Integrative, dynamic structural biology at atomic resolution—it's about time

Henry van den Bedem[1,2] & James S Fraser[3,4]

Biomolecules adopt a dynamic ensemble of conformations, each with the potential to interact with binding partners or perform the chemical reactions required for a multitude of cellular functions. Recent advances in X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and other techniques are helping us realize the dream of seeing—in atomic detail—how different parts of biomolecules shift between functional substates using concerted motions. Integrative structural biology has advanced our understanding of the formation of large macromolecular complexes and how their components interact in assemblies by leveraging data from many low-resolution methods. Here, we review the growing opportunities for integrative, dynamic structural biology at the atomic scale, contending there is increasing synergistic potential between X-ray crystallography, NMR and computer simulations to reveal a structural basis for protein conformational dynamics at high resolution.

Biochemical mechanisms often depend on macromolecules accessing transient, 'hidden' excited states[1,2]. By their very nature, these dynamic processes are difficult to characterize structurally. However, a structural understanding of conformational dynamics can shed light on fundamental and unanswered questions in structural biology: What is the role of dynamics in catalysis? What is the role of conformational entropy in allosteric and binding events? Are long-range, structural interactions in proteins facilitated by pre-existing pathways or through conformational rearrangements? To what extent does ligand-receptor binding involve induced fit or conformational selection?

In practical terms, the rewards for understanding how substates and different parts of biomolecules are coupled are large; such understanding will increase our ability to manipulate or completely redirect protein or nucleic acids function by selectively stabilizing particular conformations. Accurate, atomic-scale representations of collective motions will aid in determining the effects of mutations that are distant from functional sites and developing allosteric small-molecule modulators of protein function. Furthermore, therapeutics based on protein design and engineering are rapidly developing[3,4]. As protein function is governed by a delicate balance of structure and motion, ensuring adequate sampling of favorable interactions is an important design specification[5]. For example, nonspecific encounter complexes, which can populate up to 30% of an ensemble[6], can hierarchically facilitate formation of a productive complex by probing the binding partner conformations before establishing specific intermolecular interactions[7]. Especially dynamic proteins that only briefly adopt the precise active site conformations needed for catalysis may contribute to low activity of designed enzymes[8], which generally require many rounds of further experimental optimization by directed evolution.

Convergent developments in NMR spectroscopy and X-ray crystallography open up the possibility of bringing together structure and dynamics for atomic-resolution integrative studies of biomolecules (**Fig. 1**). Integrative structural methods that combine sparse or low-resolution data have helped to advance our understanding of many large macromolecular assemblies that cannot be characterized by any single technique alone[9]. A challenge common to both high- and low-resolution integrative structural biology is representing motionally averaged, sparse or ambiguous data with an ensemble of states. Parsimoniously representing key features of the data demands sophisticated computational procedures, often depending on techniques traditionally associated with artificial intelligence and robotics. Evidence is emerging from

[1]Joint Center for Structural Genomics, Stanford Synchrotron Radiation Lightsource, Stanford University, Menlo Park, California, USA. [2]Division of Biosciences, SLAC National Accelerator Laboratory, Stanford University, Menlo Park, California, USA. [3]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California, USA. [4]California Institute for Quantitative Biology, University of California, San Francisco, San Francisco, California, USA. Correspondence should be addressed to H.v.d.B. (vdbedem@stanford.edu) or J.S.F. (james.fraser@ucsf.edu).

RECEIVED 29 AUGUST 2014; ACCEPTED 21 JANUARY 2015; PUBLISHED ONLINE 31 MARCH 2015; DOI:10.1038/NMETH.3324

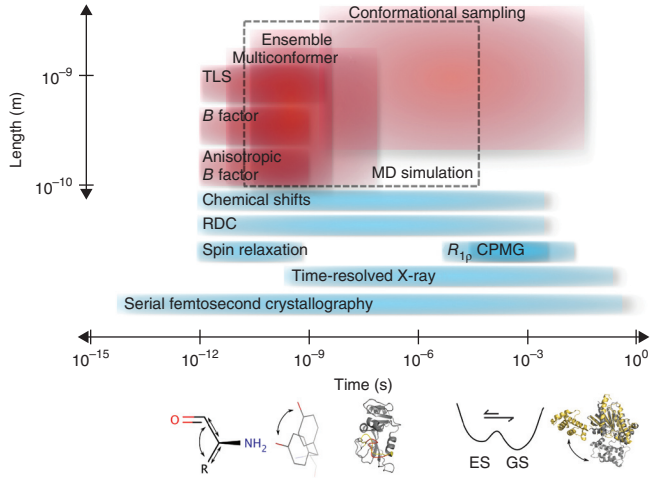

Figure 1 | Protein dynamics across temporal (*x*-axis) and spatial (*y*-axis) scales. Proteins exhibit conformational dynamics ranging from atomic vibrational motions around average positions on the picosecond timescale, (bond vibrations, leftmost cartoon at bottom) to exchanging conformational substates of rotameric side chains, to loop motions, to collective exchanges between the ground state (GS) and the excited state (ES) and increasingly larger substructures at millisecond or even longer timescales (rightmost cartoon at bottom). Experimental techniques to probe structure and dynamics are highlighted in blue, and methods to represent protein conformations or conformational ensembles are highlighted in red. Conventional, synchrotron-based X-ray data can result in different structural characterizations (red), which can additionally provide a structural basis of NMR observables. Picosecond dynamics are commonly modeled with a harmonic *B* factor (local) or translation, libration and screw (TLS) (global) model in crystal structures. Nanosecond-to-microsecond motions result in conformational (anharmonic) substates, which require multiconformer or ensemble models for visualization. Time-resolved X-ray experiments depend on conformational substates frozen into the crystal. NMR order parameters derived from spin-relaxation experiments have established a link between fast protein dynamics in solution and the crystalline state. Chemical shift and residual dipolar coupling (RDC) data measure dynamics spanning nine orders of magnitude. Molecular dynamics (MD) simulations or conformational sampling algorithms can aid in interpreting RDC data. $R_{1\rho}$ and Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments report on exchanging substates at millisecond timescales. Serial femtosecond crystallography in particular enables access to conformational ensembles across many orders of magnitude of timescales.

these representations that suggests that dynamically exchanging networks[10] are a macromolecule's evolutionary unit[11–14], linking evolutionary timescales with molecular timescales through mutual information. In this Perspective, we therefore argue for a view of macromolecules that centers on evolvable, sparse networks of functional, collectively exchanging substates. Looking ahead, new experimental techniques, such as X-ray free electron lasers and terahertz spectroscopy, will enable integrative structural biology studies at ever-higher resolution.

Solution NMR signals are generated by multiple conformations

The classical strengths of X-ray crystallography for precisely determining a unique macromolecular structure are complemented by NMR dynamics experiments that probe how macromolecules shift between conformational substates in solution. Structurally interpreting the data obtained from most NMR dynamics experiments is complicated, however, because the signals are averaged over the ensemble populated in a specific time window (**Figs. 1** and **2**). NMR three-dimensional (3D) structural ensemble determinations from nuclear Overhauser effect (NOE) distance restraints traditionally have bridged the static view from X-ray structures and NMR dynamics observables[15–17]. NMR structural ensembles determine atomic positions from a relatively small number of experimental data points that report on largely local information. Thus, these ensembles are generally underdetermined—even after additional restraints from experiments and the energy function used in structure calculation are accounted for—and contain insufficient information to determine the relative populations of conformers[18]. It can therefore be a challenge to interpret dynamical or functional mechanisms directly from the NMR structural ensemble owing to a lack of precision[19].

Recently, exact NOEs (eNOE), which are based on calibrated NOE measurements that provide a precision of ±0.07 Å for the measured distance between atoms, have been used to improve the precision of ensemble determination. eNOE-derived structural ensembles form a distribution of conformers that averages to the experimental measurements. For example, the eNOE ensemble of protein G (GB3) revealed three exchanging conformational substates that would be impossible to detect by traditional NOE structure determination alone[20]. In addition to distance information encoded by NOEs, the chemical environment information encoded by chemical shifts is becoming increasingly useful for calculating NMR structural ensembles. One such tool, CS-Rosetta, assembles fragments of locally similar sequences with similar patterns of ^{13}C, ^{15}N and ^{1}H chemical shifts to derive the 3D structure of proteins[21] and RNA[22] *de novo*. CS-Rosetta depends critically on its potential energy function to disambiguate the wide variety of structural features consistent with chemical shift values. As chemical shift data are subject to motional averaging, the resulting conformation can inadequately represent areas where the chemical shifts report on multiple conformations. However, recent advances can determine multiple substates over which the chemical shifts are averaged[23].

Putting the puzzle back together: integrative models of dynamic structures. As NMR experimental techniques grow more sensitive, they are revealing not only rich dynamics in classically disordered regions and loops but also collective exchange between conformations within folded domains. Three general strategies are emerging to connect observable NMR dynamics with plausible constituent structures: comparison to X-ray data, sample-and-select methods and biased sampling. The simplest method is to compare the observables to a series of conformations populated in X-ray crystal structures determined under different conditions (for example, mutation, ligand or crystal form) or, as has been done more recently, to conformations from a single X-ray data set collected at room temperature[24,25]. Parallel solid-state and solution NMR experiments also provide reasonable expectations for observing time-averaged data in a single X-ray electron density map.

