

Automated identification of functional dynamic contact networks from X-ray crystallography

Henry van den Bedem¹, Gira Bhabha^{2,3}, Kun Yang⁴, Peter E Wright³ & James S Fraser⁵

Protein function often depends on the exchange between conformational substates. Allosteric ligand binding or distal mutations can stabilize specific active-site conformations and consequently alter protein function. Observing alternative conformations at low levels of electron density, in addition to comparison of independently determined X-ray crystal structures, can provide mechanistic insights into conformational dynamics. Here we report a new algorithm, CONTACT, that identifies contact networks of conformationally heterogeneous residues directly from high-resolution X-ray crystallography data. Contact networks determined for *Escherichia coli* dihydrofolate reductase (ecDHFR) predict the observed long-range pattern of NMR chemical shift perturbations of an allosteric mutation. A comparison of contact networks in wild-type and mutant ecDHFR suggests that mutations that alter optimized contact networks of coordinated motions can impair catalytic function. CONTACT-guided mutagenesis can exploit the structure-dynamics-function relationship in protein engineering and design.

Proteins fluctuate between alternative conformations to mediate their biological functions^{1,2}. Perturbations that affect the relative populations of conformations caused by ligand binding³, mutation⁴, post-translational modification⁵ and temperature⁶ can affect biological mechanisms. In addition to causing global or local unfolding⁷, perturbations can affect the structure and dynamics of groups of residues within the folded ensemble⁸. Proteins can evolve to exploit these perturbations for regulation: binding of ligands or the introduction of mutations at allosteric sites can stabilize transiently populated, but functional, alternative conformations at distant active sites⁹. However, these low-population states are difficult to identify with most biophysical techniques¹⁰, making it challenging to characterize how structural fluctuations regulate protein activity by changing the populations of different conformers.

Traditionally, X-ray crystallography has provided a single static model of a ‘ground state’, which is assumed to represent the lowest-energy conformation in the crystal lattice. More ‘dynamic’

interpretations of crystallographic data have historically involved modeling an ensemble of independent conformations^{11–17}. Recent developments in time-averaged refinement represent the crystallographic data with an ensemble of tens to hundreds of structures¹⁸. However, the relationship between individual ensemble members and the representation of conformational heterogeneity in the crystal remains a subject of debate¹⁵. In contrast to an ensemble model, a multiconformer model introduces alternate conformations for a residue only if they are supported by the data (**Supplementary Fig. 1**). We previously developed qFit¹⁹, an algorithm used to generate multiconformer models to represent conformational heterogeneity from high-resolution X-ray diffraction data at levels of electron density (below 1σ , where $\sigma = \text{r.m.s. deviation}$) that are commonly ignored by manual model-building efforts and that are inadequately represented by harmonic atomic displacement parameters^{19,20}.

Multiconformer and ensemble analyses of X-ray data can provide information complementary to NMR^{8,21,22}, simulations²³ and coevolutionary analyses²⁴ to reveal how interactions between distant sites enable proteins to respond to perturbations. However, tools to uncover and interpret conformational diversity from X-ray crystallography data and connect it to structural mechanisms for biomolecular dynamics remain underdeveloped. We created an algorithm, contact networks through alternate conformation transitions (CONTACT), that automatically identifies contact networks of conformationally heterogeneous residues—also called ‘dynamic close packing’¹⁷—from X-ray data. Our method can connect functional sites, propagate chemical shift differences and reveal the structural mechanisms of mutations that affect redistributions of the conformational ensemble. CONTACT networks are determined from experimental X-ray data, and our approach complements simulation-based methods such as analyzing molecular dynamics trajectories²⁵, Monte Carlo simulations²⁶ or Rosetta-based analyses²⁷.

RESULTS

Identifying conformational coupling between residues

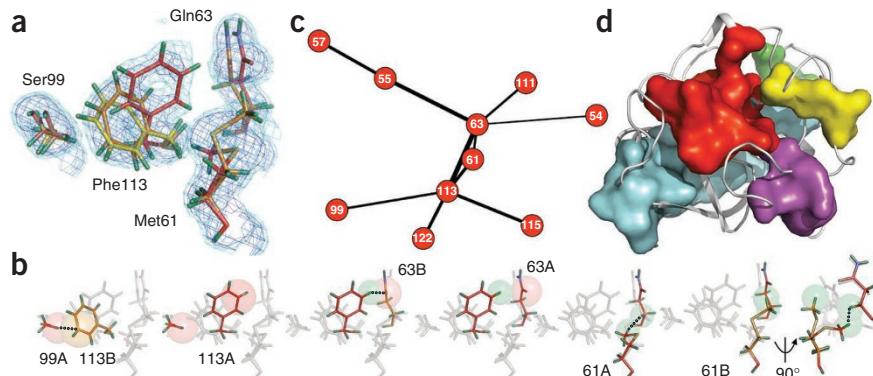
To identify interacting residues that can respond collectively to perturbations, we used our robotics-inspired algorithm

¹Joint Center for Structural Genomics, Stanford Synchrotron Radiation Lightsource, Stanford, California, USA. ²Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California, USA. ³Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California, USA. ⁴Institute for Computational and Mathematical Engineering, Stanford University, Stanford, California, USA. ⁵Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California, USA. Correspondence should be addressed to H.v.d.B. (vdbedem@slac.stanford.edu) or J.S.F. (james.fraser@ucsf.edu).

RECEIVED 8 FEBRUARY; ACCEPTED 28 JUNE; PUBLISHED ONLINE 4 AUGUST 2013; DOI:10.1038/NMETH.2592

Figure 1 | Mechanisms for conformational exchange in cyclophilin A. **(a)** X-ray electron density map contoured at 1σ (blue mesh) and 0.3σ (cyan mesh) of CYPA fit with discrete alternative conformations using qFit. Alternative conformations are colored red, orange or yellow, with hydrogen atoms added in green. **(b)** Visualizing a pathway in CYPA: atoms involved in clashes are shown in spheres scaled to van der Waals radii, and clashes between atoms are highlighted by dotted lines. This pathway originates with the OG atom of Ser99 conformation A (99A) and the CE1 atom of Phe113 conformation B (113B), which clash to 0.8 of their summed van der Waals radii.

The pathway progresses from Phe113 to Gln63, and after the movement of Met61 to conformation B introduces no new clashes, the pathway is terminated. A 90° rotation of the final panel is shown to highlight how the final move of Met61 relieves the clash with Gln63. **(c)** Networks identified by CONTACT are displayed as nodes connected by edges representing contacts that clash and are relieved by alternative conformations. The node number represents the sequence number of the residue. Line thickness between a pair of nodes represents the number of pathways that the corresponding residues are part of. The pathway in **b** forms part of the red contact network in CYPA. **(d)** The six contact networks comprising 29% of residues are mapped on the three-dimensional structure of CYPA.



qFit¹⁹ to optimally fit up to four alternative conformations per residue into electron density features, derived from experimental X-ray data, that are consistent with anharmonic disorder (**Fig. 1a**). CONTACT analyzes the repulsive van der Waals interactions across all alternative conformations in the qFit multiconformer model (**Supplementary Fig. 2**). The goal of this analysis is to define contact networks of conformationally coupled residues, in which movement between alternative conformations of one residue likely influences the conformations of all other residues in the contact network. Software for running CONTACT is available at <http://smb.slac.stanford.edu/CONTACT> and as **Supplementary Software**.

CONTACT first identifies ‘pathways’ of van der Waals overlaps. Each residue (for example, residue *i*) in turn is moved to an alternative conformation, and overlaps of van der Waals radii with any neighboring residues (for example, residue *j*) are identified (**Fig. 1b**). If neighboring residue *j* can be moved to an alternative conformation to reduce the steric overlap (or ‘clashes’) with residue *i*, the pathway is continued to neighboring residues of residue *j* (for example, residue *k*, *l* and so on) until no new clashes are introduced. The relative frequency with which residues can or cannot reduce steric overlap is also recorded (**Supplementary Fig. 3**).

Pathways can include overlapping or nearly overlapping sets of residues using different combinations of alternative conformations. Thus, pathways that share common members indicate conformational coupling even if the residues are not directly linked in a single pathway. Any pathways that share residues are grouped into a single contact network (**Fig. 1c**). A pathway is therefore a single sequence of residues that can be moved between alternative conformations such that van der Waals overlaps are reduced after each move. A contact network is a set of residues that are linked by common pathways.

CONTACT identifies multiple pathways in cyclophilin A

To test whether CONTACT can automatically identify contact networks of residues from experimental X-ray data, we first examined the human proline isomerase cyclophilin A (CYPA). Previous NMR experiments have demonstrated that CYPA undergoes conformational exchange both during its catalytic cycle and

in the absence of substrate²⁸. Room-temperature X-ray electron density maps revealed extensive alternative side-chain conformations for residues extending from the active site into the core, providing a structural rationale for the collective motion inferred by NMR⁴.

A large contact network of nine residues connects the hydrophobic core of the protein to the active site (**Fig. 1d**). This contact network (‘red’, as depicted in **Fig. 1**) generally agrees with findings obtained from a visual analysis of room-temperature X-ray data in providing a structural mechanism for collective conformational exchange detectable by NMR relaxation dispersion experiments²⁸. Consistent with the idea that perturbations in the contact network will stabilize alternative conformations of other residues in the contact network, a mutation (S99T) of a core residue leads to chemical shift changes that spread across the red contact network and results in a second-order rate constant (k_{cat}/K_m) that is $\sim 0.3\%$ that of wild type⁴.

Notably, the red contact network consists of many independent pathways that connect contact-network residues as reflected in the weights of the edges of the network graph (**Fig. 1c**). The large number of pathways suggests that the transition between the major and minor forms of the enzyme can occur by multiple structurally distinct mechanisms. The idea of multiple transition paths in CYPA is supported by recent NMR studies that incorporate evidence from the conformational dynamics of several mutants²⁹. The contributions of these transitions are difficult to separate in the collective exchange fitting procedures of the NMR experiment²⁹. Major- and minor-form end states distinguished by alternative side-chain conformations may dominate the conformational dynamics in the crystal⁴ and solution²⁸; however, CONTACT identifies multiple plausible transition paths.

