with the excited-state cluster indicates that this excited-state cluster is the excited state.

TABLE 3 Details of the Five Best Ranked Conformations by the RMSE of Normalized Chemical Shifts for the Native and Alternative Clusters

Cluster	Rank	Structure	Rosetta Energy	^{1}H RMSE (ppm)	^{15}N RMSE (ppm)	^{13}C RMSE (ppm)	nRMSE$_{excited}$[a]	nRMSE$_{ground}$[a]
Native	1	10651	−241	0.29	1.20	0.30	1.00	2.22
Native	2	14425	−243	0.30	0.86	0.44	1.01	2.05
Native	3	08919	−228	0.28	1.00	0.42	1.02	2.04
Native	4	10282	−242	0.31	0.94	0.40	1.02	2.05
Native	5	00617	−239	0.32	1.32	0.19	1.03	2.18
Alternative	1	03164	−242	0.27	0.84	0.37	0.91	2.35
Alternative	2	11874	−238	0.20	1.26	0.30	0.93	2.62
Alternative	3	10756	−242	0.21	1.13	0.37	0.94	2.43
Alternative	4	05293	−223	0.30	1.17	0.19	0.96	2.40
Alternative	5	03718	−239	0.24	0.94	0.43	0.96	2.34

[a]nRMSE is the normalized RMSE that combines multiple nuclei (see Materials and Methods).
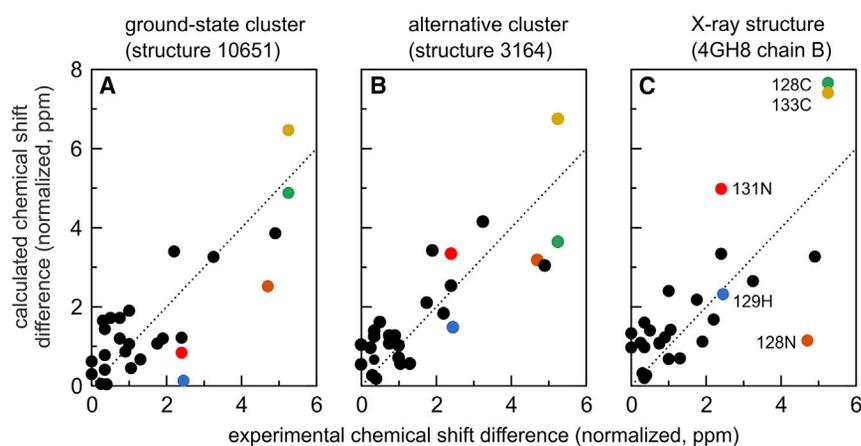
FIGURE 2　Correlation of the normalized experimental $|\delta\varpi|$ from the CPMG relaxation dispersion experiments and the normalized $|\varDelta\varpi|$ calculated from the structures using SPARTA+. The data are shown for (*A*) the lowest nRMSE structure from the ground-state cluster, (*B*) the excited-state cluster, and (*C*) for chain B of the x-ray crystal structure 4GH8 that is closest to the alternative state. Representative points are labeled with colors. To see this figure in color, go online.

This observation is also consistent with visual comparison of the protein structures in Fig. 3.

Fig. 3 *A* shows a comparison of the backbone conformation of the ground state in several structures from the Rosetta ensemble of Fig. 1 *E* with a large number of DHFR x-ray structures. The comparison shows that the ground state (-1.5 dPC 1 and -0.4 dPC 2 in Fig. 1 *E*) is consistent with the structure of the C-terminal-associated region observed in 106 of the 107 x-ray structures. Fig. 3 *B* shows the top five structures of the excited-state cluster, and Fig. 3 *C* shows a comparison of the best individual structures from the ground- and excited-state clusters.

Structural ensembles of the excited state

"Sample and select" methods have been used to select representative structural ensembles of proteins from structure-generation protocols (29,41). The "sample and select" methods were further developed to select ensembles of structures from MD simulations that agreed with NMR order parameters (7), and have subsequently been used with MD simulations and RDCs (42). In our case, we have used the Rosetta excited-state cluster of structures from which we extract a small ensemble of the excited state. To select an ensemble of structures that further improve the agreement with the excited state experimental chemical shift differences, we designed the csFit algorithm to iteratively explore the signs of the $\delta\varpi$-values to determine a

parsimonious, population-weighted ensemble of 10 or fewer weighted members that have weights $\geq 10\%$ that optimally and collectively fit the chemical shift data (14,37). The details of the approach are given in the Materials and Methods. In Table 4, we provide the details of the ensemble that best agreed with the chemical shift differences.