Alternatively, computational sampling procedures generate many plausible conformations and then select only a subset of these conformations on the basis of agreement with experimental NMR data. These sampling-based procedures allow rapid access to conformational variability, even that of excited states 'hidden' by large energy barriers. In a sample-and-select strategy, the observable is not used in the sampling stage. In contrast, biased

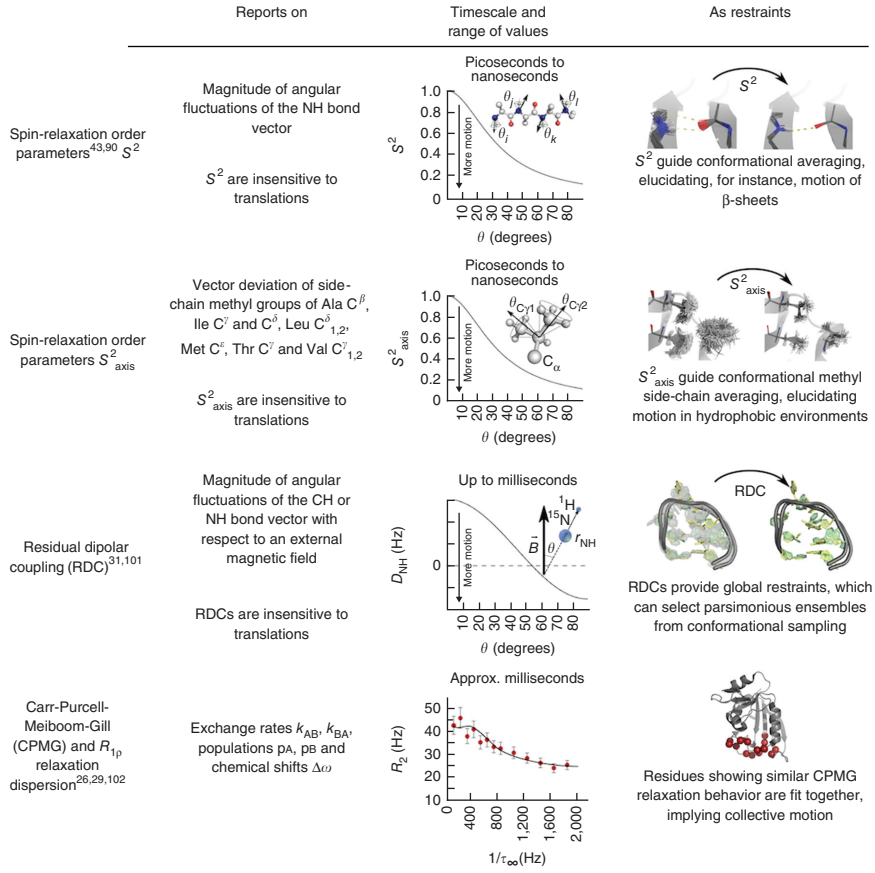

Figure 2 | NMR experiments report on motions across different timescales. The structural basis of these motions and the fitting procedures govern the conversion of these experimental observables into structural restraints. Spin-relaxation order parameters S^2 can guide conformational averaging. RDCs $^1D_{\{N,C\}H}$ provide global restraints[101]. Residues with similar CPMG or $R_{1\rho}$ relaxation dispersion behavior are often fit together in a two-state model[102].

sampling approaches restrain sampling to those conformations consistent with the NMR observable by altering the underlying energy function. Molecular dynamics (MD) simulations have the potential to associate the hierarchies of motions with timescales, but interpretation of NMR observables with biased or unbiased MD simulations requires long trajectories to ensure adequate sampling of conformational space.

Below we outline how these three strategies can yield structures consistent with NMR dynamics measurements across timescales.

Guilty by association: slower timescale data are fit collectively. Collective motions on slow, millisecond timescales underlie important biomolecular processes such as catalysis. An ingenious series of NMR relaxation experiments enable researchers to detect whether specific residues are populating multiple distinct chemical environments. Carr-Purcell-Meiboom-Gill (CPMG) spin-echo and rotating-frame $R_{1\rho}$ relaxation dispersion (RD) experiments, which report on timescales of micro- to milliseconds[26], take advantage of the sensitivity of chemical shifts to identify exchange processes between a major ground state and a minor excited state (**Fig. 1**). Residues exhibiting NMR signals on this timescale are not necessarily mobile; rearrangements of neighboring residues can change the chemical environment of the residue being probed too. For even slower timescales, chemical exchange saturation transfer provides similar information[27]. In a classical application of CMPG experiments, fitting against a dispersion curve[28] revealed residue-specific populations, exchange rates and chemical shift differences for the enzyme dihydrofolate reductase (DHFR) bound to different ligands mimicking states of its catalytic cycle (**Fig. 2**). Because populations and exchange rates are fit across groups of residues, whereas chemical shift differences are individually fit, CPMG implies collective motion. A close monitoring of the quality of the fit was used to determine whether a specific residue belongs to the group undergoing collective exchange[28,29]. But detailed conformational states and precise mechanisms of collective motion remained unresolved from these experiments.

As long timescale and relatively large energy barriers separate the substates identified by RD experiments, developing a motional model using MD simulations is challenging. A potential work-around is to use Markov state models or other acceleration procedures to assemble many short simulations and analyze the resulting model for potential excited states[30]. Alternatively, selection of sampled conformations using the chemical shifts of the excited states, as applied in CS-Rosetta, requires very accurate determination of excited-state chemical shifts (see **Box 1** and **Fig. 3** for an example applied to T4 lysozyme (T4L)). Examining conformational transitions between multiple, often perturbed, crystal structures remains the most common method of inferring the conformational exchanges monitored by RD experiments. In the DHFR example, the chemical shift changes are driven by the exchange of the Met20 loop between distinct conformations throughout the catalytic cycle. The loop residues and those predicted to have their chemical environments altered by the two loop conformations are grouped together. The dynamics of these Met20 loop proximal residues have clear implications for ligand flux, but a proline insertion designed to stabilize the loop also impaired the chemical step of catalysis. A comparative analysis of crystal structures alone was unable to explain the sharply reduced rate and the absence of active-site dispersive signal[28]. However, a carefully designed X-ray experiment and subsequent data analysis revealed perturbed conformational dynamics consistent with the NMR observations[10] (**Box 1** and **Fig. 3**).

Sample-and-select methodology captures structure and dynamics from residual dipolar couplings. Whereas chemical shifts indirectly encode the structural information accessible at slower timescales, residual dipolar couplings (RDCs) report on the structural dynamics of CH and NH bond-vector orientations with respect to a global coordinate frame on timescales up to milliseconds[31] (**Figs. 1** and **2**). In a pioneering study, Tjandra and Bax found that measured RDCs of ubiquitin agreed well with those predicted from a high-resolution crystal structure of ubiquitin[31]. The finding suggested that the crystal structure captured the average conformation of ubiquitin in solution. However, interpretation

BOX 1 NMR AND CRYSTALLOGRAPHY TOGETHER CAN BRIDGE STRUCTURE AND DYNAMICS

Same kinetic effects, but divergent mechanisms. Targeted, allosteric mutations of CypA and ecDHFR (**Fig. 3a**) superficially had the same effects: sharply reduced catalytic rates, which were difficult to explain from NMR data or crystallography alone. We recently proposed a program to integrate NMR and high-resolution crystallography[9]. By analyzing how strain owing to van der Waals overlaps propagates through a network of alternate conformations, we found highly divergent mechanisms underlying the reduced rates. In CypA, an S99T 'traffic jam' mutation resulted in an overpacked core, limiting conformational exchange[24]. Similarly, NMR studies of ecDHFR-$NADP^+$-FOL found no dispersion throughout the active site. Initial, cryogenic crystallography data sets indicated that the active site was structurally unaffected by the mutations. Subsequent room-temperature experiments, interpreted by a multiconformer model, paradoxically identified an increase in conformational heterogeneity. A further CONTACT network analysis revealed that FG loop residues had perturbed the active-site network, leading to nonproductive interactions[10]. These orthogonal interpretations of similar observations are difficult to achieve with NMR or X-ray crystallography alone and would lead to radically different mutational trajectories to restore or alter activity.

Dynamics guide structural characterization: T4 lysozyme and HIV-1 TAR. Unlike wild-type T4L, an L99A mutant[104] experienced millisecond conformational exchange between a ground and an excited state[105]. Although the ground state is structurally indistinguishable from the wild-type state, the excited state, populated to only 3%, long remained elusive. In a recent *tour de force*, Rosetta-CS provided an atomically detailed structure of the excited state from backbone ^{15}N, ^{13}C and ^{1}H chemical shifts obtained from Carr-Purcell-Meiboom-Gill (CPMG) experiments[106,107] (**Fig. 3b**). Two further mutations were shown to invert the populations, thereby evolving the protein—by design—to take on new functional roles.