Analysis of long-range perturbations in G121V ecDHFR

Dihydrofolate reductase (DHFR) is a model system for studying the relationship between conformational dynamics and catalytic activity³⁰. The solution-state dynamics of the ecDHFR are dominated by the interconversion of the Met20 loop between closed and occluded conformations, which allows optimal substrate flux through the catalytic cycle³¹. We obtained X-ray diffraction data

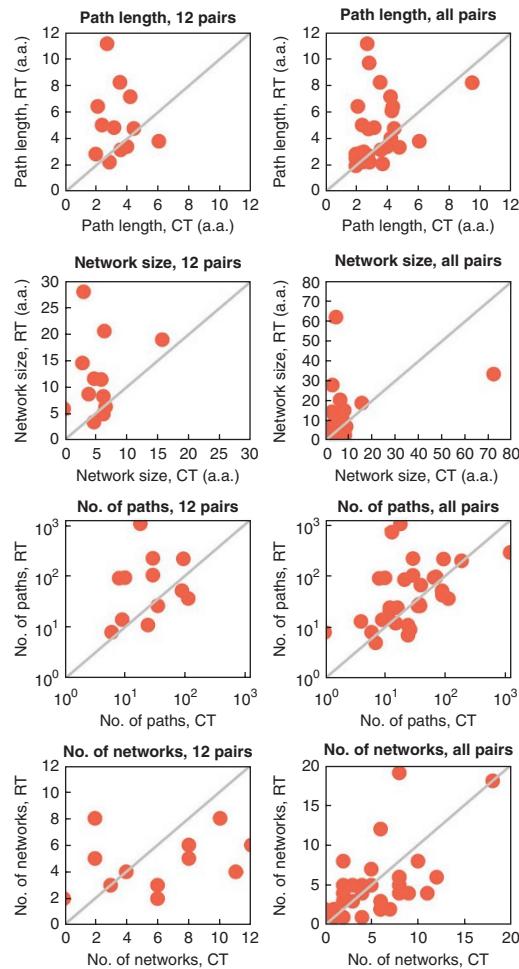
Figure 2 | Characteristics of pathways and contact networks are sensitive to temperature. Changes to average pathway lengths, contact-network sizes, number of pathways and number of contact networks across 12 closely matched high-resolution (left) or all matched (right) room-temperature (RT) and cryogenic (CT) data sets are shown. Each data point represents paired data sets, with values corresponding to room temperature along the vertical axis and cryogenic temperature along the horizontal axis. Data points are expected to lie along a 45° line if there are no differences between the room and cryogenic temperature pairs. a.a., amino acids.

sets at cryogenic (100 K; 1.15-Å resolution) and room temperature (~273 K; 1.35-Å resolution) for the model Michaelis complex E:NADP⁺:FOL (**Supplementary Table 1**). Consistent with the trends observed in a larger data set of 35 proteins taken from the Protein Data Bank (**Fig. 2, Supplementary Table 2, Supplementary Figs. 4–6 and Supplementary Note**), the room-temperature data set exhibits both more (157 versus 70) and longer (5.5 versus 4.3) pathways than the cryogenic data set.

The largest contact network from the room-temperature data set connects the functionally important FG loop to the NADP-binding pocket and the adenosine-binding domain (**Fig. 3**). This long-range connection is mediated by the cofactor NADP molecule, which uniquely connects the two subdomains (**Fig. 3a** and **Supplementary Table 3**). The electron density map is consistent with discrete disorder around the cofactor (**Fig. 4a**), providing further evidence that the bound NADP molecule is a dynamic hub. To test this model of conformational coupling of the FG loop to the adenosine-binding domain, we examined the chemical shift perturbations of a mutation in the FG loop (G121V). The NMR data for G121V is consistent with this hypothesis³². Changes in ¹⁵N and/or ¹H chemical shifts between wild-type and the G121V mutant E:NADP⁺:FOL complexes propagate from the FG loop to the adenosine-binding domain (**Fig. 4b**). However, because no G121V X-ray structure has been published, the structural origin of these long-range effects is unclear.

Modeling a valine side chain at position 121 causes severe clashes with residues 13–15 and directly impinges upon the red ecDHFR contact network (**Fig. 4c**). The pattern of contacts in the network includes FG-loop residue 125 and extends to the adenosine-binding domain, resembling the chemical shift perturbations between wild-type and G121V ecDHFR. Long-range conformational coupling generally corresponds to chemical shift propagation, whereas local effects due to increased flexibility of the FG loop in the mutant complex are absent. These results support a physical model underlying the long-range chemical shift perturbations: the G121V mutation selectively stabilizes pre-existing conformations that propagate from the FG loop to the adenosine-binding domain. Although the distance encompassed by the chemical shift changes is unexpectedly large, a single contact network mediates the direction and extent of long-range conformational coupling. In addition to these long-range effects, the allosteric mutation G121V destabilizes the catalytically competent closed conformation of the enzyme, stabilizes the occluded conformation and reduces k_{cat} to approximately ~0.6% that of wild type³³.

Several additional lines of evidence support this model. In molecular dynamics simulations, the dynamics of residue 121 are correlated with the dynamics of residues in the adenosine-binding domain³⁴. Sampling of locally unfolded states identified a similar



pattern of energetic coupling⁷. NMR spin relaxation experiments in the fast (picosecond-to-nanosecond) timescale found that the majority of residues with fast dynamics affected by G121V mutation are located in the adenosine-binding domain^{32,35}.

On the basis of our contact network analysis, we predicted that removing NADP would disrupt the conformational coupling of the FG loop to the adenosine-binding domain. To assess this idea, we compared ¹⁵N and ¹H chemical shifts for the binary wild-type and G121V E:FOL complexes. In the absence of NADP, we observed no major chemical shift perturbations for residues in the adenosine-binding domain in the folate binary complexes (**Fig. 4d** and data not shown). Thus, CONTACT provides testable hypotheses about how mutations shift conformational ensembles and induce long-range chemical shift perturbations observed in NMR experiments.

Altered conformational heterogeneity in N23PP/S148A ecDHFR

We used CONTACT to study a double N23PP and S148A mutant (N23PP/S148A) of ecDHFR, which was designed to destabilize the occluded conformation of the Met20 loop. This mutant populates a nearly wild-type structure as defined by a single cryogenic X-ray model³⁶. Notably, the N23PP/S148A mutant has a reduced rate of hydride transfer (k_{hyd}). In contrast to the wild-type enzyme, the N23PP/S148A mutant displays no evidence of conformational exchange for most active-site residues on the millisecond timescale. However, faster (picosecond-to-nanosecond)-timescale

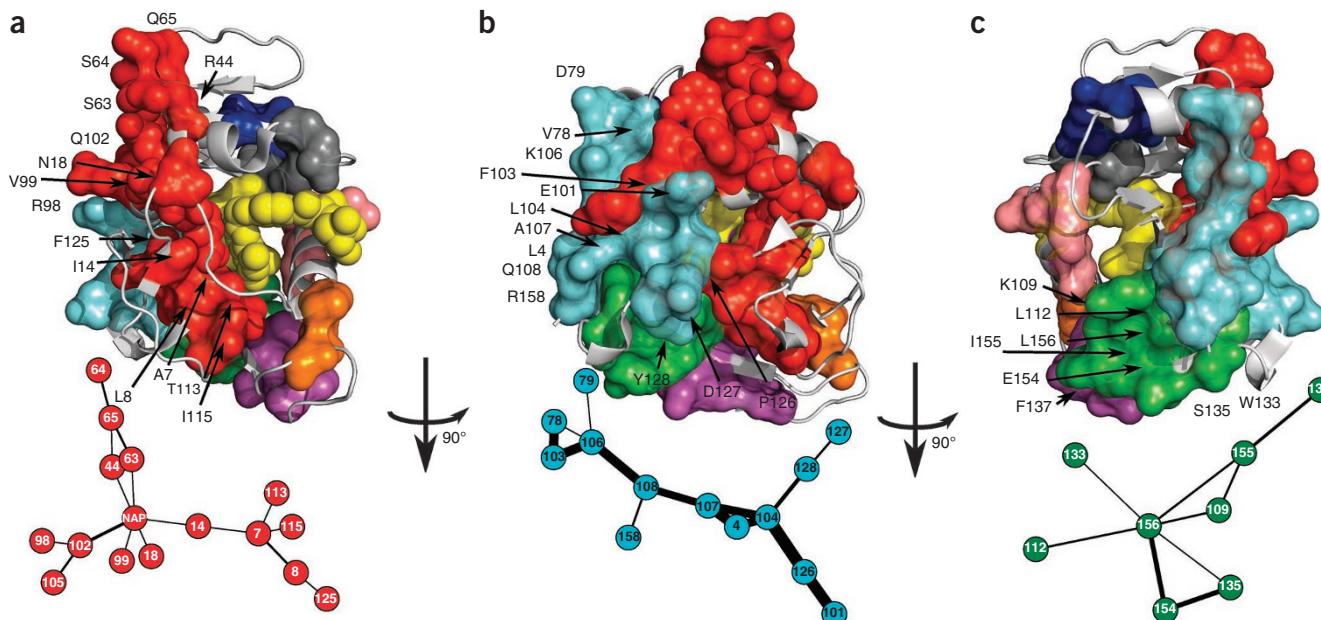


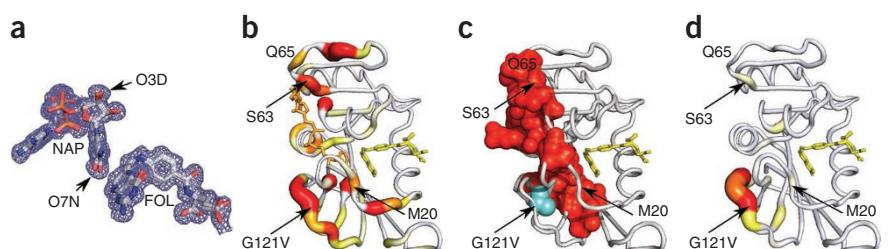
Figure 3 | All-atom contact networks in ecDHFR. **(a–c)** Contact networks are displayed in surface rendering on the crystal structure of the room-temperature E:NADP⁺:FOL complex (Protein Data Bank: 3QL3) above graphs of the largest contact networks. Nine all-atom contact networks comprise 47% of the residues in the room-temperature model of ecDHFR. **(a)** The NADP⁺ cofactor is part of the red contact network, shown in red spheres. Folate, part of the yellow contact network, is shown in yellow spheres. The red network connects residue Phe125 in the FG loop to the adenosine-binding domain (Ser63–Gln65) exclusively through the NADP cofactor. Residues in the cyan **(b)** and green **(c)** contact networks broadly agree with those identified undergoing collective exchanges in Carr-Purcell-Meiboom-Gill relaxation dispersion experiments in a process that is distinct from the conformational exchange observed near the active site. Consistent with NMR data, the cyan and green contact networks do not contact active-site residues. The salmon and blue contact networks **(c)** are implicated in hinge motions. The orange contact network is implicated in changing hydrogen bonding patterns during the closed-to-occluded transition of the Met20 loop. The yellow contact network links active site residues Ile5 and Ile95 to the folate.

backbone dynamics remain similar to those of wild type³⁶. Although previous experiments indicated that the alteration of millisecond conformational dynamics in the mutant influenced the chemical step of catalysis, the underlying mechanisms were elusive. To further investigate the origins of the reduction in hydride transfer rate, we crystallized the N23PP/S148A mutant E:NADP⁺:FOL complex in the same crystal form as wild-type ecDHFR and collected a new room-temperature data set to 1.38-Å resolution (Fig. 5, Supplementary Fig. 7 and Supplementary Tables 1 and 4).

Given the minimal structural effects of the mutation and the loss of observable millisecond conformational dynamics, we expected that CONTACT would reveal a large reduction in pathways and a lower number of residues participating in active-site

contact networks. However, we found an ~500% increase in the number of all-atom pathways in N23PP/S148A (806) compared to in wild type (157). In both wild-type and N23PP/S148A ecDHFR, a large contact network connects the FG loop to the adenosine-binding domain through the NADP cofactor. However, several residues of the Met20 loop were included in the mutant, but not wild-type, contact network (Fig. 5a). This result suggests that additional nonproductive motions surrounding the active site of N23PP/S148A ecDHFR can influence the relative positions of the NADP and FOL during the reaction cycle. A second large contact network ('cyan') in the N23PP/S148A mutant reveals an extensive set of connections across the central β-strand of the protein (Fig. 5a,b). Whereas in wild-type ecDHFR the connections across these β-strands are distributed over several contact

Figure 4 | Contact-network analysis and an allosteric mutant (G121V) of ecDHFR. **(a)** $2mF_0 - DF_c$ electron density map around the cofactor and substrate of the room-temperature E:NADP⁺:FOL complex contoured at 0.3σ . Asymmetric density profiles around oxygen atoms (O7N, O3D) of the NADP molecule support multiple conformations. **(b)** Color (yellow, small, to red, large) and thickness of the backbone tube represents the magnitude of the weighted chemical shift differences obtained from wild-type (WT) and a G121V mutant E:NADP⁺:FOL complex from 0.1–1.0 p.p.m. Residues Ser63, Gln65, Gly67 and Thr68 in the adenosine-binding domain exhibit large chemical shift changes despite their location over 23 Å away from the mutation site. **(c)** Red contact network obtained from the room-temperature WT E:NADP⁺:FOL complex (also in Fig. 3a) in the same orientation. For illustrative purposes, the G121V mutation (cyan) is modeled on the WT molecule, abutting the red contact network. **(d)** Chemical shift differences between binary WT and G121V mutant E:FOL complexes are localized to the site of mutation, confirming the central role of the NADP cofactor in coupling distant sites in ecDHFR.