In Table 5, we compare the best ensemble selected from the full 402 structures in the excited-state cluster with the ensembles of equally weighted structures generated from the top five native cluster structures, the top five excited-state structures, and an optimized ensemble of the top five excited-state structures. We observed better agreement with the chemical shifts for the csFit ensembles generated from the top five excited-state cluster structures and all 402 structures in the excited-state cluster compared with the equally weighted ensembles generated from the excited-state and native-state clusters.

The individual structures in Fig. 3 *B* are akin to the average structures from NMR or x-ray crystallography. In contrast, a csFit ensemble captures a group of representative states similar to what is obtained from molecular dynamics or multiconformer modeling of x-ray crystallographic data. The csFit approach selects a group of structures that together fit the experimental $|\delta\varpi|$-values better than does any individual structure. The csFit structural ensemble of five structures is shown in Fig. 4 *A*, and the correlation of the ensemble-averaged chemical shift differences $|\varDelta\varpi|$- and experimental $|\delta\varpi|$-values are shown in Fig. 4 *B*. The
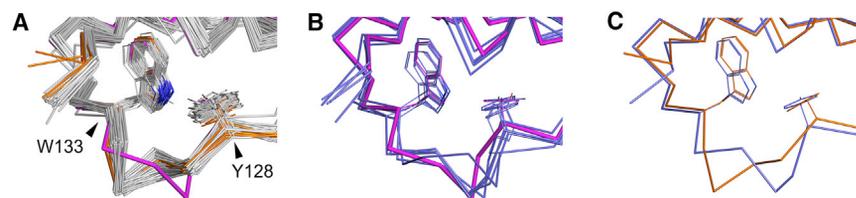
FIGURE 3　(*A*) Ground-state structures of the C-terminal region of DHFR are pictured, showing a superposition of 106 *E. coli* DHFR x-ray structures (*gray*), together with the 4GH8:B structure (*magenta*) and the five lowest-energy Rosetta structures obtained for the ground state (*orange*). Ensemble models that describe structure and dynamics have been removed (4PTJ and 4PTH). (*B*) The best five individual structures (by nRMSE) from the excited-state cluster (*blue*), together with the 4GH8:B structure (*magenta*) are shown. To see this figure in color, go online.

TABLE 4 Summary of the csFit Ensemble Fitted to the Experimental Data

Structure	Percentage	Rosetta Energy	¹H RMSE (ppm)	¹⁵N RMSE (ppm)	¹³C RMSE (ppm)	nRMSE$_{excited}$[a]	nRMSE$_{ground}$[a]
04084	10.0	−240	0.31	0.93	0.37	0.97	2.37
02434	10.0	−239	0.25	1.98	0.26	1.31	2.66
03718	36.2	−239	0.24	0.94	0.43	0.96	2.34
08606	22.1	−237	0.30	1.29	0.32	1.07	2.65
00456	21.6	−235	0.19	1.63	0.46	1.21	2.76

[a]nRMSE is the normalized RMSE that combines multiple nuclei (see Materials and Methods).

backbone conformations of the individual members of the csFit ensemble are very similar to the x-ray structure 4GH8 chain B (see Fig. 4 A) and the best individual structures in Fig. 3 B. The structure 03718 (labeled in Fig. 4 A) is the only structure that is in both the csFit ensemble and in the top five individual structures. It is surprising that the same conformation is present in both ensembles; this indicates that structural features of this conformation are important for agreement with the average experimental |δϖ|-values. Chemical shifts are very sensitive to the local structural environment, and small differences in the structures selected by csFit lead to the improvement in agreement with the experimental |Δϖ| when chemical shifts are linearly averaged over the ensemble. Fig. 4 B shows the improved agreement for the conformational ensemble, which has an nRMSE$_{excited}$ of 0.63, an improvement of almost 30% over nRMSE$_{excited}$ of 0.91 for the best single structure (03164). The structural heterogeneity in the ensemble shown in Fig. 4 A may indicate that the excited state is relatively flexible, consistent with RMSDs of 1.7 and 2.3 Å, respectively, for the backbone and heavy atoms of the remodeled residues (126–135). The structure 03718, which accounts for 36.3% of the csFit ensemble, indicates that, in the excited state, the structure of the C-terminal-associated region of *E. coli* DHFR samples configurations are similar to the state found in structure 03718 and the x-ray crystal structure 4GH8 chain B.