The highly flexible apical loop of the 59-nt stem-loop HIV-1 TAR binds human cyclin T1 and viral *trans*-activator protein Tat, which activate and enhance transcription of the HIV-1 genome[108,109]. NMR $R_{1\rho}$ relaxation dispersion (RD) measurements combined with mutagenesis suggested conformational exchange between the ground state of the apical loop and a hidden excited state, populated to 13% (ref. 110). A conformational sampling procedure, KGSrna, structurally characterized the loop in solution by resolving averaged residual dipolar coupling (RDC) data into contributions of constituent members via a small ensemble obtained by fitting the experimental data[40] (**Fig. 3c**).

In these cases, transient excited states, inaccessible by crystallography, were revealed from solution data through a computational intermediary. As our computational abilities are rapidly developing but far from mature, those successes often rely on the availability of high-resolution structural data. In L99A T4L, the computational complexity was reduced by remodeling and matching to observations only those parts of the crystal structure that showed divergence from the ground state in relaxation data. Likewise, in HIV-1 TAR, the stem served as a scaffold to probe the conformational landscape of the loop.

Active sites revealed by reduced conformational diversity. Wüthrich and coworkers developed a systematic procedure to compare high-resolution crystal and NMR structures[111]. They examined data from two globular proteins, a *Thermotoga maritima* anti-σ-factor antagonist and a mouse γ-glutamylamine cyclotransferase[112], obtained with NMR and cryogenic X-ray crystallography. Their analysis determined that sites that exhibited conformational exchange on the millisecond timescale and that showed elevated conformational variability in NMR solution but not in the crystal coincided with active sites. In the crystal structures, polar residues generally adopted conformations to satisfy hydrogen-bonding interactions with each other or with the components of the buffer. Lattice formation, as well as cryogenic temperatures, likely played a role in stabilizing these conformations. In γ-glutamylamine cyclotransferase, a collapsed cavity surrounding the catalytic site further stabilized amino acids (**Fig. 3d**). In contrast, chain termini and surface-exposed loops retained their conformational diversity in the crystal. Coordinated, conformational exchange of the millisecond timescale is often indicative of functional activity, suggesting that a comparative and complementary analysis of solution and cryogenic crystal structures can be a first step to identify molecular mechanisms.

of RDCs by comparative crystal analysis is impeded by the wide range of timescales—up to 12 orders of magnitude—spanned by RDCs. It was soon realized that not just structural interpretations but also slow, diffusive motions could be extracted from RDCs[32].

For this to happen, multiple experiments with different alignment media are generally required as RDC restraints are underdetermined. Although, in principle, an exact solution for the orientation of a peptide is available with a minimum of two RDCs per residue, this procedure does not accommodate motionally averaged data[33]. Ensemble refinement improves substantially when RDCs from multiple alignment media are used as restraints alongside NOEs[34]. RDCs measurements share a common, global reference frame, which implicitly provides long-range structural information that is lacking in interatomic NOEs. To derive a conformational ensemble that recovers averaged RDCs, researchers can generate many possible ensembles,

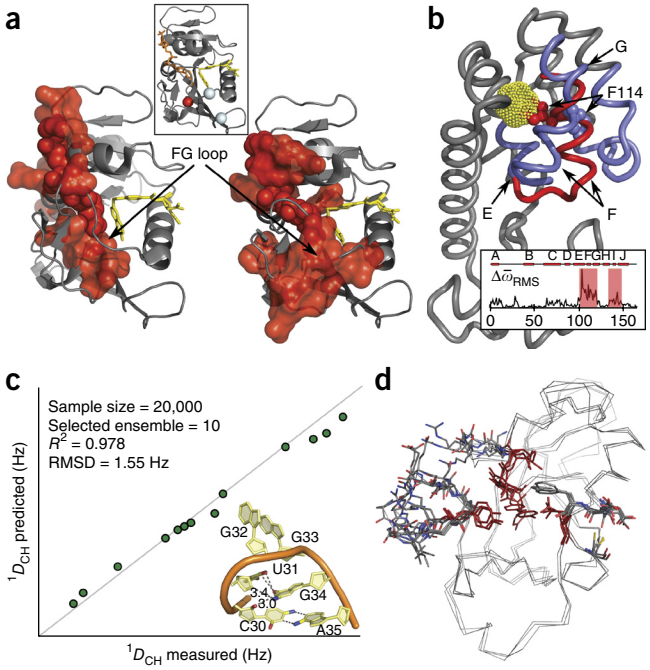

Figure 3 | Examples of synergistic insights from NMR and X-ray crystallography. (**a**) Allosteric mutations to wild-type ecDHFR-NADP+-FOL (left; mutations shown as cyan spheres in inset) abrogated the chemical step of catalysis. RD measurements indicated that millisecond exchange in the active site was absent except for one amino acid (inset, red sphere). CONTACT network analysis revealed that FG loop amino acids led to frustration of active-site functional motions (right). (**b**) Ground state (blue) and excited state (red) of the C-terminal domain of L99A T4L. The mutation results in a cavity, shown in yellow contour. Fragments 100–120 and 132–146 were remodeled with CS-Rosetta from CPMG RD chemical shifts $\Delta\bar{\omega}_{RMS}$ (inset), starting from the ground-state crystal structure. Helices F and G adopt different conformations between the ground and excited states. (**c**) Excited state of HIV-1 TAR characterized from RDC measurements. A sample-and-select procedure (KGSrna) identified a ten-member ensemble from 20,000 samples that agreed with RDC measurements to within experimental error. A representative from the ensemble resembling the excited state was further optimized to obtain base pairs consistent with the RDCs $^1D_{CH}$ (inset). RMSD, r.m.s. deviation. (**d**) Twenty-four amino acids in the NMR bundle of mouse γ-glutamylamine cyclotransferase have per-residue displacements exceeding the mean values for the entire polypeptide chain but low *B* factors in the crystal structure, while exhibiting millisecond conformational exchange. These amino acids include six of nine catalytic residues (red) and surround the active site. This suggests that sites with elevated structural disorder and slow exchange in solution, while ordered in the crystal structure, can indicate functional relevance.

and the maximum-entropy principle can then be used to select a representative ensemble that minimizes force field bias[35]. Simulations with replica-averaged RDCs, which satisfy the maximum-entropy principle, accurately characterized domain motions in hen lysozyme[36]. However, an atomically detailed structural interpretation of collective motion is not straightforward unless the structural ensemble is restricted to a few members.

By contrast, a sample-and-select procedure can identify a parsimonious set of distinct conformational substates from a large pool of sampled conformers by matching back-calculated RDCs to measured ones[37,38]. In one example, this approach identified one major and two minor conformational substates sampled with accelerated MD in the protein SH3C from CD2AP's free-energy landscape from a large set of RDC measurements and established a hierarchy of motions and states[39]. Moreover, this approach directly addresses the challenge of separating fast, subnanosecond, within-state motions from slower, between-state motions.

Additionally, the large energy barriers associated with the long timescales of RDCs favor conformational sampling procedures over MD simulations. A sampling procedure for RNA that treats hydrogen bonds of Watson-Crick pairs as distance constraints was shown to probe the conformational landscape according to a distribution similar to that of RNA in solution. Together with an exact RDC-based ensemble selection procedure that determines both the size and the weights of the structural ensemble, the procedure resulted in an atomistic characterization of an excited state of the highly flexible HIV-1 TAR apical loop[40]. A similar sparse ensemble sampling and selection procedure was applied to ubiquitin[41]. A hybrid approach, in which sampling-based seed conformations provided structural diversity that were subsequently relaxed with RDC-biased MD simulations, resulted in a complete characterization of the recognition dynamics of ubiquitin[42]. Calculating a structural ensemble consistent with RDCs across multiple alignment media was key to structurally interpreting the dynamics, and comparisons to multiple crystal structures validated the functional connections to the binding mechanisms of ubiquitin.

Independent motions at fast timescales report on conformational entropy. Whereas RDCs contain some structural information to constrain the modeling of dynamics, model-free Lipari-Szabo (LS) spin-relaxation order parameters S^2 and S^2_{axis} report on an extent of order but not the conformations encoding that order. This class of motion characterizes deviations of backbone amide and methyl-containing side chains at the picosecond-to-nanosecond timescale[43] (**Figs. 1** and **2**).

Wand and coworkers used the model system Ca^{2+}-calmodulin, which recognizes hundreds of peptides of diverse sequences, to test the hypothesis that changes in side-chain order parameters are indicative of changes in the conformational entropy of the resulting complexes[44,45]. Their work showed a linear correlation between the calculated change in conformational entropy based on methyl side-chain order parameters and the overall binding entropy measured by calorimetry. Similar trends were extended to catabolite activator protein[46], a transcription factor, and galectin, a carbohydrate-binding protein[47].

