

# Crystal Cryocooling Distorts Conformational Heterogeneity in a Model Michaelis Complex of DHFR

Daniel A. Keedy,<sup>1</sup> Henry van den Bedem,<sup>2</sup> David A. Sivak,<sup>3</sup> Gregory A. Petsko,<sup>4,5</sup> Dagmar Ringe,<sup>5</sup> Mark A. Wilson,<sup>6,\*</sup> and James S. Fraser<sup>1,\*</sup>

<sup>1</sup>Department of Bioengineering and Therapeutic Sciences and California Institute for Quantitative Biology, University of California, San Francisco, San Francisco, CA 94158, USA

<sup>2</sup>Joint Center for Structural Genomics, Stanford Synchrotron Radiation Lightsource, Stanford, CA 94025, USA

<sup>3</sup>Center for Systems and Synthetic Biology, University of California, San Francisco, San Francisco, CA 94158, USA

<sup>4</sup>Department of Biochemistry and Chemistry, Brandeis University, Waltham, MA 02454, USA

<sup>5</sup>Department of Neurology and Center for Neurologic Diseases, Harvard Medical School and Brigham and Women's Hospital, Boston, MA 02139, USA

<sup>6</sup>Department of Biochemistry and the Redox Biology Center, University of Nebraska, Lincoln, NE 68588, USA

\*Correspondence: [mwilson13@unl.edu](mailto:mwilson13@unl.edu) (M.A.W.), [james.fraser@ucsf.edu](mailto:james.fraser@ucsf.edu) (J.S.F.)

<http://dx.doi.org/10.1016/j.str.2014.04.016>

## SUMMARY

Most macromolecular X-ray structures are determined from cryocooled crystals, but it is unclear whether cryocooling distorts functionally relevant flexibility. Here we compare independently acquired pairs of high-resolution data sets of a model Michaelis complex of dihydrofolate reductase (DHFR), collected by separate groups at both room and cryogenic temperatures. These data sets allow us to isolate the differences between experimental procedures and between temperatures. Our analyses of multiconformer models and time-averaged ensembles suggest that cryocooling suppresses and otherwise modifies side-chain and main-chain conformational heterogeneity, quenching dynamic contact networks. Despite some idiosyncratic differences, most changes from room temperature to cryogenic temperature are conserved and likely reflect temperature-dependent solvent remodeling. Both cryogenic data sets point to additional conformations not evident in the corresponding room temperature data sets, suggesting that cryocooling does not merely trap preexisting conformational heterogeneity. Our results demonstrate that crystal cryocooling consistently distorts the energy landscape of DHFR, a paragon for understanding functional protein dynamics.

## INTRODUCTION

Although protein molecules are often depicted as static objects, they must adopt multiple conformations to play their physiological roles. Indeed, mutations that preserve the average structure but disrupt protein conformational heterogeneity can impair enzyme function (Bhabha et al., 2011; Fraser et al., 2009).

However, multiple conformations in proteins are difficult to characterize in detail with X-ray crystallography, mostly because a crystallographic electron density map represents an average of the ensemble of conformations populated throughout the crystal (Burnley et al., 2012; Lang et al., 2010; Smith et al., 1986; van den Bedem et al., 2009). Crystallographic B-factors combine positional uncertainty and thermal displacements by assuming harmonic motion about average atomic coordinates, but they are incapable of modeling the anharmonic motions that are common in proteins (Fenwick et al., 2014; Kuriyan et al., 1986). The difficulty of interrogating near-native protein conformational heterogeneity from crystallographic data has been a roadblock for applications such as small-molecule inhibitor discovery (Lorber and Shoichet, 1998) and protein engineering (Mandell and Kortemme, 2009).

Recent computational approaches explicitly model previously hidden conformational heterogeneity in all-atom detail from high-resolution ( $\leq 1.8$  Å) crystallographic data. For example, Ringer samples low-level electron density to reveal secondary side-chain conformations (Lang et al., 2010). Similarly, qFit automatically builds multiconformer models with one conformation for rigid regions, but a few (typically two to four) conformations that optimally explain the local electron density for flexible regions (van den Bedem et al., 2009). An alternative approach, PHENIX ensemble refinement, fuses molecular dynamics (MD) simulations with crystallographic refinement to obtain an ensemble of tens to hundreds of conformations that collectively fit the data (Burnley et al., 2012). These methods define multiple conformations, but do not define correlations between conformations. To fill this gap, the program CONTACT identifies steric clashes and reliefs between adjacent alternate conformations in multiconformer models to build networks of energetically coupled residues, which are often relevant to protein function (van den Bedem et al., 2013). Similarly, the program MutInf (McClendon et al., 2012) identifies statistically correlated torsion angles in traditional molecular dynamics simulations, and could be adapted to identify correlations in crystallographic ensemble models (Burnley et al., 2012).