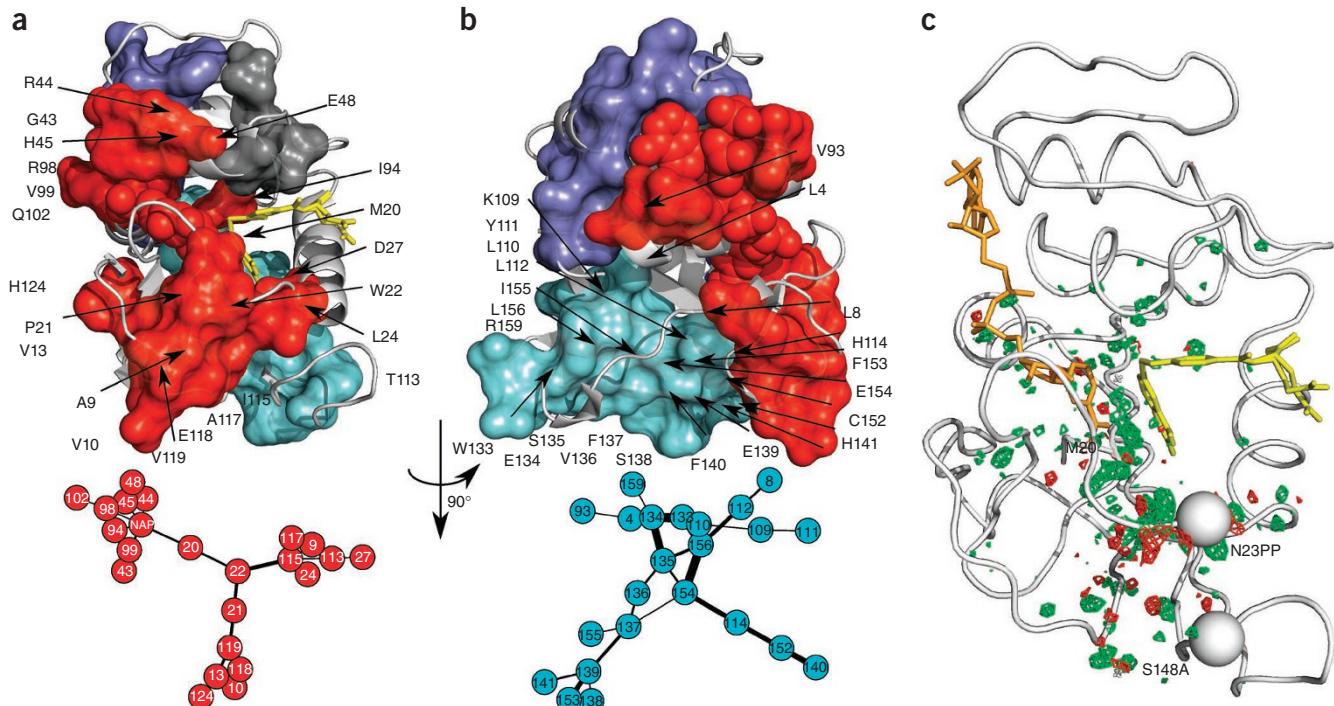


Figure 5 | Increased conformational heterogeneity at the active site of the crystal structure of the E:NADP⁺:FOL complex of the N23PP/S148A mutant. **(a)** The red contact network shares similarities with its wild-type (WT) counterpart, but several residues in the Met20 loop now participate in this contact network. The connection between the FG loop and NAP to the adenosine-binding domain through residue Ile14 has been replaced by Met20 loop residues Met20, Pro21 and Trp22. **(b)** The cyan contact network is composed largely of three β-strands toward the C terminus of the molecule. **(c)** An isomorphous difference electron density map of the WT and N23PP/S148A data sets phased with the WT crystal structure. Positive difference density (4.0σ; green) in the absence of negative difference density (-4.0σ; red) indicates elevated heterogeneity. Mutation sites are shown with spheres: NADP⁺ in orange and folate in yellow (see also **Supplementary Fig. 7b**).

networks (**Fig. 4b,c**), the entire β-sheet forms a single contact network in N23PP/S148A ecDHFR. These results suggest that the mutant enzyme exhibits increased active-site heterogeneity despite the loss of detectable conformational averaging on the millisecond timescale.

To complement the results of CONTACT, we generated an isomorphous difference map between the wild-type and N23PP/S148A data sets. This map revealed that the most prominent difference features were located immediately adjacent to the site of mutation, corresponding to the change in amino acid identity (Asn23 to Pro) or proline insertion (**Fig. 5c**). However, several other features surrounding the active site suggest that the mutations had shifted the conformational distributions of neighboring residues. The presence of positive difference density without a corresponding negative difference density peak implies increased disorder in the mutant electron density map. A single model would fail to reveal any differences between the wild-type and mutant enzyme, as the mean positions are not altered. Rather, one major effect of the mutation is to broaden the conformational distributions relative to wild type, thus increasing frustrated, nonproductive motions, involving competing low-free-energy conformations that cannot be mutually satisfied, in the mutant enzyme. Indeed, difference density around the Met20 loop residues Met20 and Trp22 revealed elevated conformational heterogeneity, and these residues participate in N23PP/S148A, but not wild-type, contact networks (**Figs. 3** and **5a**). Additional difference density is observed across the β-sheet contacting the C terminus. These results provide a further indication that the mutations can

alter dynamic properties in distant regions of the molecule and that the N23PP/S148A mutation increases conformational heterogeneity observable by X-ray crystallography.

DISCUSSION

Our analysis reveals that conformations accessible in the crystal are sensitive to perturbations such as temperature and mutations. Researchers behind recent experiments have postulated that intramolecular pathways of signaling exist within a protein structure²⁴. The physical basis of these pathways and how they are insulated from and supported by the surrounding protein structure remains unclear. Whereas previous studies have focused largely on using the functional response to mutations to delineate the boundaries of intramolecular pathways, our results suggest that changes to contact networks as a function of data collection temperature may reveal new insights. For most proteins, we observed a decrease in the number and lengths of pathways at cryogenic temperature, a result implying that the broader conformational ensembles of residues at room temperature go beyond ‘filling the voids’ that are provided by expansion of the lattice⁶. Rather, networks of residues collectively sample alternative conformational substates. Some cryo-cooling procedures stabilize contact networks into specific conformational substates from those sampled at room temperature. It is important to note that exposure to X-ray radiation can introduce serious radiation damage artifacts that may complicate these analyses. Radiation damage is dramatically reduced at cryogenic temperatures. New advances in serial femtosecond crystallography may mitigate some of these

concerns, retaining the advantages of noncryogenic collection with radiation damage protection that exceeds cryogenic techniques³⁷. Thus, new biophysical tools may help circumvent the complex trade-off between collecting data below the glass transition³⁸, which alters conformational heterogeneity, and protecting against radiation damage.

In CYPA, a large contact network consists of residues that collectively sample different conformations during the catalytic cycle. A mutation (S99T) outside of the active site reduces the extent of the contact network, impairs the kinetics of interconversion and reduces catalytic efficiency. In contrast, in ecDHFR, a mutant (N23PP/S148A) with a rigidified Met20 loop has the surprising effect of increasing nonproductive, frustrated conformational heterogeneity in the active site. Indeed, frustration is thought to play a key role in native state dynamics³⁹. The loss of detectable millisecond conformational exchange in the active site is due to a decrease in backbone flexibility of the Met20 loop. However, local anharmonic side-chain motions, which do not generate a large enough chemical shift difference or which occur on a timescale inaccessible for detection by a relaxation dispersion experiment, have likely increased around the active site and may inhibit progress toward the transition state. Further NMR experiments, particularly for side-chain methyl groups, may provide additional insights.

Superficially, the S99T mutation of CYPA and the N23PP/S148A mutation of ecDHFR are similar: both result in losses of catalytic activity and micro- to millisecond-timescale conformational dynamics. However, the CONTACT results offer distinct interpretations of these mutations. In CYPA, an overpacked core decreases the kinetics of interconversion between conformational substates and catalytic rate. In contrast, in ecDHFR, the stabilization of Met20-loop dynamics increases heterogeneity in the active site, which likely decreases the rate of hydride transfer. Enzyme engineering and design efforts could target residues in contact networks for simultaneous sequence optimization based on the principle that network residues exert a substantial influence on the conformations at other contact-network sites. As protein design methods are optimized, the conformational dynamics necessary for catalytic cycles (for example, to occlude water or prevent product inhibition) will need to be defined and engineered. Recently, *de novo* designed enzymes appear to suffer from packing imperfections that allow too much conformational heterogeneity, suggesting that contact networks can be targeted to improve packing⁴⁰.

As multiconformer and ensemble models are more widely adopted, several improvements to our method could reveal further links between conformational flexibility and mechanism. Minor improvements in R_{free} ⁴¹ should be carefully evaluated against model fit, stereochemistry and, importantly, a biologist's ability to develop structure-based hypotheses of biomolecular function. A particular challenge is to determine the extent of conformational coupling caused by steric mechanisms, as modeling of atomic interactions with a repulsive hard-sphere potential is an egregious simplification. As in ecDHFR, in which the NADP ligand connects distant sites, it is likely that nonprotein atoms play key roles in coupling distant sites; however, qFit does not currently automatically fit multiconformer small molecules or water molecules into the electron density. Despite these simplifications, CONTACT analysis of a multiconformer model from a single

X-ray experiment complements NMR²², long-timescale molecular dynamics⁴², Rosetta sampling²⁷ and comparison of multiple independent crystal structures²¹. Integrated analyses combining these approaches will help researchers to derive structural bases for conformational dynamics of proteins and to develop new hypotheses about how protein motions relate to function.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. Protein Data Bank: [4KJL](#), [4KJK](#), [4KJJ](#).

Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).

ACKNOWLEDGMENTS

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 SLAC National Accelerator Laboratory LDRD (Laboratory Directed Research and Development) grant SLAC-LDRD-0014-13-2; G.B. is supported as a Merck Fellow of the Damon Runyon Cancer Research Foundation (DRG-2136-12); K.Y. is supported by the General Wang Yaowu Stanford graduate fellowship; P.E.W. is supported by the US National Institutes of Health (GM75995); J.S.F. is supported by the US National Institutes of Health Early Independence Award (DP5OD009180). We acknowledge T. Alber, T. Kortemme, D. Keedy, D. Kern, D. Tawfik and R. Woldeyes for helpful discussions; N. Echols for advice on different treatments of hydrogen atoms in Phenix; J. Holton, J. Tanamachi and G. Meigs at Advanced Light Source Beamline 8.3.1 for support with X-ray data collection; B. Duggan and D. Boehr for chemical shift data on G121V ecDHFR complexes; and H. Axelrod and C. Trame for data collection and refinement of RT RBM39.