Among NMR observables, order parameters are most easily accessed with MD simulations owing to the similarity of experimental and simulated timescales. Yet, poor correlations are generally observed for S^2 spin-relaxation order parameters from unrestrained simulations[48], with slight improvements when the overall rotational diffusion of the molecule is considered[49]. Restrained ensemble simulations allow averaging of the S^2 spin-relaxation order parameters, in space and time, to closely match true motional averaging in solution and greatly improve correlation between observed and simulated values[48].

Like S^2 order parameters, side-chain order parameters, S^2_{axis}, predicted from MD trajectories, generally agree weakly with experimental data[50,51] unless the dynamics are biased with S^2_{axis}-based restraints[48]. Only modest agreement (coefficient of determination $R^2 \approx 0.65$) of site-to-site correlations between methyl side chains were reported in a study reporting MD simulations from 60-ns unbiased trajectories for seven proteins[45]. However, the total side-chain conformational entropy changes calculated from the trajectories generally agreed with the trends observed by NMR and calorimetry. This suggests that nonmethyl side chains experience comparable conformational averaging and that the fast motion that is measured experimentally is a good proxy for the overall flexibility of protein side chains.

Simultaneous interpretation of NOEs and S^2 order parameters of ubiquitin by an ensemble MD simulation dramatically improved agreement with experimentally determined S^2 order parameters[52]. The atomistic, motional model obtained from the simulation revealed a highly dynamic ensemble of conformations, with liquid-like behavior for side chains, even in the core, and solid-like behavior for the backbone. The ensembles interconverted at the pico- to nanosecond timescale, which is consistent with independent RDC observations. Interestingly, although Monte Carlo sampling usually has no timescale, it can potentially be recovered by matching the samples with experimental data. For instance, excellent correlations, exceeding 92% on average, between calculated and measured S^2 order parameters were observed for four well-characterized proteins using a sampling technique[53,54]. Similarly, calculating order parameters from ensembles of independently determined high–sequence similarity crystal structures provided good agreement with experimental values[55]. Thus, simulations, sampling and multiple crystal structures are all generally useful for providing motional models of fast-timescale dynamics.

The connections between fast-timescale dynamics and entropy are an important experimental validation of the "dynamic allostery" paradigm put forward by Cooper and Dryden[56]. They proposed an explanation for how entropic changes arise without detectable changes to the average conformation, i.e., small enthalpic changes are dwarfed by the residual entropy changes. For example, a binding-interface side chain that adopts slightly different backbone conformations upon recognizing distinct binding partners can have minor enthalpic consequences. However, distinct backbone conformations can alter the packing of all the residues nearby. This scenario will lead to global changes in side-chain flexibility that propagate away from the binding site with very little effect on binding enthalpy or direct interactions. Residue conformational distributions must be coupled sufficiently strongly so that the changes are distributed throughout the protein but not so strongly as to reduce the number of functionally relevant microstates accessible to the protein. Despite the potential to optimize the potency of specific interactions in drug design and molecular recognition, modulating (marginal) conformational entropy by allosteric point mutations or binding events remains a major challenge[57].

Characterizing dynamic processes in crystals

The successes of leveraging multiple independent crystal structures to provide a structural basis for dynamics raise an intriguing connection between the specific time windows probed by NMR and the time-independent view of X-ray crystallography. Solid-state NMR in particular allows us to ask: what kind of motions can we expect to see in the crystal, and how do the barriers between conformations change in the crystal lattice?

Surprisingly, the influence of the crystalline environment on fast dynamics is modest at physiological temperatures. Magic-angle-spinning solid-state NMR (MAS ssNMR) enables the study of structural dynamics of crystalline proteins[58,59]. It reveals motions on timescales spanning 9–10 orders of magnitude through relaxation data R_1, $R_{1\rho}$ and dipolar coupling[60,61]. In solid-state experiments, dipolar coupling data, which are completely determined by local dynamics, are unaffected by rotational tumbling, thus making this timescale (5–15 ns) accessible. Solution-state picosecond-nanosecond order parameters agree very well with those measured in the crystalline state[62]. The solution $R_{1,solution}$ and crystalline MAS ssNMR $R_{1,solid}$ relaxation rates of side-chain methyl groups in the hydrophobic core of R-spectrin SH3 domain were also highly correlated[63] (**Fig. 4a**).

However, slower, microsecond-millisecond conformational exchange quantified by CPMG RD experiments revealed subtle differences between microcrystals and solution-state processes for ubiquitin[64]. V70, which exhibits conformational exchange in solution, showed no dispersion in the crystal. A cluster of residues (I23–N25 and T55) showed exchange in solution and in the crystal, but the solid-state exchange rates were tenfold slower. A potential explanation for this discrepancy is that T55 is part of a β-turn that adopts a type I conformation in solution (**Fig. 4b**) but a type II conformation in the crystal (**Fig. 4c**). CPMG measurements indicated an exchange between type I and type II conformations for this β-turn in solution[65,66]. A microcrystal X-ray structure collected at cryogenic temperature used to interpret the MAS ssNMR measurements suggests that the ground and excited states are reversed in solution and microcrystals. The type II conformation is stabilized by crystal contacts, which likely means that a free-energy barrier slows the exchange process (**Fig. 4b**,**c**). However, a room-temperature X-ray crystal structure

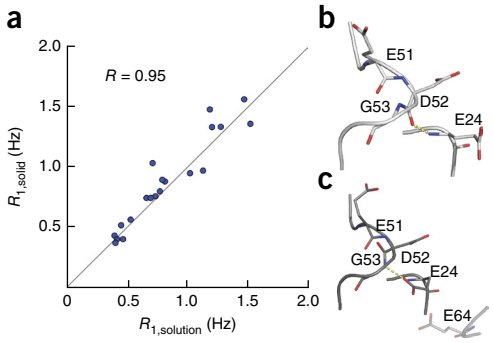

Figure 4 | At physiological temperatures, crystalline environments mildly affect biomolecular motions. (**a**) R_1 relaxation rates for methyl side chains of R-spectrin SH3 in solid and solution state[63]. The correlation coefficient R between the data obtained from the two environments is 0.95, suggesting highly similar motions. Data points are expected to lie along a 45° line if there are no differences between the crystalline and the solution state. (**b**) Solution structure of human ubiquitin exhibiting a type I conformation β-turn. (**c**) A type II conformation β-turn in microcrystalline human ubiquitin. A peptide flip of D52 is stabilized by a hydrogen bond to E24 and by a water-mediated hydrogen bond to crystal contact K63 (not shown). Furthermore, E24 is stabilized by crystal-contact E64.

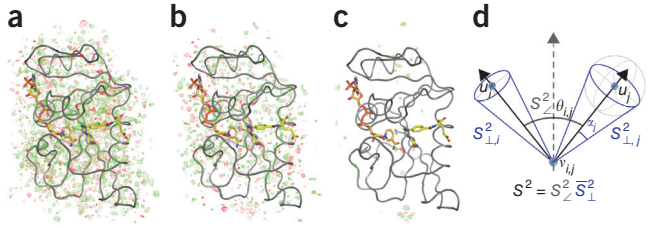

Figure 5 | Cryo-cooling of protein crystals irregularly selects conformational substates. Isomorphous $F_o - F_o$ maps of two independently collected pairs (in 2005 and 2013) of room-temperature (RT) and cryogenic data sets (Cryo) of ecDHFR-NADP$^+$-FOL are shown, contoured at 0.4 $e^-/Å^3$ (ref. 103). (**a**) The RT_{2005} – $Cryo_{2005}$ map shows widespread positive difference peaks (green; red indicates negative peaks), demonstrating that RT data sets exhibit elevated conformational heterogeneity. (**b**,**c**) The $Cryo_{2005}$ – $Cryo_{2013}$ difference map (**b**) shows peaks of both signs, which are absent in the RT_{2005} – RT_{2013} data (**c**), thus pointing to irregular and unpredictable changes in structure and dynamics owing to cryo-cooling. (**d**) The Fenwick-Wright framework relates crystallographic, isotropic atomic-displacement parameters obtained from anharmonic substates to order parameters in solution. The angular order parameter, $S^2_{\angle}$, reports on angular diffusion between discrete states of atoms u_i and u_j through angle $\theta_{i,j}$. The orthogonal order parameter, $S^2_{\perp}$, reports on angular diffusion of bond vectors within states through, for instance, angle α_j. The method revealed excellent agreement for atomic displacements measured with X-ray crystallography and NMR in solution[89].

of human ubiquitin at a comparable pH but determined in a different space group adopts a type I conformation for this β-turn[67], leaving unresolved the effects of cryo-cooling and lattice formation in reversing the ground and excited states[68].

These ssNMR results hint that comparing multiple conformations determined within a single X-ray diffraction experiment might be a fruitful avenue for providing a structural basis for dynamics measured in solution. In parallel, new experimental and computational techniques in crystallography are expanding the potential for modeling low-occupancy conformations and directly proposing models of conformational coupling.