Given that the crystal structure 4GH8 contains both the ground-state (chain A) and excited-state (chain B) structures in the asymmetric unit, we looked for interactions that might stabilize the excited state in the x-ray model. Residues in the 127–134 loops of chains A and B undergo different crystal lattice contacts. In chain A, in which the loop adopts the usual ground-state conformation, a calcium ion is bound to the loop. The calcium ion does not appear to influence the loop conformation because x-ray structures that lack

ions bound in this region (e.g., 4NX6) have essentially identical backbone structures. In contrast, the side chain of E129 and backbone carbonyl of P130 in chain B form intermolecular hydrogen bonds to side chains in a neighboring molecule that could potentially stabilize the excited-state conformation in the crystal lattice.

DISCUSSION

The Rosetta calculations indicate that the C-terminal-associated region of DHFR populates two possible conformational states and that these two states are well defined by the combination of the ground-state chemical shifts and the chemical shift differences from CPMG relaxation dispersion experiments. Our calculations, based on the relaxation dispersion data, suggest that this region is rearranged in the excited-state structure into a different conformation that resembles that seen in one particular x-ray structure. Rosetta is a structure prediction program that searches structure space by replacing structural fragments in the current model with new fragments in a predefined fragment library derived from known structures. The structural propensities of the peptide fragments may already include information on the alternative state, and unsigned chemical shift differences may be sufficient to extract the correct alternative-state structure using a "sample and select" approach.

Computational screening for molecular inhibitors could improve if convenient methods to determine alternative configurations of biomolecules become available. Here, we have described the study of a "sample and select" method to yield structural ensembles of alternative states enriched using information from relaxation dispersion experiments. The position of the alternative conformational state is conserved in all simulations, except those performed without chemical shift restraints. Comparing the top five structures with the lowest nRMSE for the excited state,

TABLE 5 Summary of the Various Ensembles Fitted to the Experimental Data

Ensemble	¹H RMSE (ppm)	¹⁵N RMSE (ppm)	¹³C RMSE (ppm)	nRMSE$_{excited}$[a]	nRMSE$_{ground}$[a]
Top five from native cluster[b]	0.31	1.49	0.28	1.13	1.92
Top five from excited cluster[b]	0.22	1.47	0.26	1.03	2.27
csFit 2 structures (top five excited cluster structures)[c]	0.20	0.96	0.26	0.79	2.43
csFit 5 structures (top 402)	0.22	0.50	0.24	0.63	2.32

[a]nRMSE is the normalized RMSE that combines multiple nuclei (see Materials and Methods).
[b]Ensemble generated using 20% weight for each member.
[c]Ensemble fitted with csFit (65% 11874, 35% 03718).
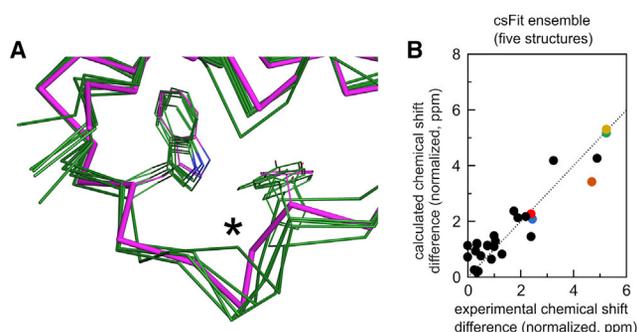
FIGURE 4 The ensemble of the excited state. (*A*) The csFit ensemble of structures selected from the top 402 structures in the excited-state cluster (*green*), with the structure 03718 indicated with a star, together with the 4GH8:B structure (*magenta*) are shown for comparison to Fig. 3. (*B*) The agreement with the ensemble generated using csFit for is shown comparison with Fig. 2 (see Materials and Methods). The ensemble captures the chemical shifts of the excited state as compared with the individual structures (Fig. 2 *B*). Colors represent the same representative points as in Fig. 2. To see this figure in color, go online.