AUTHOR CONTRIBUTIONS

H.v.d.B., G.B., P.E.W. and J.S.F. designed and performed experiments, analyzed data and prepared the manuscript; G.B. and J.S.F. collected data; and H.v.d.B., K.Y. and J.S.F. developed analytical tools.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Fraser, J.S. & Jackson, C.J. Mining electron density for functionally relevant protein polysterism in crystal structures. *Cell. Mol. Life Sci.* **68**, 1829–1841 (2011).
- Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA* **102**, 6679–6685 (2005).
- Marlow, M.S., Dogan, J., Frederick, K.K., Valentine, K.G. & Wand, A.J. The role of conformational entropy in molecular recognition by calmodulin. *Nat. Chem. Biol.* **6**, 352–358 (2010).
- Fraser, J.S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673 (2009).
- Yu, B. *et al.* Structural and energetic mechanisms of cooperative autoinhibition and activation of Vav1. *Cell* **140**, 246–256 (2010).
- Fraser, J.S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. USA* **108**, 16247–16252 (2011).
- Whitten, S.T., García-Moreno, E.B. & Hilser, V.J. Local conformational fluctuations can modulate the coupling between proton binding and global structural transitions in proteins. *Proc. Natl. Acad. Sci. USA* **102**, 4282–4287 (2005).
- Mittermaier, A.K. & Kay, L.E. Observing biological dynamics at atomic resolution using NMR. *Trends Biochem. Sci.* **34**, 601–611 (2009).
- Goodey, N.M. & Benkovic, S.J. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* **4**, 474–482 (2008).
- Kern, D. & Zuiderweg, E.R. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **13**, 748–757 (2003).

11. Burling, F. & Brünger, A. Thermal motion and conformational disorder in protein crystal structures: comparison of multi-conformer and time-averaging models. *Isr. J. Chem.* **34**, 165–175 (1994).
12. Levin, E.J., Kondrashov, D.A., Wesenberg, G.E. & Phillips, G.N. Jr. Ensemble refinement of protein crystal structures. *Structure* **15**, 1040–1052 (2007).
13. Knight, J.L. et al. Exploring structural variability in X-ray crystallographic models using protein local optimization by torsion-angle sampling. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 383–396 (2008).
14. Kuriyan, J. et al. Exploration of disorder in protein structures by X-ray restrained molecular dynamics. *Proteins* **10**, 340–358 (1991).
15. 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).
16. Wilson, M.A. & Brunger, A.T. The 1.0 Å crystal structure of Ca²⁺-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. *J. Mol. Biol.* **301**, 1237–1256 (2000).
17. Rader, S.D. & Agard, D.A. Conformational substates in enzyme mechanism: the 120K structure of α -lytic protease at 1.5 Å resolution. *Protein Sci.* **6**, 1375–1386 (1997).
18. Burnley, B.T., Afonine, P.V., Adams, P.D. & Gros, P. Modelling dynamics in protein crystal structures by ensemble refinement. *eLife* **1**, e00311 (2012).
19. 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).
20. Lang, P.T. et al. Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci.* **19**, 1420–1431 (2010).
21. 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).
22. 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).
23. Young, M.A., Gonfloni, S., Superti-Furga, G., Roux, B. & Kuriyan, J. Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **105**, 115–126 (2001).
24. Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138**, 774–786 (2009).
25. McClendon, C.L., Friedland, G., Mobley, D.L., Amirkhani, H. & Jacobson, M.P. Quantifying correlations between allosteric sites in thermodynamic ensembles. *J. Chem. Theory Comput.* **5**, 2486–2502 (2009).
26. DuBay, K.H., Bothma, J.P. & Geissler, P.L. Long-range intra-protein communication can be transmitted by correlated side-chain fluctuations alone. *PLoS Comput. Biol.* **7**, e1002168 (2011).
27. Kidd, B.A., Baker, D. & Thomas, W.E. Computation of conformational coupling in allosteric proteins. *PLoS Comput. Biol.* **5**, e1000484 (2009).
28. Eisenmesser, E.Z. et al. Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–121 (2005).
29. Schlegel, J., Armstrong, G.S., Redzic, J.S., Zhang, F. & Eisenmesser, E.Z. Characterizing and controlling the inherent dynamics of cyclophilin-A. *Protein Sci.* **18**, 811–824 (2009).
30. Boehr, D.D., McElheny, D., Dyson, H.J. & Wright, P.E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638–1642 (2006).
31. Sawaya, M.R. & Kraut, J. Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. *Biochemistry* **36**, 586–603 (1997).
32. Boehr, D.D. et al. A distal mutation perturbs dynamic amino acid networks in dihydrofolate reductase. *Biochemistry* **52**, 4605–4619 (2013).
33. Cameron, C.E. & Benkovic, S.J. Evidence for a functional role of the dynamics of glycine-121 of *Escherichia coli* dihydrofolate reductase obtained from kinetic analysis of a site-directed mutant. *Biochemistry* **36**, 15792–15800 (1997).
34. Rod, T.H., Radkiewicz, J.L. & Brooks, C.L. Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proc. Natl. Acad. Sci. USA* **100**, 6980–6985 (2003).
35. Mauldin, R.V., Sapienza, P.J., Petit, C.M. & Lee, A.L. Structure and dynamics of the G121V dihydrofolate reductase mutant: lessons from a transition-state inhibitor complex. *PLoS ONE* **7**, e33252 (2012).
36. Bhabha, G. et al. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **332**, 234–238 (2011).
37. Boutet, S. et al. High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362–364 (2012).
38. Halle, B. Biomolecular cryocooling: structural changes during flash-cooling. *Proc. Natl. Acad. Sci. USA* **101**, 4793–4798 (2004).
39. Zhuravlev, P.I. & Papoian, G.A. Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. *Q. Rev. Biophys.* **43**, 295–332 (2010).
40. Khersonsky, O. et al. Optimization of the *in-silico*-designed kemp eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.* **407**, 391–412 (2011).
41. Brünger, A.T. Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
42. Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H. & Shaw, D.E. Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* **41**, 429–452 (2012).

ONLINE METHODS

qFit. To model heterogeneous features in the electron density, qFit computes an optimal fit of 1–4 conformations per residue together with their occupancies¹⁹. No explicit information about nonbonded contacts is included in the initial assignment of conformations or occupancy. These conformations are subsequently connected, allowing for backbone heterogeneity into a multiconformer model. The multiconformer models can provide evidence of discrete heterogeneity of populations down to ~10%. qFit is available as a web server at <http://smb.slac.stanford.edu/qFitServer/qFit.jsp>.

We used Phenix v.1.8-1069 to add hydrogens to qFit models and for subsequent conventional positional and ADP refinement steps⁴³. The qFit model is prepared for refinement with phenix.ready_set using the default parameters, which add riding hydrogens. The model is refined without manual intervention with phenix.refine using the flags “optimize_xyz_weight=true optimize_adp_weight=true main.strategy.individual_sites_real_space=false main.number_of_macro_cycles=5.”

All other phenix.refine parameters are set to the default values, including the occupancy refinement parameter. For models with diffraction to greater than 1.55-Å resolution, we also include the flag “adp.individual.anisotropic=not water or element H.”

Because phenix.refine optimizes the hydrogen placement to the center of X-ray scattering rather than the likely nuclear position, hydrogens are stripped after refinement with phenix.pdbtools using the flags “modify.remove=element H.”

Hydrogens were then restored to their nuclear positions with phenix.ready_set using the default parameters.

Model building and calculation of pathways. The CONTACT algorithm (available at <http://smb.slac.stanford.edu/CONTACT>) calculates the most severe van der Waals clash between atoms from different residues that are separated by less than the sum of the van der Waals radii. This calculation takes into account all conformations of all pairs of residues with alternative conformations, excluding main-chain hydrogen bonded atoms (identified by MMDB) and pairs of cysteine residues. In evaluating a multiconformer model, we use the most severe 30% of these van der Waals overlaps to define the threshold value for clashes (T_{stress}). The 30th percentile of van der Waal's overlap is an adjustable parameter in CONTACT. Functional biological insights may be obtained at other values of this parameter subject to considering resolution, data quality and crystal environment.

Starting at each residue in succession, if we have obtained a pathway up to residue i , a residue j is added to the pathway if the following two conditions are satisfied: (i) changing conformation u_i at residue position i to conformation r_i creates substantial ($\geq T_{\text{stress}}$) overlap E on a conformation u_j at residue position j : $E(r_i, u_j) \geq T_{\text{stress}}$ and (ii) changing conformation u_j to conformation r_j at residue position j reduces the most substantial overlap: $E(r_i, r_j) < E(u_i, r_j)$.

A pathway then continues for the pairwise interactions between j and a residue k that satisfies these conditions. This process is continued until no more van der Waals overlaps are introduced. The CONTACT algorithm can calculate pathways using either side-chain atoms only or both main-chain and side-chain atoms. The parameter T_{stress} can be varied (Supplementary Fig. 5). Although these alternative conformations need not be kinetically

or thermodynamically mutually exclusive, the overlap of van der Waals radii indicates a likely coupling between the relative populations of conformations at each site.

Calculation of contact networks for CYPA and ecDHFR. Pathways identified by CONTACT are included in the network analysis of CYPA, WT ecDHFR and N23PP/S148A ecDHFR in the main text at the worst 30th percentile of van der Waals overlap, corresponding to 13%, 14% and 14% overlaps, respectively. We furthermore required that all clashes in a pathway were reduced to 10% overlap or less. We built a network graph in which nodes were residues with edges indicating contacts identified by CONTACT. In analyzing the properties of these connected networks, we considered only subgraphs with more than three nodes and computed the edge weight as the number of distinct pathways between nodes.

Pathways and contact networks calculated from 35 pairs of cryogenic and room-temperature data sets. We used qFit to rebuild 35 pairs of protein models with available high-resolution (2.0-Å or better) X-ray diffraction data collected at room and cryogenic temperatures (Supplementary Table 3). The majority were crystallized in nearly identical or very similar conditions. The data sets have R_{merge} values indicating no unusual radiation damage⁶. When no R_{free} set was deposited or could be extracted, we chose a test set using the standard parameters in phenix.refine. Structures were rebuilt using qFit and then refined as described above.

So that we could examine the effect of severity of van der Waals overlap on pathway discovery, the parameter T_{stress} was incremented in 5% steps starting at the worst 5th percentile of van der Waals overlaps. The search for additional pathways was terminated when its number exceeded 1,000,000. The effect of considering a higher percentile is that generally more van der Waals overlaps are identified but that these are also less likely to be relieved below the threshold to continue the pathway. The number of all-atom pathways and their lengths diverge substantially between room and cryogenic temperatures, reaching a peak at the 30th–35th percentile overlap (Supplementary Fig. 5). For side chain-only pathways, the largest differences in the number of pathways between room and cryogenic temperature were observed at the 15th and 25th percentiles, but no significant difference was found for their lengths (Supplementary Fig. 2).