An increasing role for X-ray crystallography to provide the structural basis for dynamics. High-resolution X-ray crystallography data suggest that a macromolecular ensemble can lead to conformational substates in the crystal, manifested as lattice disorder or dynamically exchanging sites[69]. For example, high-resolution data of crystalline Ca^{2+}-calmodulin revealed a hierarchy of motions, from harmonic displacements to a discrete conformational ensemble[70]. This hierarchy is based on modeling to fit an electron density map, which is a 3D distribution of electronic charge derived from an X-ray diffraction experiment. The electron density value at each point of the unit cell is calculated from a global transformation involving all measured, diffracted intensities and is therefore affected by all atoms of the structural model. This global influence of local modeling brings its own challenges in developing a motional model for the crystalline ensemble[69]. Additionally, diffraction data are routinely collected at cryogenic temperature to reduce radiation damage. As modeling and refinement techniques have advanced, it became apparent that flash cooling idiosyncratically alters the structure and dynamics of crystalline samples[25,68] (**Fig. 5**).

Traditionally, a crystal structure is presented as a single, unique conformer with isotropic or anisotropic atomic-displacement parameters, or *B* factors (**Box 2**). These displacement parameters account for harmonic deviations from an average position. A translation, libration and screw model can additionally account for anisotropic rigid-body deviations for groups of atoms[71]. Elastic normal mode[72] or deformable elastic network[73,74] refinement add additional physicochemical restraints to accommodate sparse, low-resolution data. Supplemental NMR restraints can aid interpretation of crystallography data with a traditional, single-conformer model[75,76]. Although these harmonic parameterizations improve agreement of the model with the data, identifying collective motion is hindered by imposed motional models, which generally assume equilibrium deviations and statistical independence of anisotropic displacement parameters[77].

In twin[78] or fixed-ensemble[79,80] refinement (**Box 2**), multiple independent copies of the molecule are subjected to restrained MD simulations. The conformational sampling provided by the simulations then simultaneously accounts for the data. First applied to high-resolution RNase A and crambin X-ray data collected at subphysiological temperature, but well above the glass-transition temperature, the models captured conformational substates encoded by the X-ray diffraction data. Distinct hydrogen-bonding interactions were observed between the conformers, suggesting that important functional information is hidden in the single-conformer treatment of X-ray data. Fixed-ensemble refinement revealed correlated motions and early evidence of conformational selection in a binding pocket of α-lytic protease, even at cryogenic temperature[81].

Time-averaged, rather than instantaneous, X-ray restraints in MD simulations result in a structural ensemble weighted by the Boltzmann distribution. In time-averaged ensemble refinement, the pseudo-energy is augmented by a weighted time-averaged ensemble of tens to hundreds of conformations from MD simulations[82]. Although its early applications were shown to overfit the data[83], a recent implementation has improved cross-validation statistics[84]. A highly complementary approach underlies the multiconformer model, as implemented in qFit, which determines an optimal set of 1–4 conformers that, together with their occupancies, collectively best fit the local electron density around each amino acid[85] (**Box 2** and **Fig. 6**).

Interpreting ensembles. Deviating coordinates among an ensemble obtained from repeatedly and individually rebuilt, independent, single-conformer interpretations of data were found to reflect uncertainty rather than actual heterogeneity[86]. However, a fixed or time-averaged ensemble is irreducible, which makes interpretation of its coordinates more complex. Fixed or time-averaged structural ensembles of conformers and a multiconformer model are complementary techniques in two important respects. First, their strengths are at opposite ends of the data resolution spectrum. At high resolution, better than 1.5 Å, the improvement in the cross-validation measure R_{free} (ref. 87) resulting from refinement of multiple, independent copies of the molecule appears to be diminished[84,85]. In contrast, the largest improvement in cross-validation statistics occurs at high resolution for a multiconformer model, with correlated alternate conformations[85]. At medium to low resolution, where structural ensemble models generally result in improved data fit, the

BOX 2 SINGLE-CONFORMER, ENSEMBLE AND MULTICONFORMER MODEL INTERPRETATIONS OF CONFORMATIONAL HETEROGENEITY

Single-conformer model. A single-conformer model with *B* factors represents harmonic vibrations of atoms. *B* factors require one (**Fig. 6**, isotropic, left column) or six (anisotropic, middle column) additional variables to parameterize an atom. *B* factors are visualized by a circle or ellipse, scaled such that it contains the atom with 50% probability. A single conformer often accounts for distinct substates by modeling atoms at the highest levels of (overlapping) density (middle column). Regions with weak or ambiguous experimental restraints (right column) are modeled with a single conformer and elevated *B* factors.

Ensemble model. An ensemble model uses the same number of conformers for all residues, irrespective of local experimental restraints. Conformers and *B* factors simultaneously represent harmonic deviations. Each conformer accounts for a preset fraction of the total experimental data. The relative number of conformers in anharmonic substates can reveal populations (middle column). Weak or ambiguous experimental restraints often lead to indiscriminate conformations (right column).

Multiconformer model. A multiconformer model introduces up to four conformations for each residue as needed to collectively, locally explain the experimental data. *B* factors represent harmonic deviations (left column), whereas conformers represent anharmonic deviations (middle column). Occupancies *q* represent the fraction of experimental data accounted for by the conformer. Weak experimental restraints result in a single conformer with elevated *B* factors, as in the single-conformer model (right column).

multiconformer approach is unable to distinguish discrete, alternate conformations owing to a loss in multimodality of electron density profiles[88]. Second, coordinates of a structural ensemble diverge in areas where experimental restraints are lacking, whereas their conformers tend to cluster in areas where experimental restraints are abundant. Owing to the very low occupancy of each conformer, contributions to the total scattering of a locally divergent structural ensemble could approximate that of noise or bulk-solvent levels, improving agreement with observations compared to a single conformation at full occupancy. By contrast, a qFit multiconformer model includes additional conformers for a residue only when required for locally fitting the data. If conformational heterogeneity is not clearly present in the data, qFit will model an average conformation instead, with elevated *B* factors distinguishing the relative disorder about that conformation.

Collective motions are the next frontier. One of the major challenges of experimental biophysics is directly measuring correlated motions within macromolecules. The conformations populated in a multiconformer model[85] can provide a structural interpretation of uncoupled solution-state backbone NH (S^2) and methyl axis S^2_{axis} order parameters[89]. A recent study expanded the original Brüschweiler-Wright framework[90] relating atomic-displacement parameters to order parameters to include conformational substates from a qFit multiconformer model[89] (**Fig. 5d**). Atomic-displacement parameters obtained from room-temperature crystallography data sets were in excellent agreement with solution-state motions, confirming an earlier analysis based on MD simulations[91]. By contrast, atomic-displacement parameters obtained from cryogenic data sets showed little agreement. The multiconformer crystal structure thus provides a motional model of NMR measurements, structurally characterizing the solution-state NMR dynamics parameters at fast timescales. The challenge is to subsequently refine these conformations into a representation of correlated heterogeneity to provide a structural basis for slower motions occurring within folded domains. Contact networks through alternate conformation transitions (CONTACT networks), which are composed of dynamically interacting residues identified directly from high-resolution X-ray data, visualize intramolecular conformational ensembles[10]. CONTACT networks of residues undergoing collective motion have provided a structural basis for NMR relaxation data and explained how mutations affect both catalytic rate and protein conformational dynamics for the *Escherichia coli* DHFR (ecDHFR)-$NADP^+$-FOL complex. Distinct networks were implicated in particular enzymatic activities or mechanisms in the complex (**Box 1**). The largest network connects the functionally important FG loop to the active site. Other networks are known to rearrange hydrogen-bonding networks to facilitate ligand flux or encompass hinge residues. CONTACT networks

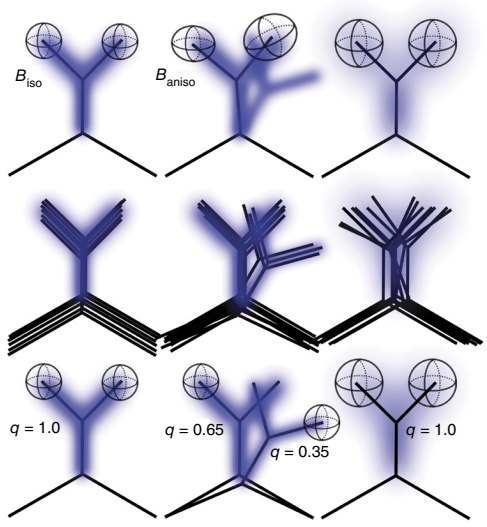

Figure 6 | In X-ray crystallography, resolution and model selection interact to affect the interpretation of conformational heterogeneity. Electron density (blue) at progressively worse resolution can be fit by different classes of models (black lines).