we observe that the structures show remarkable agreement with the structure of 4HG8 model B, even though we have only included ambiguous backbone chemical shift information. We note that the side chains of the 4GH8 model B structure also have similar packing to our selected models, further suggesting that this conformation is representative of the excited state. The backbone structure of the alternative conformational-state ensemble from csFit determined from the relaxation dispersion restraints (Fig. 4 *A*, *green*) closely resembles that of chain B in the x-ray structure 4GH8 (Fig. 4 *A*, *magenta*). Residues 130–132 in the majority of x-ray structures, exemplified by 1RX2 (43) form a 3_{10}-helix, as identified by the secondary structure assignment program STRIDE (44) and 3QL3 (21), where the region is helical between residues 129 and 131. There was, however, poorer agreement between the SPARTA+-calculated chemical shifts for the structure of 4GH8:B and the experimental shift differences derived from the CPMG experiments (Fig. 2 *C*). Both of these interactions could be perturbed by experimental conditions, such as cryo-cooling of the crystals and variations in buffer conditions, or the differences could be due to the difficulty in accurately predicting nitrogen chemical shifts from a single conformation (34,45).

NMR chemical shifts are sensitive to small structural changes, and the careful selection of a set of native-state structures can enhance the agreement of calculated excited-state structures with the experimental data (46–51). For this reason, we selected an ensemble of structures that maximize the agreement with the chemical shift differences derived from the relaxation dispersion measurements. In Table 4, we present the parameters for the ensemble that best fit the chemical shift differences; it can be seen that the combined ensemble gives superior agreement with the experimental chemical shift differences than do the individual structures

from the cluster. Given that chemical shifts calculated from molecular dynamics trajectories are highly variable and averaging can cause dramatic improvements in agreement with experimental data (52), it is maybe not surprising that the csFit ensemble shows better agreement. The cluster of structures that have lost the helical character of residues 129–132 is consistent with the excited state. Indeed a high degree of structural heterogeneity was observed in the multiple conformer ensemble generated using crystallographic data collected at room temperature (53,54).

Conformational fluctuations in the C-terminal-associated region have been identified in all *E. coli* DHFR complexes studied to date and appear to be linked to the intrinsic dynamics of the enzyme (55). These studies show that the 128–134 loop samples two distinct backbone conformations in solution and exchanges between them on a ms timescale. Although the biological function of these conformational fluctuations remains unknown, this region is connected allosterically to the active site (56), and interactions with other proteins could potentially modulate DHFR activity.

Code availability

The code for the algorithm csFit has been uploaded to the website of the Wright lab and to the GitHub server (https://www.scripps.edu/wright and https://github.com/ExcitedStates/csFit).

CONCLUSION

Here, we have demonstrated the use of a simplified "sample and select" method to generate configurations of the excited state of the C-terminal-associated region of *E. coli* DHFR that is populated through conformational fluctuations on the μs-ms timescale. Using unsigned $\delta\varpi$-values from relaxation dispersion measurements with ground-state chemical shifts and a structural model, it is possible to model the local hidden structures and heterogeneity of the excited state. This work offers a streamlined approach to the use of dispersion experiments, even in cases where signs of $\delta\varpi$ cannot be determined experimentally, to generate alternative conformational states of macromolecules to provide new insights into molecular mechanisms or for in silico drug discovery.

SUPPORTING MATERIAL

AUTHOR CONTRIBUTIONS

R.B.F. and P.E.W. designed the research. R.B.F. and D.O. acquired and analyzed NMR data. All authors discussed the results and wrote the manuscript.

ACKNOWLEDGMENTS

REFERENCES

1. Akke, M., and A. G. Palmer. 1996. Monitoring macromolecular motions on microsecond to millisecond timescales by $R_{1\rho}$-R_1 constant relaxation time NMR spectroscopy. *J. Am. Chem. Soc.* 118:911–912.

2. Palmer, A. G., III, C. D. Kroenke, and J. P. Loria. 2001. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* 339:204–238.

3. van den Bedem, H., and J. S. Fraser. 2015. Integrative, dynamic structural biology at atomic resolution–it's about time. *Nat. Methods.* 12:307–318.

4. Barette, J., A. Velyvis, …, L. E. Kay. 2012. Cross-validation of the structure of a transiently formed and low populated FF domain folding intermediate determined by relaxation dispersion NMR and CS-Rosetta. *J. Phys. Chem. B.* 116:6637–6644.

5. Bouvignies, G., P. Vallurupalli, …, L. E. Kay. 2011. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature.* 477:111–114.

6. Vallurupalli, P., D. F. Hansen, and L. E. Kay. 2008. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci. USA.* 105:11766–11771.