DHFR chemical shift perturbations. Backbone amide assignments for the E:NADP⁺:FOL⁴⁴ and E:FOL⁴⁵ complexes of ¹³C,¹⁵N-labeled or ¹³C,¹⁵N,²H-labeled WT and G121V ecDHFR were made using standard triple-resonance methods⁴⁴. Weighted average of chemical shift differences were calculated using the formula

$$\sqrt{\left(\Delta^{13}\text{C}\right)^2 + \left(\frac{\Delta^{15}\text{N}}{5}\right)^2}$$

ecDHFR X-ray crystallography. WT (E:NADP⁺:FOL) and N23PP/S148A (E:NADP⁺:FOL) were purified and concentrated to 30 mg/mL (ref. 36). Both WT and N23PP/S148A crystals were obtained by the hanging drop method by mixing protein 1:1 with well solution (100 mM HEPES, pH 7.5, 21% PEG8000, 200 mM MgCl₂). Data were collected at BL 8.3.1 of the Advanced Light Source. For room-temperature data collection, the crystal is

mounted using a loop and covered with RT tubing (MITEGEN) with 15 μ L of a 3:1 mixture of mother liquor to water in the tip. The cryojet is set to 273 K and moved such that it lies just outside the capillary tube. The beam is heavily attenuated during test shots to establish a data collection strategy. During data collection, the attenuation is reduced by removing all but the aluminum foil, and the radiation dose is spread across the large ($600 \mu\text{m} \times 200 \mu\text{m} \times 100 \mu\text{m}$) needle-shaped crystals by translation after each shot using a custom TCL script that interfaces with the standard BLUEICE interface. There are no unusual indications of radiation damage or problems with mosaic spread due to the translation: unit-cell parameters and scale factors are stable during data processing, and the overall R_{sym} is low. The cryogenic data set was collected after the addition of 10% PEG400 to the mother liquor. Data sets were processed using XIA2 (ref. 46) using the flag

“-3dii” for XDS⁴⁷ pipeline. Molecular replacement was performed using 1RX2 using Phaser⁴⁸ through the Phenix GUI⁴³.

43. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
44. Osborne, M.J., Schnell, J., Benkovic, S.J., Dyson, H.J. & Wright, P.E. Backbone dynamics in dihydrofolate reductase complexes: role of loop flexibility in the catalytic mechanism. *Biochemistry* **40**, 9846–9859 (2001).
45. Falzone, C.J. *et al.* ^1H , ^{15}N and ^{13}C resonance assignments, secondary structure, and the conformation of substrate in the binary folate complex of *Escherichia coli* dihydrofolate reductase. *J. Biomol. NMR* **4**, 349–366 (1994).
46. Winter, G., Lobley, C.M.C. & Prince, S.M. Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–1273 (2013).
47. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
48. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

ONLINE METHODS

qFit. To model heterogeneous features in the electron density, qFit computes an optimal fit of 1–4 conformations per residue together with their occupancies¹⁹. No explicit information about nonbonded contacts is included in the initial assignment of conformations or occupancy. These conformations are subsequently connected, allowing for backbone heterogeneity into a multiconformer model. The multiconformer models can provide evidence of discrete heterogeneity of populations down to ~10%. qFit is available as a web server at <http://smb.slac.stanford.edu/qFitServer/qFit.jsp>.

We used Phenix v.1.8-1069 to add hydrogens to qFit models and for subsequent conventional positional and ADP refinement steps⁴³. The qFit model is prepared for refinement with phenix.ready_set using the default parameters, which add riding hydrogens. The model is refined without manual intervention with phenix.refine using the flags “optimize_xyz_weight=true optimize_adp_weight=true main.strategy.individual_sites_real_space=false main.number_of_macro_cycles=5.”

All other phenix.refine parameters are set to the default values, including the occupancy refinement parameter. For models with diffraction to greater than 1.55-Å resolution, we also include the flag “adp.individual.anisotropic=not water or element H.”

Because phenix.refine optimizes the hydrogen placement to the center of X-ray scattering rather than the likely nuclear position, hydrogens are stripped after refinement with phenix.pdbtools using the flags “modify.remove=element H.”

Hydrogens were then restored to their nuclear positions with phenix.ready_set using the default parameters.

Model building and calculation of pathways. The CONTACT algorithm (available at <http://smb.slac.stanford.edu/CONTACT>) calculates the most severe van der Waals clash between atoms from different residues that are separated by less than the sum of the van der Waals radii. This calculation takes into account all conformations of all pairs of residues with alternative conformations, excluding main-chain hydrogen bonded atoms (identified by MMDB) and pairs of cysteine residues. In evaluating a multiconformer model, we use the most severe 30% of these van der Waals overlaps to define the threshold value for clashes (T_{stress}). The 30th percentile of van der Waal's overlap is an adjustable parameter in CONTACT. Functional biological insights may be obtained at other values of this parameter subject to considering resolution, data quality and crystal environment.

Starting at each residue in succession, if we have obtained a pathway up to residue i , a residue j is added to the pathway if the following two conditions are satisfied: (i) changing conformation u_i at residue position i to conformation r_i creates substantial ($\geq T_{\text{stress}}$) overlap E on a conformation u_j at residue position j : $E(r_i, u_j) \geq T_{\text{stress}}$ and (ii) changing conformation u_j to conformation r_j at residue position j reduces the most substantial overlap: $E(r_i, r_j) < E(u_i, r_j)$.

A pathway then continues for the pairwise interactions between j and a residue k that satisfies these conditions. This process is continued until no more van der Waals overlaps are introduced. The CONTACT algorithm can calculate pathways using either side-chain atoms only or both main-chain and side-chain atoms. The parameter T_{stress} can be varied (Supplementary Fig. 5). Although these alternative conformations need not be kinetically

or thermodynamically mutually exclusive, the overlap of van der Waals radii indicates a likely coupling between the relative populations of conformations at each site.

Calculation of contact networks for CYPA and ecDHFR. Pathways identified by CONTACT are included in the network analysis of CYPA, WT ecDHFR and N23PP/S148A ecDHFR in the main text at the worst 30th percentile of van der Waals overlap, corresponding to 13%, 14% and 14% overlaps, respectively. We furthermore required that all clashes in a pathway were reduced to 10% overlap or less. We built a network graph in which nodes were residues with edges indicating contacts identified by CONTACT. In analyzing the properties of these connected networks, we considered only subgraphs with more than three nodes and computed the edge weight as the number of distinct pathways between nodes.

Pathways and contact networks calculated from 35 pairs of cryogenic and room-temperature data sets. We used qFit to rebuild 35 pairs of protein models with available high-resolution (2.0-Å or better) X-ray diffraction data collected at room and cryogenic temperatures (Supplementary Table 3). The majority were crystallized in nearly identical or very similar conditions. The data sets have R_{merge} values indicating no unusual radiation damage⁶. When no R_{free} set was deposited or could be extracted, we chose a test set using the standard parameters in phenix.refine. Structures were rebuilt using qFit and then refined as described above.

So that we could examine the effect of severity of van der Waals overlap on pathway discovery, the parameter T_{stress} was incremented in 5% steps starting at the worst 5th percentile of van der Waals overlaps. The search for additional pathways was terminated when its number exceeded 1,000,000. The effect of considering a higher percentile is that generally more van der Waals overlaps are identified but that these are also less likely to be relieved below the threshold to continue the pathway. The number of all-atom pathways and their lengths diverge substantially between room and cryogenic temperatures, reaching a peak at the 30th–35th percentile overlap (Supplementary Fig. 5). For side chain-only pathways, the largest differences in the number of pathways between room and cryogenic temperature were observed at the 15th and 25th percentiles, but no significant difference was found for their lengths (Supplementary Fig. 2).

DHFR chemical shift perturbations. Backbone amide assignments for the E:NADP⁺:FOL⁴⁴ and E:FOL⁴⁵ complexes of ¹³C,¹⁵N-labeled or ¹³C,¹⁵N,²H-labeled WT and G121V ecDHFR were made using standard triple-resonance methods⁴⁴. Weighted average of chemical shift differences were calculated using the formula

$$\sqrt{\left(\Delta^{13}\text{C}\right)^2 + \left(\frac{\Delta^{15}\text{N}}{5}\right)^2}$$

ecDHFR X-ray crystallography. WT (E:NADP⁺:FOL) and N23PP/S148A (E:NADP⁺:FOL) were purified and concentrated to 30 mg/mL (ref. 36). Both WT and N23PP/S148A crystals were obtained by the hanging drop method by mixing protein 1:1 with well solution (100 mM HEPES, pH 7.5, 21% PEG8000, 200 mM MgCl₂). Data were collected at BL 8.3.1 of the Advanced Light Source. For room-temperature data collection, the crystal is

mounted using a loop and covered with RT tubing (MITEGEN) with 15 μ L of a 3:1 mixture of mother liquor to water in the tip. The cryojet is set to 273 K and moved such that it lies just outside the capillary tube. The beam is heavily attenuated during test shots to establish a data collection strategy. During data collection, the attenuation is reduced by removing all but the aluminum foil, and the radiation dose is spread across the large ($600 \mu\text{m} \times 200 \mu\text{m} \times 100 \mu\text{m}$) needle-shaped crystals by translation after each shot using a custom TCL script that interfaces with the standard BLUEICE interface. There are no unusual indications of radiation damage or problems with mosaic spread due to the translation: unit-cell parameters and scale factors are stable during data processing, and the overall R_{sym} is low. The cryogenic data set was collected after the addition of 10% PEG400 to the mother liquor. Data sets were processed using XIA2 (ref. 46) using the flag

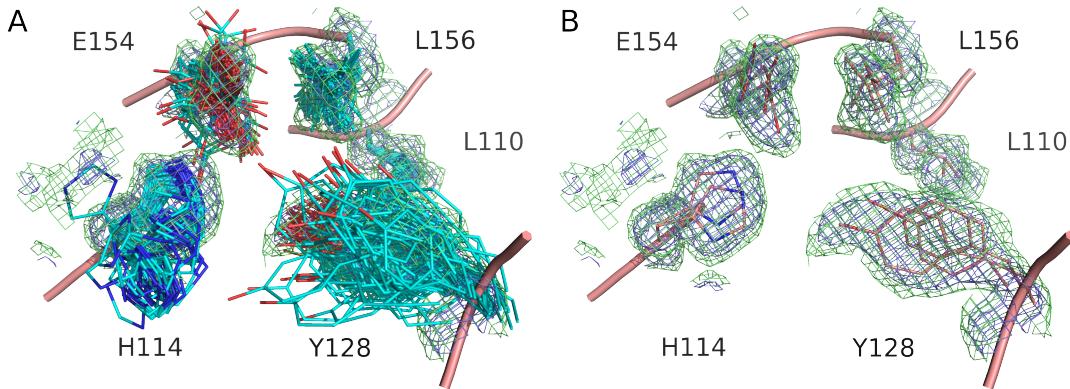
“-3dii” for XDS⁴⁷ pipeline. Molecular replacement was performed using 1RX2 using Phaser⁴⁸ through the Phenix GUI⁴³.

43. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
44. Osborne, M.J., Schnell, J., Benkovic, S.J., Dyson, H.J. & Wright, P.E. Backbone dynamics in dihydrofolate reductase complexes: role of loop flexibility in the catalytic mechanism. *Biochemistry* **40**, 9846–9859 (2001).
45. Falzone, C.J. *et al.* ^1H , ^{15}N and ^{13}C resonance assignments, secondary structure, and the conformation of substrate in the binary folate complex of *Escherichia coli* dihydrofolate reductase. *J. Biomol. NMR* **4**, 349–366 (1994).
46. Winter, G., Lobley, C.M.C. & Prince, S.M. Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–1273 (2013).
47. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
48. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

Automated identification of functional dynamic contact networks from X-ray crystallography

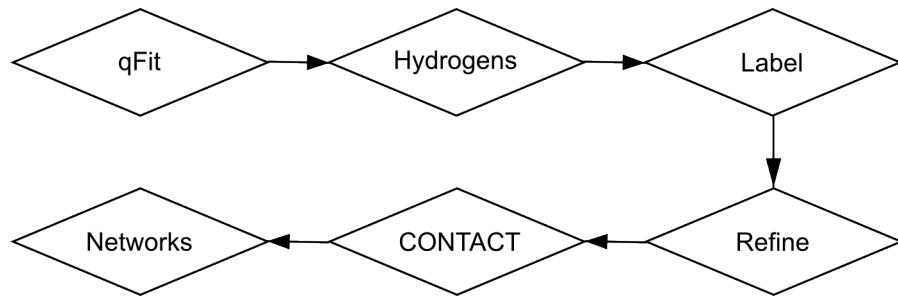
Henry van den Bedem, Gira Bhabha, Kun Yang, Peter E. Wright and James S. Fraser

Supplementary Figure 1



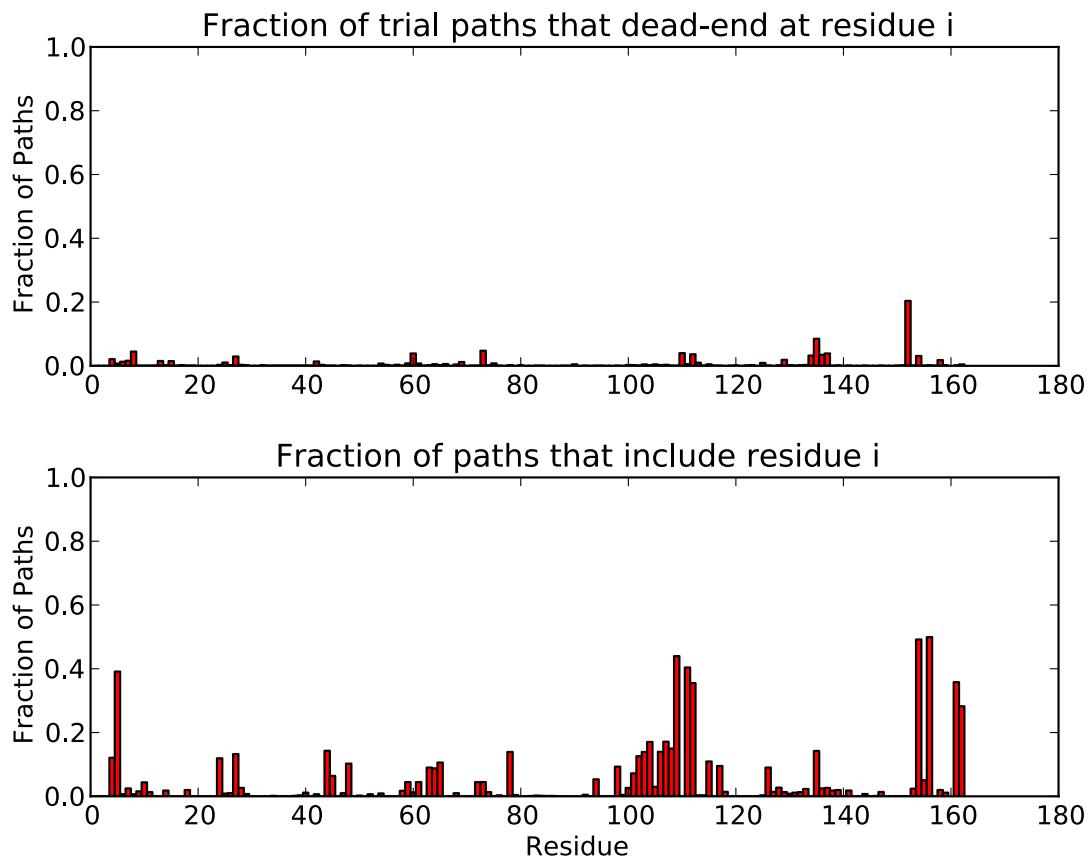
Supplementary Figure 1 | Contrasting an ensemble model and a qFit model of *E. Coli* dihydrofolate reductase ternary complex (E:NADP+:FOL). We generated a phenix.ensemble_refinement model and a qFit model from a 1.35Å room temperature wild-type ecDHFR (E:NADP+:FOL) data set (Supplementary Table 1). Using phenix.ensemble_refine with the protocol described by Gros and coworkers¹⁸, we performed a grid search that explored the parameters pTLS, wxray_coupled_tbath_offset, and tx. This procedure resulted in ensembles ranging in size from 28 to 167 models. The ensemble with the lowest Rfree (0.1917) contained 167 models (pTLS = 0.775, wxray = 4.375, tx = 1.5). The qFit model has Rfree = 0.1652, 2.65% lower than the ensemble model. To illustrate the differences in the models, five side chains of ecDHFR in a conformationally heterogeneous neighborhood are shown. (a) Results from phenix.ensemble_refinement with electron density calculated from $2mFo-Dfc$ coefficients corresponding to the ensemble model contoured at 1 σ (blue) and at 0.3 σ (green). While the density contours for Leu110, His114, and Glu154 seem well-defined, even at 0.3 σ , these side chains exhibit considerably heterogeneity in the ensemble. (b) The qFit model, shown for comparison in the same density, models these side chains with 1 (Leu110), 2 (His114) or 3 (Glu154) conformers that explain the spread in the electron density. The different rotamers of Leu156 are clearly distinguishable in the density, and modeled accordingly by qFit. These discrete states of Leu156 are difficult to observe in the ensemble model. Similarly, the density of Tyr128 is well-represented by the qFit multi-conformer model. The three Tyr128 conformations are furthermore supported by significant spread in the main-chain density (not shown).

Supplementary Figure 2



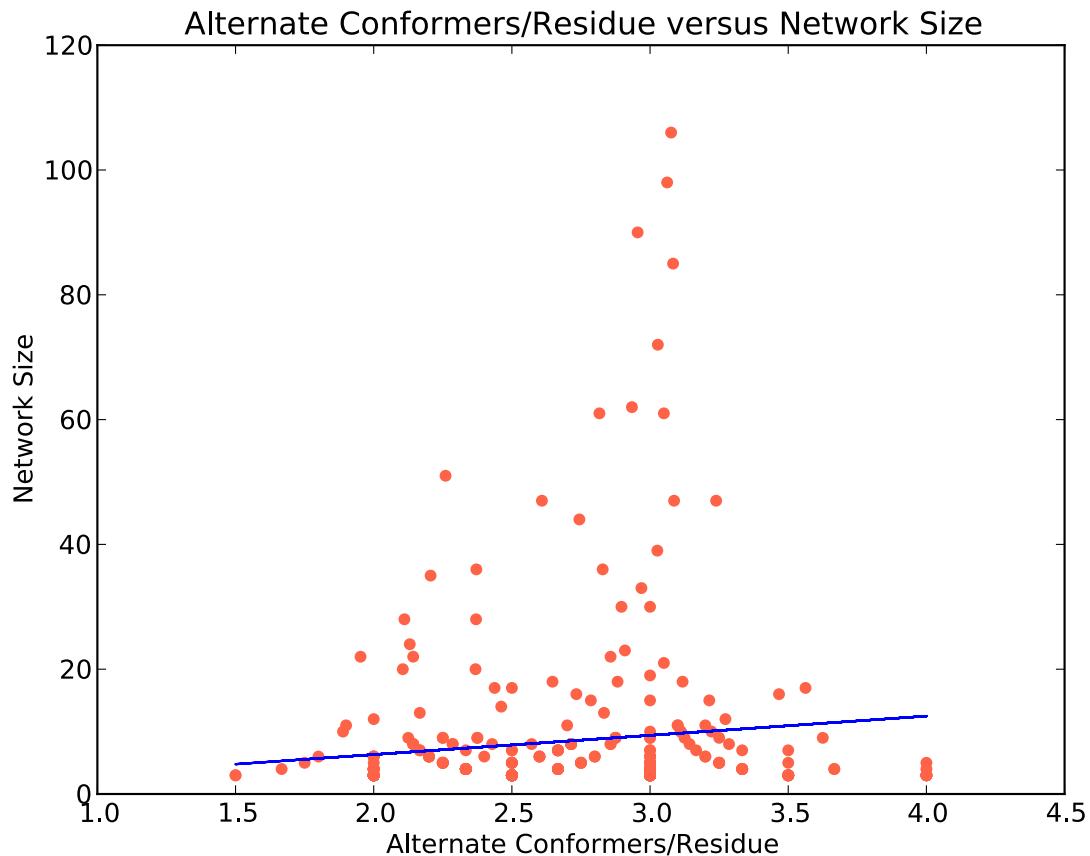
Supplementary Figure 2 | Flowchart of our algorithm to identify pathways. We first rebuild a starting protein model with qFit to identify alternate conformations in the electron density. Next, hydrogen atoms are added to the qFit model. At the 'Label' stage, alternate conformations in the qFit model are assigned a label A, B, etc. We use Monte Carlo Minimization for label assignment, i.e. labels are optimized (re-assigned) to obtain globally internally consistent alternate conformations by minimizing clashes within labels. This is followed by multiple rounds of conventional positional and atomic displacement parameter crystallographic refinement with phenix.refine. Then we apply CONTACT to calculate pathways based on overlapping van der Waals radii. Finally, residues that share pathways are grouped into networks.

Supplementary Figure 3



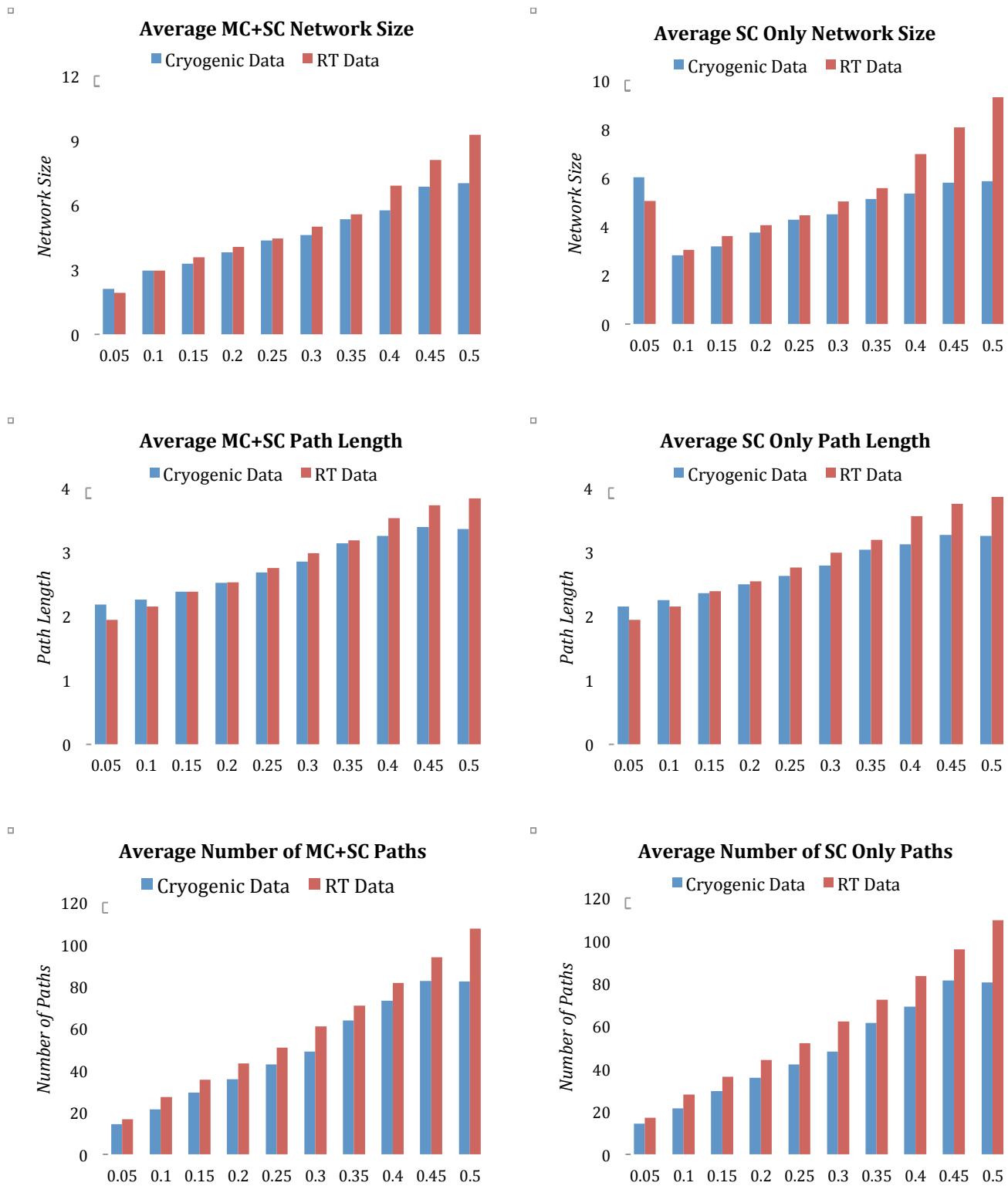
Supplementary Figure 3 | Example of frequencies of steric clash relief by residue in CYPA. The top panel shows the fraction of trial paths that dead-end at residue i because it is unable to relieve a steric clash. In this case, the entire path is invalidated. The bottom panel shows the fraction of final paths that include residue i .