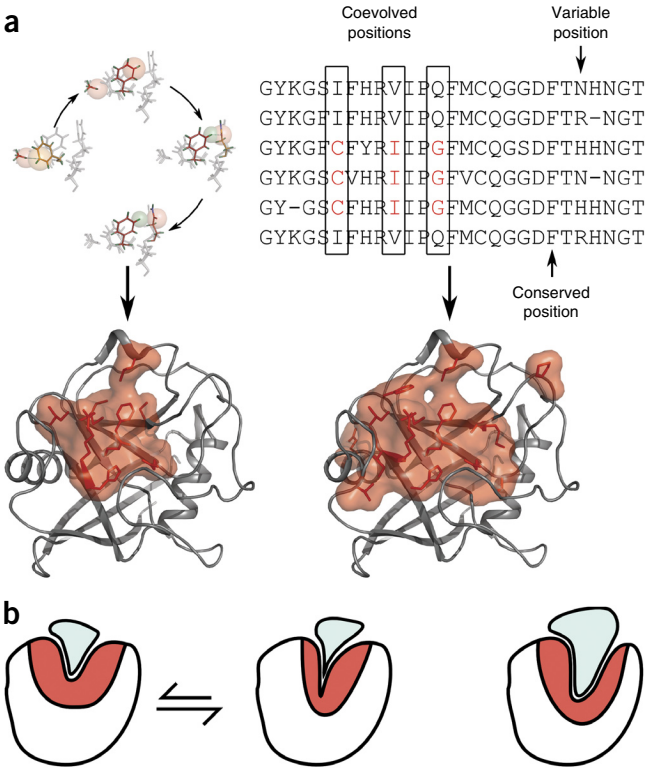

Figure 7 | Networks of conformational exchange are evolution's engines. (**a**) CONTACT networks are identified from clash-and-relieve pathways of alternate main- and side-chain conformers in a multiconformer qFit model (top left). Pathways that share residues are grouped into networks. A CONTACT network representing conformational exchange in the enzyme cyclophilin A is shown as a red surface on the molecule (bottom left). Sectors are networks composed of coevolving amino acids identified from statistical coupling analysis (right). The similarities between the CONTACT network and the sector are striking, suggesting that conformational exchange may possibly be a phylogenetic instrument that enables members of the family to evolve toward specific functions and accommodating a wide variety of ligands. (**b**) Two distinct scenarios follow from this theory. Within species, networks (red) are optimized to evolve new function by enabling conformational exchange between substates to bind to (functionally) different and new partner molecules (left). Across species, networks could be further specialized, enabling new functions and/or losing their ability to exchange with substates associated with previous functions (right). Data partially provided by D.R. Hekstra and K.I. White (UT Southwestern Medical Center, personal communication).

allow us to view biomolecules as interconnected, deformable and functional networks rather than just linear, branched polymers, fully integrating NMR relaxation data with X-ray crystallography in a 4D space-and-time representation.

Intriguingly, similar functional networks, or 'sectors', were revealed independently by studies of coevolving amino acids in PDZ domains[13] and the S1A family of serine proteases[12]. Sectors, like CONTACT networks, are functional units of contiguous residues in the protein fold independent of sequence number or tertiary structure. Sectors and CONTACT networks for cyclophilin A are remarkably alike, suggesting shared functional mechanisms among evolutionarily related proteins despite differentiated specificity. An important question that requires immediate attention is whether agreement with statistical coupling analysis results from a conserved packing geometry, a conserved subset of motions, or both. Establishing a direct correspondence between sectors and CONTACT networks would further experimentally validate a dynamic view of protein evolvability[11] (**Fig. 7**). Additionally, these findings offer the tantalizing possibility of integrating evolutionary data in characterizing conformational dynamics from crystallography and NMR experiments.

Outlook

New techniques and sources to illuminate conformational dynamics. As new conformational dynamics data emerge, new tools for their representation will be necessary. It is extraordinarily difficult for a single researcher to synthesize complementary data sources into a comprehensive, biophysical model. Multiconformer models and CONTACT networks, or ensembles of multiple conformations, similar to those currently used in NMR or for time-averaged X-ray refinement, show great promise to shed light on conformational dynamics and functional mechanisms. New, more rigorous methods are facilitating selection and comparison of these ensembles[92–94]. The relative weighting of different conformational states in the ensemble, currently not often performed, will be necessary[40,85].

Moreover, these ensembles need to help uncover how the distinct conformations relate to each other in time. If the ensemble

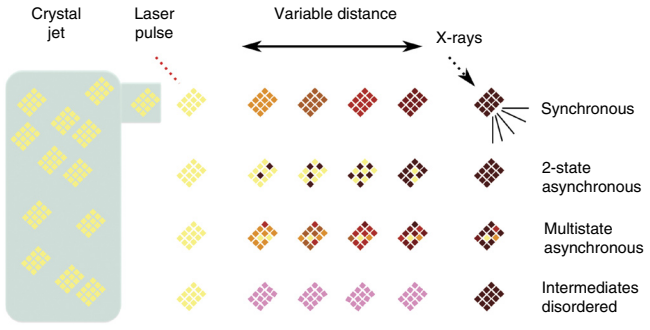

Figure 8 | Illustration of serial femtosecond crystallography. Nanocrystals are extruded from a jet into the X-ray free-electron laser beam. In time-resolved studies, a 'pump' laser beam is placed in the path of the crystal. The laser pulse can uncage a substrate or excite a naturally occurring chromophore, starting a chemical process. The small crystal size can also allow rapid mixing of substrates, enabling the possibility of monitoring enzymatic reactions. Varying the distance between the laser and X-ray pulses would intercept the process at different times, resulting in a molecular movie. We depict four possible scenarios of the conformational transitions, represented by colors, after excitation by the laser. In the top scenario, all unit cells are synchronized through a series of conformational changes represented by the different colors. This scenario gives a straightforward interpretation: as the distance is varied, the electron density map changes from one state to the next. Next, we show a two state system in which some of the unit cells switch to the dark state with no detectible intermediates. Here, occupancy refinement at high resolution can determine the relative populations. Complications arise when conformational changes are asynchronous, as depicted in the third scenario. Prior knowledge of the conformational landscape is essential to determine the shifting occupancies of different states. In the fourth scenario, the lattice becomes disordered as the conformational changes occur, resulting in a loss of diffraction resolution (pink crystals). If the lattice stabilizes in a new conformation, information about the kinetics, but not the intermediate structures, can be extracted from the experiment.

is sequenced like a movie, how can we model multiple potential transitional pathways between conformations? Serial femtosecond X-ray crystallography at X-ray free-electron laser (XFEL) sources opens up unprecedented opportunities to study irreversible reactions, limited only by our imagination (**Fig. 8**). A 'diffraction before destruction' strategy illuminates millions of nanosized crystals injected into the path of ultrafast, femtosecond X-ray pulses that terminate before radiation damage occurs[95]. A recent development avoids the requirement for large quantities of crystals by mimicking a synchrotron diffraction experiment using a handful of microsized or larger crystals (serial femtosecond rotation-oscillation X-ray crystallography), greatly expanding the accessibility of XFEL crystallography[96]. Major challenges and opportunities lie ahead in designing and engineering time-resolved experiments at XFELs, in both the experimental hardware and biochemistry of the samples. For instance, designing samples that can be reversibly cross-linked upon a photon trigger or that include cavities containing photocaged substrates requires knowledge about how the protein fluctuates at equilibrium. Terahertz lasers may also allow a general strategy for probing excited correlated motional modes of the protein in the crystal[97], and diffuse scattering can be used to distinguish between potential models of correlated motion[98]. Currently, time-resolved X-ray studies decompose diffraction data into time-independent structures and transition rates between them[99]. This type of mathematical model can be extended to a more comprehensive framework such as Markov state models that are currently used to summarize MD trajectories in a network representation[100]. Successfully weighting diverse data types (for example, NMR, X-ray and simulation data) in these network representations will be essential for formulating and testing new hypotheses about the role of the conformational transitions in biological function.