7. Chen, Y., S. L. Campbell, and N. V. Dokholyan. 2007. Deciphering protein dynamics from NMR data using explicit structure sampling and selection. *Biophys. J.* 93:2300–2306.

8. Frank, A. T., A. C. Stelzer, …, I. Andricioaei. 2009. Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition. *Nucleic Acids Res.* 37:3670–3679.

9. Fenwick, R. B., S. Esteban-Martín, and X. Salvatella. 2011. Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur. Biophys. J.* 40:1339–1355.

10. Korzhnev, D. M., T. L. Religa, …, L. E. Kay. 2010. A transient and low-populated protein-folding intermediate at atomic resolution. *Science.* 329:1312–1316.

11. Neudecker, P., P. Robustelli, …, L. E. Kay. 2012. Structure of an intermediate state in protein folding and aggregation. *Science.* 336:362–366.

12. Nerli, S., A. C. McShan, and N. G. Sgourakis. 2018. Chemical shift-based methods in NMR structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* 106–107:1–25.

13. Nerli, S., and N. G. Sgourakis. 2019. Cs-Rosetta. *Methods Enzymol.* 614:321–362.

14. van Zundert, G. C. P., B. M. Hudson, …, H. van den Bedem. 2018. qFit-ligand reveals widespread conformational heterogeneity of drug-like molecules in X-ray electron density maps. *J. Med. Chem.* 61:11183–11198.

15. Buonfiglio, R., M. Recanatini, and M. Masetti. 2015. Protein flexibility in drug discovery: from theory to computation. *ChemMedChem.* 10:1141–1148.

16. Antunes, D. A., D. Devaurs, and L. E. Kavraki. 2015. Understanding the challenges of protein flexibility in drug design. *Expert Opin. Drug Discov.* 10:1301–1313.

17. Lee, G. M., and C. S. Craik. 2009. Trapping moving targets with small molecules. *Science.* 324:213–215.

18. Boehr, D. D., D. McElheny, …, P. E. Wright. 2006. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science.* 313:1638–1642.

19. McElheny, D., J. R. Schnell, …, P. E. Wright. 2005. Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc. Natl. Acad. Sci. USA.* 102:5032–5037.

20. Fenwick, R. B., D. Oyen, and P. E. Wright. 2016. Multi-probe relaxation dispersion measurements increase sensitivity to protein dynamics. *Phys. Chem. Chem. Phys.* 18:5789–5798.

21. Bhabha, G., J. Lee, …, P. E. Wright. 2011. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science.* 332:234–238.

22. Skrynnikov, N. R., F. W. Dahlquist, and L. E. Kay. 2002. Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments. *J. Am. Chem. Soc.* 124:12352–12360.

23. Auer, R., D. F. Hansen, …, L. E. Kay. 2010. Measurement of signs of chemical shift differences between ground and excited protein states: a comparison between H(S/M)QC and R1ρ methods. *J. Biomol. NMR.* 46:205–216.

24. Bouvignies, G., D. M. Korzhnev, …, L. E. Kay. 2010. A simple method for measuring signs of (1)H (N) chemical shift differences between ground and excited protein states. *J. Biomol. NMR.* 47:135–141.

25. Bhabha, G., L. Tuttle, …, P. E. Wright. 2011. Identification of endogenous ligands bound to bacterially expressed human and *E. coli* dihydrofolate reductase by 2D NMR. *FEBS Lett.* 585:3528–3532.

26. Ishima, R., J. Baber, …, D. A. Torchia. 2004. Carbonyl carbon transverse relaxation dispersion measurements and ms-micros timescale motion in a protein hydrogen bond network. *J. Biomol. NMR.* 29:187–198.

27. Lundström, P., D. F. Hansen, and L. E. Kay. 2008. Measurement of carbonyl chemical shifts of excited protein states by relaxation dispersion NMR spectroscopy: comparison between uniformly and selectively (13)C labeled samples. *J. Biomol. NMR.* 42:35–47.

28. Sugase, K., T. Konuma, …, P. E. Wright. 2013. Fast and accurate fitting of relaxation dispersion data using the flexible software package GLOVE. *J. Biomol. NMR.* 56:275–283.

29. Shen, Y., O. Lange, …, A. Bax. 2008. Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. USA.* 105:4685–4690.

30. van den Bedem, H., G. Bhabha, …, J. S. Fraser. 2013. Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat. Methods.* 10:896–902.