Supplementary Figure 4



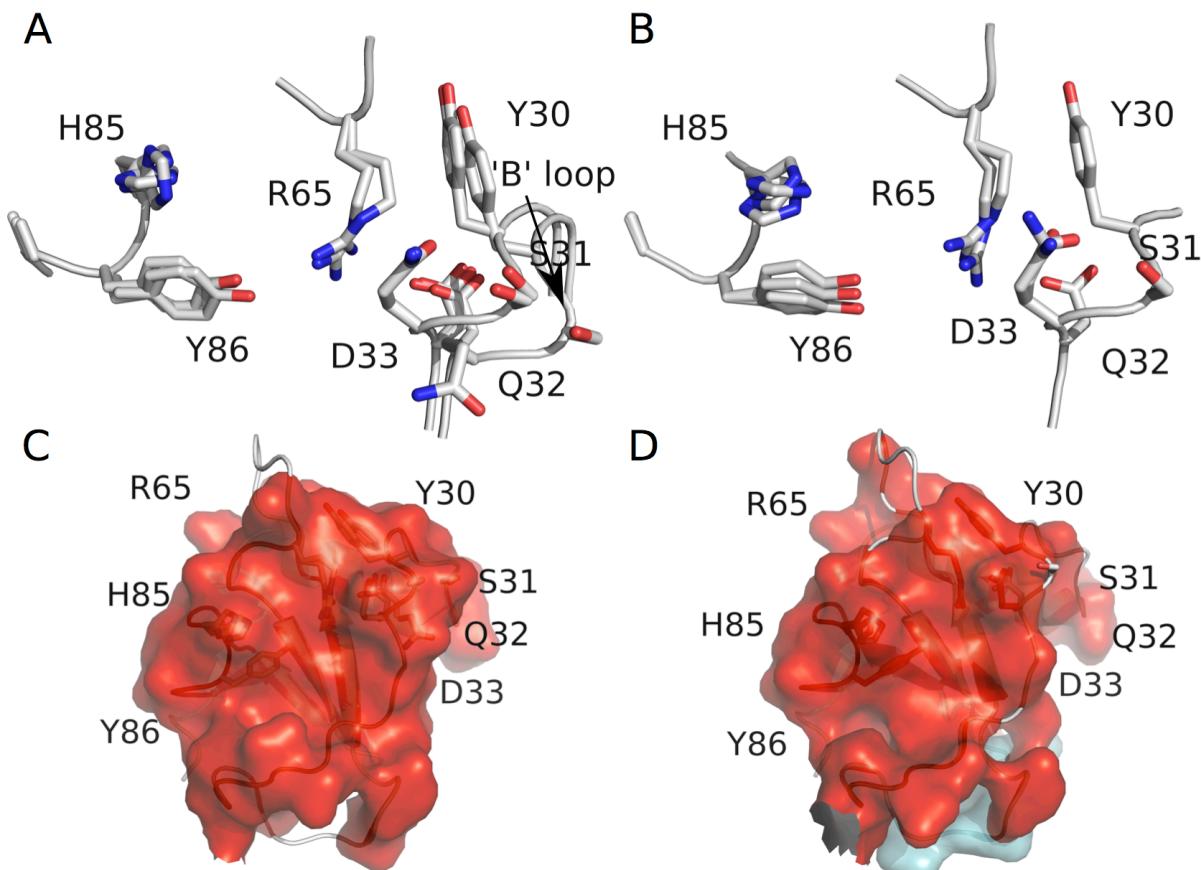
Supplementary Figure 4 | A scatterplot of the average number of conformers per residue versus networks size for 35 matched pairs of room temperature and cryogenic data sets. The average number is 2.5 conformers/residue per network at RT vs. 2.7 at cryogenic temperature (two-sided t-test $p=0.02$). For the 12 matched pairs that diffract to within 0.2Å the average number of alternate conformers/residue per network is 3.1 for RT vs. 2.8 for cryo ($p = 0.02$). We find a small correlation between the size of a network and the average number of conformers (Spearman Rank Correlation Coefficient = 0.14, $p < 0.01$).

Supplementary Figure 5



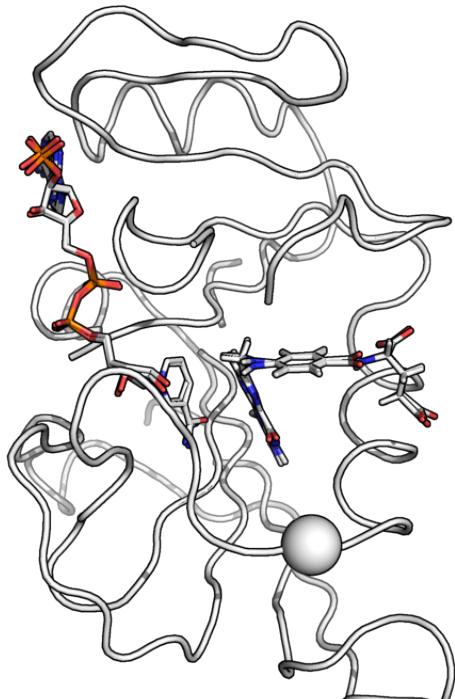
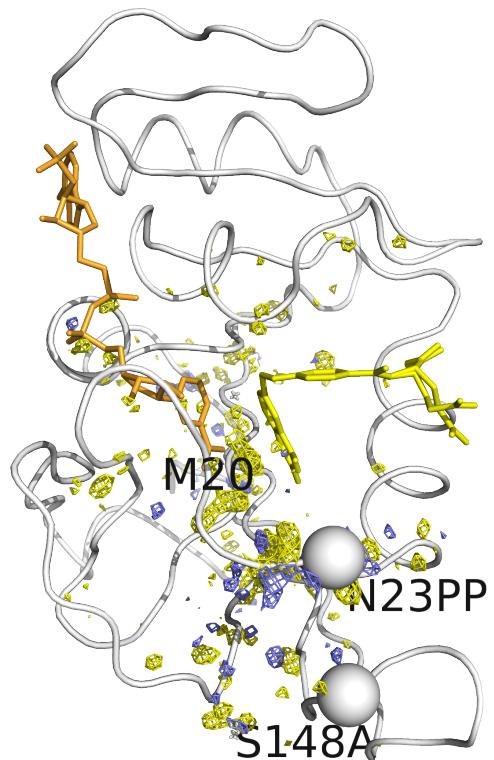
Supplementary Figure 5 | Summary statistics of pathways and networks in 35 matched pairs. Room temperature (red bars) and cryogenic temperature (blue bars) are indicated in each data set. We calculated the average size of networks, the number of pathways and the average length of pathways for each data set at percentiles of van der Waals overlaps ranging from 0.05 to 0.5 (horizontal axes). The graphs depict averages of these statistics for the 70 data sets. Statistics in the left column above were obtained with main- and side chain (MC+SC) interactions in CONTACT turned on, while the right column depicts side chain only interactions (SC Only). Overall, the graphs strongly indicate that for both MC+SC and SC Only interactions the size of networks and the length and number of pathways is greater in RT data than in cryogenic data across all levels of van der Waals overlaps.

Supplementary Figure 6



Supplementary Figure 6 | Atomic contacts in cryogenic and room temperature RNase Sa. **(a)** The cryogenic data set 1lni exhibits an alternate main-chain conformation in molecule A around residues A28-A33. The main conformation of this fragment is modeled by qFit, but it incorrectly models the alternate Ser31 and omits alternate conformations for Gln32 and Asp33 owing to large main-chain amplitude. **(b)** The alternate main-chain fragment is absent in the room temperature data set 1rgg. Alternate conformations for these residues result in a large number of additional pathways. In the qFit models Tyr30 is involved in 28% of pathways in 1lni vs. 1% in 1rgg, which propagate across the active site to neighboring residues Arg65 and Tyr86 that in turn are involved in more pathways in 1lni than in 1rgg (25% vs. 6% resp. 56% vs. 33%). When the correct ‘B’ fragment is inserted in the qFit model, the number of pathways exceeds the threshold. Despite the large difference in pathways, networks at cryogenic **(c)** and room temperature **(d)** are similar and in both cases comprise nearly all residues in the molecule. Cryocooling of RNase Sa may increase static disorder within the crystal. This behavior may reflect restraints in the lattice or the details of the glass transition specific to these crystals.

Supplementary Figure 7

A**B**

Supplementary Figure 7 | Additional views of ecDHFR and difference maps. **(a)** The ecDHFR ternary complex is shown in ribbons with folate and NADP⁺ rendered in sticks. The site of the proline mutation and insertion (residue 23) for the N23PPS148A mutant is indicated by a sphere. **(b)** Alternative coloring for **Fig. 5c** with yellow indicating positive and blue indicating negative difference density.

Supplementary Table 1

	N23PPS148A RT	WT RT	WT CRYO
PDB	4KJL	4KJK	4KJJ
Temperature(K)	273	273	100
Resolution range (Å)	50 - 1.38 (1.42 - 1.38)	50 - 1.35 (1.4 - 1.35)	50 - 1.15 (1.19 - 1.15)
Space group	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell	34.64 45.76 98.71	34.32 45.51 98.91	34.04 44.82 98.2
Total reflections	120725	242424	264643
Unique reflections	32720	31891	51212
Multiplicity	3.7 (2.5)	7.6 (3.4)	4.9 (1.2)
Completeness (%)	98.88 (91.85)	91.58 (48.73)	94.42 (64.25)
Mean I/sigma(I)	17.74 (2.17)	21.65 (2.42)	22.58 (5.82)
Wilson B-factor	11.73	12.07	7
R-sym	0.044 (0.451)	0.067 (0.451)	0.046 (0.281)
R-factor	0.1476 (0.1854)	0.1255 (0.1864)	0.1314 (0.1229)
R-free	0.1788 (0.2307)	0.1652 (0.2888)	0.1545 (0.1805)
Number of atoms	6125	6182	5967
macromolecules (qFit)	2831	2875	2679
macromolecules (single model)	1356	1350	1350
ligands	160	162	160
water	173	167	346
Protein residues	157	157	153
RMS(bonds)	0.011	0.013	0.014
RMS(angles)	1.32	1.57	1.88
Ramachandran favored (%)	98	98	98
Ramachandran outliers (%)	0	0	0
Average B-factor	12.3	12.4	8
macromolecules	11.1	11.3	6.6
solvent	29.4	29.3	19.5

Supplementary Table 1 | X-ray crystallography data collection and refinement statistics. E:NADP⁺:FOL complexes at room temperature (N23PP/S148A and WT ecDHFR) and at cryogenic temperature for WT ecDHFR. Note that the total number of atoms includes all atoms included in the model, but the macromolecule, ligand, and water counts include only those atoms refined individually and not hydrogens refined in riding positions.