ACKNOWLEDGMENTS

We are grateful to M. Vendruscolo and R. Ranganathan for stimulating discussions and to D.R. Hekstra and K.I. White, Green Center for Systems Biology, UT Southwestern Medical Center, for providing data for **Figure 7**. H.v.d.B. is supported by the US National Institute of General Medical Sciences Protein Structure Initiative (U54GM094586) at the Joint Center for Structural Genomics and by SLAC National Accelerator Laboratory LDRD (Laboratory Directed Research and Development) grant SLAC-LDRD-0014-13-2. J.S.F. is supported by US National Institutes of Health grants OD009180 and GM110580, US National Science Foundation grant STC-1231306, the Kinship Foundation Searle Scholar Program, the Pew Charitable Trusts Scholars Program and the David and Lucile Packard Foundation Fellowship.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Boehr, D.D., McElheny, D., Dyson, H.J. & Wright, P.E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638–1642 (2006).
CPMG relaxation dispersion experiments characterize the structural dynamics of the catalytic cycle of DHFR as a sequence of 'linked' ground and excited states.
2. Henzler-Wildman, K.A. *et al.* Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838–844 (2007).
3. Fleishman, S.J. *et al.* Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816–821 (2011).
4. Whitehead, T.A. *et al.* Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* **30**, 543–548 (2012).
5. Linder, M., Johansson, A.J., Olsson, T.S.G., Liebeschuetz, J. & Brinck, T. Computational design of a Diels-Alderase from a thermophilic esterase: the importance of dynamics. *J. Comput. Aided Mol. Des.* **26**, 1079–1095 (2012).
6. Volkov, A.N., Worrall, J.A.R., Holtzmann, E. & Ubbink, M. Solution structure and dynamics of the complex between cytochrome *c* and cytochrome *c* peroxidase determined by paramagnetic NMR. *Proc. Natl. Acad. Sci. USA* **103**, 18945–18950 (2006).
7. Tang, C., Iwahara, J. & Clore, G.M. Visualization of transient encounter complexes in protein-protein association. *Nature* **444**, 383–386 (2006).
8. Privett, H.K. *et al.* Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. USA* **109**, 3790–3795 (2012).
9. Ward, A.B., Sali, A. & Wilson, I.A. Integrative structural biology. *Science* **339**, 913–915 (2013).
10. van den Bedem, H., Bhabha, G., Yang, K., Wright, P.E. & Fraser, J.S. Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat. Methods* **10**, 896–902 (2013).
This study identifies networks of coordinated motion directly from room-temperature X-ray crystallography data, which rationalize NMR data.
11. Tokuriki, N. & Tawfik, D.S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
12. Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138**, 774–786 (2009).
13. McLaughlin, R.N. Jr., Poelwijk, F.J., Raman, A., Gosal, W.S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491**, 138–142 (2012).
This study identifies functional networks of coevolving amino acids from statistical coupling analysis, rationalized by deep mutational scanning.
14. Bhabha, G. *et al.* Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat. Struct. Mol. Biol.* **20**, 1243–1249 (2013).
15. Williamson, M.P., Havel, T.F. & Wüthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182**, 295–315 (1985).
16. Wüthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley, 1986).
17. Torda, A.E., Scheek, R.M. & van Gunsteren, W.F. Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat. *J. Mol. Biol.* **214**, 223–235 (1990).
18. Bonvin, A.M.J.J. & Brünger, A.T. Do NOE distances contain enough information to assess the relative populations of multi-conformer structures? *J. Biomol. NMR* **7**, 72–76 (1996).
19. Rieping, W., Habeck, M. & Nilges, M. Inferential structure determination. *Science* **309**, 303–306 (2005).
20. Vögeli, B., Kazemi, S., Güntert, P. & Riek, R. Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. *Nat. Struct. Mol. Biol.* **19**, 1053–1057 (2012).
21. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. USA* **105**, 4685–4690 (2008).
22. Sripakdeevong, P. *et al.* Structure determination of noncanonical RNA motifs guided by ^{1}H NMR chemical shifts. *Nat. Methods* **11**, 413–416 (2014).
23. Camilloni, C. & Vendruscolo, M. Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. *J. Am. Chem. Soc.* **136**, 8982–8991 (2014).
24. Fraser, J.S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673 (2009).
25. Fraser, J.S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. USA* **108**, 16247–16252 (2011).
26. Hansen, D.F., Vallurupalli, P. & Kay, L.E. Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J. Biomol. NMR* **41**, 113–120 (2008).
27. Long, D. *et al.* A comparative CEST NMR study of slow conformational dynamics of small GTPases complexed with GTP and GTP analogues. *Angew. Chem. Int. Ed. Engl.* **52**, 10771–10774 (2013).
28. Bhabha, G. *et al.* A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **332**, 234–238 (2011).
29. McElheny, D., Schnell, J.R., Lansing, J.C., Dyson, H.J. & Wright, P.E. Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc. Natl. Acad. Sci. USA* **102**, 5032–5037 (2005).

30. Morcos, F. *et al.* Modeling conformational ensembles of slow functional motions in Pin1-WW. *PLoS Comput. Biol.* **6**, e1001015 (2010).
31. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114 (1997).
32. Tolman, J.R., Flanagan, J.M., Kennedy, M.A. & Prestegard, J.H. NMR evidence for slow collective motions in cyanometmyoglobin. *Nat. Struct. Biol.* **4**, 292–297 (1997).
33. Zeng, J. *et al.* High-resolution protein structure determination starting with a global fold calculated from exact solutions to the RDC equations. *J. Biomol. NMR* **45**, 265–281 (2009).
34. Clore, G.M. & Schwieters, C.D. 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 (2004).
35. Cavalli, A., Camilloni, C. & Vendruscolo, M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* **138**, 094112 (2013).
36. De Simone, A., Montalvao, R.W., Dobson, C.M. & Vendruscolo, M. Characterization of the interdomain motions in hen lysozyme using residual dipolar couplings as replica-averaged structural restraints in molecular dynamics simulations. *Biochemistry* **52**, 6480–6486 (2013).
37. Salmon, L., Bascom, G., Andricioaei, I. & Al-Hashimi, H.M. A general method for constructing atomic-resolution RNA ensembles using NMR residual dipolar couplings: the basis for interhelical motions revealed. *J. Am. Chem. Soc.* **135**, 5457–5466 (2013).
38. Emani, P.S. *et al.* Elucidating molecular motion through structural and dynamic filters of energy-minimized conformer ensembles. *J. Phys. Chem. B* **118**, 1726–1742 (2014).
39. Guerry, P. *et al.* Mapping the population of protein conformational energy sub-states from NMR dipolar couplings. *Angew. Chem. Int. Ed. Engl.* **52**, 3181–3185 (2013).
40. Fonseca, R., Pachov, D.V., Bernauer, J. & van den Bedem, H. Characterizing RNA ensembles from NMR data with kinematic models. *Nucleic Acids Res.* **42**, 9562–9572 (2014).
41. Berlin, K. *et al.* Recovering a representative conformational ensemble from underdetermined macromolecular structural data. *J. Am. Chem. Soc.* **135**, 16595–16609 (2013).
42. Lange, O.F. *et al.* Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475 (2008).
A comprehensive, integrative computational procedure reveals the conformational ensemble of ubiquition, validated by comparative crystal structure analysis.
43. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559 (1982).
44. Frederick, K.K., Marlow, M.S., Valentine, K.G. & Wand, A.J. Conformational entropy in molecular recognition by proteins. *Nature* **448**, 325–329 (2007).
This study establishes a linear relationship between conformational entropy and binding entropy for calmodulin. Ref. 45 then proposes S^2_{axis} as a proxy for binding entropy.
45. Kasinath, V., Sharp, K.A. & Wand, A.J. Microscopic insights into the NMR relaxation-based protein conformational entropy meter. *J. Am. Chem. Soc.* **135**, 15092–15100 (2013).
46. Tzeng, S.-R. & Kalodimos, C.G. Protein activity regulation by conformational entropy. *Nature* **488**, 236–240 (2012).
47. Diehl, C. *et al.* Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J. Am. Chem. Soc.* **132**, 14577–14589 (2010).
48. Best, R.B. & Vendruscolo, M. Determination of protein structures consistent with NMR order parameters. *J. Am. Chem. Soc.* **126**, 8090–8091 (2004).
49. Maragakis, P. *et al.* Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. *J. Phys. Chem. B* **112**, 6155–6158 (2008).
50. Showalter, S.A., Johnson, E., Rance, M. & Brüschweiler, R. Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. *J. Am. Chem. Soc.* **129**, 14146–14147 (2007).
51. Scouras, A.D. & Daggett, V. The Dynameomics rotamer library: amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. *Protein Sci.* **20**, 341–352 (2011).
52. Lindorff-Larsen, K., Best, R.B., Depristo, M.A., Dobson, C.M. & Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128–132 (2005).
53. Shehu, A., Clementi, C. & Kavraki, L.E. Modeling protein conformational ensembles: from missing loops to equilibrium fluctuations. *Proteins* **65**, 164–179 (2006).
54. Shehu, A., Kavraki, L.E. & Clementi, C. On the characterization of protein native state ensembles. *Biophys. J.* **92**, 1503–1511 (2007).
55. Best, R.B., Lindorff-Larsen, K., DePristo, M.A. & Vendruscolo, M. Relation between native ensembles and experimental structures of proteins. *Proc. Natl. Acad. Sci. USA* **103**, 10901–10906 (2006).
56. Cooper, A. & Dryden, D.T.F. Allostery without conformational change. A plausible model. *Eur. Biophys. J.* **11**, 103–109 (1984).
A landmark study proposing that entropic changes can drive binding events.
57. Motlagh, H.N., Wrabl, J.O., Li, J. & Hilser, V.J. The ensemble nature of allostery. *Nature* **508**, 331–339 (2014).
58. Castellani, F. *et al.* Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, 98–102 (2002).
59. Giraud, N., Bo, A., Lesage, A. & Blackledge, M. Site-specific backbone dynamics from a crystalline protein by solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **126**, 11422–11423 (2004).
60. Zinkevich, T., Chevelkov, V., Reif, B., Saalwächter, K. & Krushelnitsky, A. Internal protein dynamics on ps to μs timescales as studied by multi-frequency ^{15}N solid-state NMR relaxation. *J. Biomol. NMR* **57**, 219–235 (2013).
61. Schanda, P., Huber, M., Boisbouvier, J., Meier, B.H. & Ernst, M. Solid-state NMR measurements of asymmetric dipolar couplings provide insight into protein side-chain motion. *Angew. Chem. Int. Ed. Engl.* **50**, 11005–11009 (2011).
62. Haller, J.D. & Schanda, P. Amplitudes and time scales of picosecond-to-microsecond motion in proteins studied by solid-state NMR: a critical evaluation of experimental approaches and application to crystalline ubiquitin. *J. Biomol. NMR* **57**, 263–280 (2013).
63. Agarwal, V., Xue, Y., Reif, B. & Skrynnikov, N.R. Protein side-chain dynamics as observed by solution- and solid-state NMR spectroscopy: a similarity revealed. *J. Am. Chem. Soc.* **130**, 16611–16621 (2008).
64. Tollinger, M., Sivertsen, A.C., Meier, B.H., Ernst, M. & Schanda, P. Site-resolved measurement of microsecond-to-millisecond conformational-exchange processes in proteins by solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **134**, 14800–14807 (2012).
65. Sidhu, A., Surolia, A., Robertson, A.D. & Sundd, M. A hydrogen bond regulates slow motions in ubiquitin by modulating a β-turn flip. *J. Mol. Biol.* **411**, 1037–1048 (2011).
66. Salvi, N., Ulzega, S., Ferrage, F. & Bodenhausen, G. Time scales of slow motions in ubiquitin explored by heteronuclear double resonance. *J. Am. Chem. Soc.* **134**, 2481–2484 (2012).
67. Vijay-Kumar, S., Bugg, C.E. & Cook, W.J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194**, 531–544 (1987).
68. Keedy, D.A. *et al.* Crystal cryocooling distorts conformational heterogeneity in a model Michaelis complex of DHFR. *Structure* **22**, 899–910 (2014).
69. Rejto, P.A. & Freer, S.T. Protein conformational substates from X-ray crystallography. *Prog. Biophys. Mol. Biol.* **66**, 167–196 (1996).
70. Wilson, M.A. & Brunger, A.T. The 1.0 Å crystal structure of Ca^{2+}-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. *J. Mol. Biol.* **301**, 1237–1256 (2000).
71. Willis, B.T.M. & Pryor, A.W. *Thermal Vibrations in Crystallography* (Cambridge University Press, 1975).
72. Poon, B.K. *et al.* Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42-Å crystallographic resolution. *Proc. Natl. Acad. Sci. USA* **104**, 7869–7874 (2007).
73. Schröder, G.F., Levitt, M. & Brunger, A.T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
74. Brunger, A.T. *et al.* Improving the accuracy of macromolecular structure refinement at 7Å resolution. *Structure* **20**, 957–966 (2012).
75. Schiffer, C.A., Huber, R., Wüthrich, K. & van Gunsteren, W.F. Simultaneous refinement of the structure of BPTI against NMR data measured in solution and X-ray diffraction data measured in single crystals. *J. Mol. Biol.* **241**, 588–599 (1994).
76. Rinaldelli, M. *et al.* Simultaneous use of solution NMR and X-ray data in REFMAC5 for joint refinement/detection of structural differences. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 958–967 (2014).