31. Vernon, R., Y. Shen, …, O. F. Lange. 2013. Improved chemical shift based fragment selection for CS-Rosetta using Rosetta3 fragment picker. *J. Biomol. NMR.* 57:117–127.

32. Altis, A., P. H. Nguyen, …, G. Stock. 2007. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* 126:244111.

33. Abraham, M. J., T. Murtola, …, E. Lindahl. 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 1–2:19–25.

34. Shen, Y., and A. Bax. 2010. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR.* 48:13–22.

35. Yao, J., H. J. Dyson, and P. E. Wright. 1997. Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett.* 419:285–289.

36. Gärtner, B., and S. Schönherr. 2000. An efficient, exact, and generic quadratic programming solver for geometric optimization. In Sixteenth Annual Symposium on Computational Geometry - SCG '00. ACM Press, pp. 110–118.

37. van den Bedem, H., A. Dhanik, …, A. M. Deacon. 2009. Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. *Acta Crystallogr. D Biol. Crystallogr.* 65:1107–1117.

38. Fonseca, R., D. V. Pachov, …, H. van den Bedem. 2014. Characterizing RNA ensembles from NMR data with kinematic models. *Nucleic Acids Res.* 42:9562–9572.

39. Diamond, S., and S. Boyd. 2016. CVXPY: a Python-embedded modeling language for convex optimization. *J. Mach. Learn. Res.* 17:83.

40. Liu, C. T., P. Hanoian, …, S. J. Benkovic. 2013. Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. *Proc. Natl. Acad. Sci. USA.* 110:10159–10164.

41. Rule, G. S., and T. K. Hitchens. 2006. Fundamentals of Protein NMR Spectroscopy. Springer, Dordrecht, the Netherlands.

42. Stelzer, A. C., A. T. Frank, …, H. M. Al-Hashimi. 2009. Constructing atomic-resolution RNA structural ensembles using MD and motionally decoupled NMR RDCs. *Methods.* 49:167–173.

43. Sawaya, M. R., and J. Kraut. 1997. Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. *Biochemistry.* 36:586–603.

44. Heinig, M., and D. Frishman. 2004. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* 32:W500–W502.

45. Markwick, P. R. L., C. F. Cervantes, …, J. A. McCammon. 2010. Enhanced conformational space sampling improves the prediction of chemical shifts in proteins. *J. Am. Chem. Soc.* 132:1220–1221.

46. Torda, A. E., R. M. Scheek, and W. F. van Gunsteren. 1989. Time-dependent distance restraints in molecular dynamics simulations. *Chem. Phys. Lett.* 157:289–294.

47. Bonvin, A. M., R. Boelens, and R. Kaptein. 1994. Time- and ensemble-averaged direct NOE restraints. *J. Biomol. NMR.* 4:143–149.

48. Hess, B., and R. M. Scheek. 2003. Orientation restraints in molecular dynamics simulations using time and ensemble averaging. *J. Magn. Reson.* 164:19–27.

49. Best, R. B., and M. Vendruscolo. 2004. Determination of protein structures consistent with NMR order parameters. *J. Am. Chem. Soc.* 126:8090–8091.

50. Clore, G. M., and C. D. Schwieters. 2004. How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? *J. Am. Chem. Soc.* 126:2923–2938.

51. Clore, G. M., and C. D. Schwieters. 2004. Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. *Biochemistry.* 43:10678–10691.

52. Li, D.-W., and R. Brüschweiler. 2010. Certification of molecular dynamics trajectories with NMR chemical shifts. *J. Phys. Chem. Lett.* 1:246–248.

53. Keedy, D. A., H. van den Bedem, …, J. S. Fraser. 2014. Crystal cryo-cooling distorts conformational heterogeneity in a model Michaelis complex of DHFR. *Structure.* 22:899–910.

54. Fenwick, R. B., H. van den Bedem, …, P. E. Wright. 2014. Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. *Proc. Natl. Acad. Sci. USA.* 111:E445–E454.

55. Boehr, D. D., H. J. Dyson, and P. E. Wright. 2008. Conformational relaxation following hydride transfer plays a limiting role in dihydrofolate reductase catalysis. *Biochemistry.* 47:9227–9233.

56. Reynolds, K. A., R. N. McLaughlin, and R. Ranganathan. 2011. Hot spots for allosteric regulation on protein surfaces. *Cell.* 147:1564–1575.