Supplementary Table 2

Room Temperature												Cryogenic Temperature							
	ID	res	All-Atom paths				SC Only Paths				All-Atom Paths				SC Only Paths				
			R/R _{free}	R/R _{free}	Number	Length	Number	Length	ID	res	R/R _{free}	R/R _{free}	Number	Length	Number	Length			
			re-refined	qFit						re-refined	qFit								
1	1a2p	1.5	11.9/13.6	11.5/15.5	26	2.31	42	2.33	1b2x	1.8	16.1/21.0	15.8/20.9	38	2.45	43	2.35			
2	2dn2	1.25	13.9/15.4	14.5/16.5	95	3.01	104	2.53	1a3n	1.8	15.5/19.7	15.6/19.8	70	2.49	69	2.32			
3	1b2r	1.15	11.7/13.3	13.4/16.4	26	3.19	58	3.22	1a6g	1.15	12.9/14.3	14.4/17.1	35	3.60	33	2.85			
4	1b2z6	1.2	9.4/10.6	10.7/13.4	216	8.24	85	3.81	1a6k	1.1	12.7/13.7	13.8/16.2	29	3.55	36	3.14			
5	1b2p	1.15	11.4/13.1	13.0/16.2	89	6.44	69	3.17	1a6n	1.15	13.2/14.7	14/15.8	8	2.13	13	2.23			
6	1bx8	1.4	14.1/18.0	13.1/17.1	7	2.14	8	2.00	1bx7	1.2	14.6/16.4	15/18.4	24	3.71	14	2.00			
7	3tgp	1.31	17.0/21.8	17.3/22.1	-	-	135	5.80	1ctq	1.26	14.3/17.8	14.9/18.7	203	6.86	96	4.54			
8	1fdn	1.84	12.9/15.9	12.6/16.1	5	2.20	2	2.50	2fdn	0.94	9.6/10.3	12.5/13.5	7	2.14	10	2.60			
9	4gcr²	1.47	27.4/31.2	18.4/24.3	9	2.78	17	2.00	1gcs	2	17.8/22.8	17.2/23.8	26	2.77	27	2.56			
10	1gdu	1.07	11.0/12.5	10.8/13.6	91	5.04	69	2.62	1gdq	0.93	11.6/12.3	11.7/12.9	10	2.40	18	2.56			
11	1hbz	1.5	8.8/11.8	9.6/12.2	226	5.60	132	2.88	1gwe	0.88	9.2/9.8	9.9/10.6	-	-	183	3.63			
12	1i1x	1.11	9.6/11.2	10.4/12.3	438	8.45	103	3.34	1i1w	0.89	9.6/10.2	10.3/11.1	-	-	146	4.82			
13	2jcw	1.7	14.5/16.5	14.7/16.7	28	6.11	39	2.36	1jcv	1.55	15.8/18.4	15.2/18.5	37	4.30	38	2.84			
14	1jxu	0.99	21.7/23.8	22.0/23.8	8	2.88	6	2.50	1jxt	0.89	22.1/23.4	21.7/23.3	6	2.00	9	2.11			
15	1kwn	1.2	10.5/11.4	11.2/12.6	51	4.78	88	3.57	1ly0	1.36	13.0/15.2	12.7/14.8	88	4.44	78	2.45			
16	1l63	1.75	14.1/17.5	14.3/18.3	13	2.31	23	2.22	1lw9	1.45	15.5/19.3	15.8/18.8	4	2.50	4	2.50			
17	1l90	1.75	14.8/19.5	18.8/23.3	8	2.00	15	2.13	3dmv	1.65	15.9/19.1	15.9/19.8	1	2.00	4	2.00			
18	1rgg	1.2	10.2/11.3	11.0/13.2	283	8.21	168	5.04	1lni	1	12.0/12.9	13.1/14.6	1179	9.43	143	5.10			
19	1pic	1.33	12.7/16.1	13.0/16.6	702	9.68	111	5.32	1pnc	1.6	12.1/17.1	14/19.3	13	2.85	20	2.45			
20	1xvm	1.1	13.7/14.8	14.6/16.7	84	4.75	68	2.93	1pq5	0.85	9.8/10.2	11.3/12.2	21	2.81	29	2.93			
21	1tgt	1.8	15.8/19.0	16.3/21.3	14	2.57	10	2.00	1tgt	1.7	17.0/20.9	16.5/21	10	2.00	14	2.00			
22	1wme	1.5	10.9/13.1	11.2/14.0	195	6.41	128	2.78	1wmd	1.47	12.1/14.8	13.1/15.9	185	4.36	136	2.74			
23	1x6p	1.63	14.3/16.5	14.3/16.8	24	3.04	24	2.29	1x6q	1.51	16.1/19.0	15.2/20.2	12	2.50	16	2.19			
24	1yv6	1.78	14.6/19.5	14.6/19.5	12	2.50	15	2.13	1yv4	1.51	13.3/18.2	13/17.5	15	2.67	22	2.64			
25	3qyu	1.54	15.2/17.2	14.0/18.2	21	2.48	17	2.59	2bit	0.99	11.1/15.2	12.2/16.8	12	2.42	12	2.50			
26	2dfc	1.19	10.0/10.3	11.1/12.7	1011	11.12	115	4.69	2dfb	1.11	11.7/12.3	13.3/14.8	18	2.72	30	2.57			
27	3btk	1.85	16.3/19.2	16.0/18.7	42	3.38	64	2.64	2ftl	1.62	18.0/19.9	17.8/20	89	4.78	69	2.29			
28	2wt4	1.8	13.5/17.4	13.4/17.6	92	4.05	57	2.67	2wlt	1.4	11.5/13.6	12.9/15.4	65	4.23	58	2.41			
29	3djjg	1.8	14.6/19.5	14.8/19.6	66	2.55	84	2.31	3djj	1.1	10.9/11.5	12.8/14.5	39	2.28	62	2.37			
30	3k0n	1.39	11.8/15.3	12.1/15.1	102	4.85	85	3.65	3k0m	1.25	12.9/14.8	13.2/15.2	29	3.17	35	2.89			

31	3k0o	1.55	10.4/13.9	10.9/14.0	24	2.96	28	2.64	3k0p	1.65	14.9/17.8	15/17.8	16	2.31	25	2.52
32	3kyw	1.1	13.6/15.6	13.3/15.3	14	3.43	24	2.67	3kyv	1.1	12.0/13.1	11.8/13.4	9	4.00	22	2.95
33	4lzt	0.95	10.6/11.8	11.7/12.9	11	2.27	24	2.67	3lzt	0.93	9.9/10.7	10.7/12.1	24	2.88	31	2.35
34	3zsm	1.25	10.8/14.2	11.7/15.4	36	3.83	55	3.16	3zsl	1.08	15.3/17.9	16.8/19.8	112	6.05	61	3.92
35	4j5o	1.11	12.3/15.0	12.5/16.1	211	7.16	102	3.45	3s6e	0.95	11.6/12.7	12.1/13.5	93	4.22	79	3.35

Supplementary Table 2 | Refinement and pathway statistics of 35 matched pairs of protein models collected at room and cryogenic temperature. The pathway statistics were calculated at the 30th percentile of van der Waals overlaps. Entries '-' in the table indicate that the number of pathways exceeded 1,000,000. The average Rfree of qFit models was slightly worse (0.9%) than re-refined deposited models. However, the geometry of the qFit models was slightly better, with average bond length RMSDs improved by 0.0005Å and average bond angle RMSDs improved by 0.085 degrees. [#]Our automated re-refinement protocol led to unusually high R/R_{free} values for the 4gcr PDB model, but visual inspection of the resulting model revealed no gross inaccuracies or omissions.

Supplementary Table 3

Network	Residues
Red	7,8,14,18,44,63,64,65,98,99,102,105,113,115,125, NAP
Cyan	4,78,79,101,103,104,106,107,108,126,127,128,158
Green	109,112,133,135,137,154,155,156
Magenta	138,139,141,153
Yellow	5,94,FOL
Orange	25,26,147
Salmon	34,36,38
Grey	40,52,54
Blue	58,60,84

Supplementary Table 3 | Residues in WT E:NADP⁺:FOL networks. Colors refer to **Fig. 3**.

Supplementary Table 4

Network	Residues
Red	9,10,13,20,21,22,24,27,43,44,45,48,94,98,99,102,113,115,117,118,119,124,NAP
Cyan	4,8,93,109,110,111,112,114,133,134,135,136,137,138,139,140,141,152,153,154,155,156,159
Purple	63,74,75,76,77,78,80,81,82,103,106
Grey	50,54,59,72

Supplementary Table 4 | Residues in N23PP/S148A E:NADP⁺:FOL networks. Numbering corresponds to WT.
Colors refer to Fig. 5.

Supplementary Note

X-ray data collection temperature perturbs contact networks

To examine the effect of X-ray data collection temperature on conformational heterogeneity, we used CONTACT to analyze matched pairs of room and cryogenic temperature data sets (**Supplementary Table 4**). Because of the combined effect of reduced heterogeneity and radiation damage protection, most cryogenic datasets diffract to higher resolution than the corresponding room temperature datasets. Since the ability to resolve discrete alternative conformations depends on the resolution of diffraction^{19,20}, we first examined 12 pairs of high resolution (better than 1.5 Å) data sets that diffracted to within 0.2 Å, as was the case with CypA. We found that pathways are 54% longer (3.43 vs. 5.27 residues, p-value = 0.03) at room temperature than at cryogenic temperature, and contact networks are over two times larger (11.94 vs. 5.63 residues, p-value = 0.01) (**Fig. 2, left column**). Networks at room temperature on average contain slightly more alternate conformations per residue than those at cryogenic temperature (3.1 vs. 2.8, p = 0.02). (**Supplementary Fig. 4**.) Consistent with a recently proposed idea that protein domains are subdivided into multiple, functionally distinct groups of residues (“sectors”), our analysis found that domains are rarely globally coupled²⁴. Rather, domains are generally subdivided into a small number of contact networks that are more numerous (6 vs. 4.7) but smaller in size at cryogenic temperature (**Fig. 2, left column**).

Next, we expanded this analysis to examine 35 matched pairs of datasets diffracting to better than 2.0 Å resolution. We found that the number of residues in pathways (4.3 vs 3.3, p-value = 0.01) and contact networks (11.1 vs. 7.3) was greater at room temperature (**Fig. 2, right column** and **Supplementary Fig. 5**). This further supports the idea that cryogenic cooling can have a more widespread effect than simply reducing harmonic motion. A notable exception to this trend is a pair of RNase Sa datasets (1rgg/1lni) (**Supplementary Table 4**). The number of pathways at cryogenic temperature is nearly 5 times that of room temperature despite only a modest improvement in resolution (1.0 Å vs 1.2 Å) and ability to resolve alternative conformations. However, the residues implicated in these pathways are mostly shared between the two datasets (**Supplementary Fig. 6**).