77. Kuzmanic, A., Pannu, N.S. & Zagrovic, B. X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals. *Nat. Commun.* **5**, 3220 (2014).
78. Kuriyan, J. *et al.* Exploration of disorder in protein structures by X-ray restrained molecular dynamics. *Proteins* **10**, 340–358 (1991).
79. Burling, F.T. & Brünger, A.T. Thermal motion and conformational disorder in protein crystal structures: comparison of multi-conformer and time-averaging models. *Isr. J. Chem.* **34**, 165–175 (1994).
80. Levin, E.J., Kondrashov, D.A., Wesenberg, G.E. & Phillips, G.N. Jr. Ensemble refinement of protein crystal structures. *Structure* **15**, 1040–1052 (2007).
81. Rader, S.D. & Agard, D.A. Conformational substates in enzyme mechanism: the 120 K structure of α-lytic protease at 1.5 Å resolution. *Protein Sci.* **6**, 1375–1386 (1997).
82. Gros, P., van Gunsteren, W.F. & Hol, W.G. Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics. *Science* **249**, 1149–1152 (1990).
83. Clarage, J.B. & Phillips, G.N. Cross-validation tests of time-averaged molecular dynamics refinements for determination of protein structures by X-ray crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 24–36 (1994).
84. Burnley, B.T., Afonine, P.V., Adams, P.D. & Gros, P. Modelling dynamics in protein crystal structures by ensemble refinement. *eLife* **1**, e00311 (2012).
85. van den Bedem, H., Dhanik, A., Latombe, J.-C. & Deacon, A.M. Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1107–1117 (2009).
86. Terwilliger, T.C. *et al.* Interpretation of ensembles created by multiple iterative rebuilding of macromolecular models. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 597–610 (2007).
87. Brünger, A.T. Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
88. Stenkamp, R.E. & Jensen, L.H. Resolution revisited: limit of detail in electron density maps. *Acta Crystallogr. A* **40**, 251–254 (1984).
89. Fenwick, R.B., van den Bedem, H., Fraser, J.S. & Wright, P.E. Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. *Proc. Natl. Acad. Sci. USA* **111**, E445–E454 (2014).
This study establishes that a multiconformer crystallographic ensemble (ref. 85) at room temperature accurately reflects fast dynamics in solution by relating *B* factors to NMR order parameters.
90. Brüschweiler, R. & Wright, P.E. NMR order parameters of biomolecules: a new analytical representation and application to the Gaussian axial fluctuation model. *J. Am. Chem. Soc.* **116**, 8426–8427 (1994).
91. Clore, G.M. & Schwieters, C.D. Concordance of residual dipolar couplings, backbone order parameters and crystallographic *B*-factors for a small α/β protein: a unified picture of high probability, fast atomic motions in proteins. *J. Mol. Biol.* **355**, 879–886 (2006).
92. Schneider, R. *et al.* Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol. Biosyst.* **8**, 58–68 (2012).
93. Krzeminski, M., Marsh, J.A., Neale, C., Choy, W.-Y. & Forman-Kay, J.D. Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* **29**, 398–399 (2013).
94. Yang, S., Salmon, L. & Al-Hashimi, H.M. Measuring similarity between dynamic ensembles of biomolecules. *Nat. Methods* **11**, 552–554 (2014).
95. Chapman, H.N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 (2011).
96. Hirata, K. *et al.* Determination of damage-free crystal structure of an X-ray–sensitive protein using an XFEL. *Nat. Methods* **11**, 734–736 (2014).
97. Acbas, G., Niessen, K.A., Snell, E.H. & Markelz, A.G. Optical measurements of long-range protein vibrations. *Nat. Commun.* **5**, 3076 (2014).
98. Wall, M.E., Adams, P.D., Fraser, J.S. & Sauter, N.K. Diffuse X-ray scattering to model protein motions. *Structure* **22**, 182–184 (2014).
99. Schmidt, M. *et al.* Protein kinetics: structures of intermediates and reaction mechanism from time-resolved X-ray data. *Proc. Natl. Acad. Sci. USA* **101**, 4799–4804 (2004).
100. Bowman, G.R., Huang, X. & Pande, V.S. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* **49**, 197–201 (2009).
101. Bouvignies, G. *et al.* Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *Proc. Natl. Acad. Sci. USA* **102**, 13885–13890 (2005).
102. Tollinger, M., Skrynnikov, N.R., Mulder, F.A., Forman-Kay, J.D. & Kay, L.E. Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123**, 11341–11352 (2001).
103. Lang, P.T., Holton, J.M., Fraser, J.S. & Alber, T. Protein structural ensembles are revealed by redefining X-ray electron density noise. *Proc. Natl. Acad. Sci. USA* **111**, 237–242 (2014).
104. Eriksson, A.E. *et al.* Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**, 178–183 (1992).
105. Mulder, F.A.A., Mittermaier, A., Hon, B., Dahlquist, F.W. & Kay, L.E. Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.* **8**, 932–935 (2001).
106. Bouvignies, G. *et al.* Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–114 (2011).
107. Sekhar, A. & Kay, L.E. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl. Acad. Sci. USA* **110**, 12867–12874 (2013).
108. Aboul-ela, F., Karn, J. & Varani, G. Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res.* **24**, 3974–3981 (1996).
109. Kulinski, T. *et al.* The apical loop of the HIV-1 TAR RNA hairpin is stabilized by a cross-loop base pair. *J. Biol. Chem.* **278**, 38892–38901 (2003).
110. Dethoff, E.A., Petzold, K., Chugh, J., Casiano-Negroni, A. & Al-Hashimi, H.M. Visualizing transient low-populated structures of RNA. *Nature* **491**, 724–728 (2012).
111. Serrano, P. *et al.* Comparison of NMR and crystal structures highlights conformational isomerism in protein active sites. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 1393–1405 (2010).
112. Klock, H.E. *et al.* Crystal structure of a conserved hypothetical protein (gi: 13879369) from Mouse at 1.90 Å resolution reveals a new fold. *Proteins* **61**, 1132–1136 (2005).