These computational methods provide an opportunity to move beyond a single coordinate set and to link multiple conformations of proteins to biological function. Our previous work has revealed significantly increased heterogeneity and conformational coupling in models based on crystallographic data collected at room temperature (RT) instead of the much more common cryogenic temperature (cryo) across more than 30 proteins (Fraser et al., 2011; van den Bedem et al., 2013). Although cryocooling reduces radiation damage during data collection (Juers and Matthews, 2004), detailed analysis for a smaller number of RT/cryo pairs suggested that the conformational heterogeneity suppressed by cryocooling may often be important to the protein's biological activity (Fraser et al., 2009, 2011; Rasmussen et al., 1992; van den Bedem et al., 2013). These studies extend previous work that inferred quenched dynamic motions from harmonic crystallographic B-factors. For example, four structures of metmyoglobin at temperatures ranging from 220 to 300 K showed that mean-square atomic displacements increased with temperature (Frauenfelder et al., 1979). Later, data sets at temperatures ranging from 80 to 320 K established the relationship between protein contraction and the glass transition at  $\sim 200$  K (Frauenfelder et al., 1987; Tilton et al., 1992).

Despite additional analyses of existing RT/cryo pairs (Fraser et al., 2011; Juers and Matthews, 2001) and theoretical predictions of the effects of cryocooling on protein conformational heterogeneity (Halle, 2004), there remain many significant questions about cryocooling. Does the cryocooling process simply trap a subset of the disorder present at room temperature, or does it establish new favorable conformations by perturbing the free energy landscape of the protein in the crystal? Are the structural effects of cryocooling robust and repeatable, or do they depend strongly on the specifics of the process, such as the experimenters and laboratory environment? These questions are difficult to answer by studying a single RT/cryo pair—ideally, multiple data sets at both temperatures would be analyzed to account for experimental variability.

Here, we examine two independently collected pairs of high-resolution data sets for a model Michaelis complex of dihydrofolate reductase (DHFR:folate:NADP<sup>+</sup>). DHFR is an excellent model system: the current paradigm of catalytically linked conformational flexibility in enzymes has been strongly influenced by studies of structural dynamics in DHFR with crystallography and nuclear magnetic resonance (NMR; Boehr et al., 2006; Falzone et al., 1994; Sawaya and Kraut, 1997). Although many structures of DHFR complexes are available in the Protein Data Bank (PDB) as a result of these studies, they include a variety of substrate mimics, inhibitors, etc., and almost all were collected at cryogenic temperature. In contrast, the data sets we compare here all feature the same model Michaelis complex and were collected by our different laboratories at different synchrotrons in different decades, but have matching space group and unit cell parameters. Overall, our results show that the standard practice of crystal cryocooling idiosyncratically yet reproducibly disrupts the structures of proteins, and thus may limit our ability to derive useful information about the relationship of conformational heterogeneity to biological function.

## RESULTS

### Isomorphous Difference Density Maps Reveal Idiosyncratic Effects of Cryocooling

To determine whether laboratory or temperature had a larger effect on the conformational heterogeneity of DHFR, we first computed four isomorphous  $F_o1-F_o2$  difference electron density maps: 13cryo minus 05cryo, 13RT minus 05RT, 13RT minus 13cryo, and 05RT minus 05cryo (where, e.g., 13RT indicates the 2013 RT data set; Figure 1). We phased each map with the  $F_o1$  qFit model (Table 1).

To facilitate a direct comparison of the four difference maps, we contoured them at the same absolute electron density value of  $0.4 \text{ e}^-/\text{\AA}^3$  (Lang et al., 2014; Figure 1). Interestingly, this threshold corresponded to a much lower SD electron density value for the cryo-cryo map ( $2.43 \sigma$ ) than for the RT-RT map ( $4.55 \sigma$ ) and intermediate values for the two RT-cryo maps ( $3.01$  and  $4.55 \sigma$ ). Thus, the distribution of absolute electron density differences in the cryo-cryo map is broadened relative to the RT-RT map. The electron density differences are also amplified: the cryo-cryo map has very strong and widespread peaks of both signs, but the RT-cryo map is comparatively devoid of peaks (Figure 1), suggesting that cryocooling affected both the 2013 and 2005 crystals to similar extents, but in markedly different ways. The pronounced features of the map comparing cryogenic data sets collected by different labs, in contrast to the relative flatness of the map comparing RT data sets collected by different labs, show that protein structure in crystallo is affected more by cryocooling than by the lab performing the experiment.

Both RT-cryo maps have widespread positive and negative peaks with absolute value  $> 0.4 \text{ e}^-/\text{\AA}^3$  (Figure 1). These differences are not attributable to gross structural changes because the qFit models superimpose closely overall (all  $< 0.25 \text{ \AA } C_\alpha$  root-mean-square deviation [rmsd]). Rather, they are likely due to altered near-native conformational heterogeneity. The largest peaks in both RT-cryo maps are of similar magnitudes and suggest that the protein is more conformationally heterogeneous in both RT data sets. Nevertheless, the presence of many negative peaks with absolute value  $> 0.4 \text{ e}^-/\text{\AA}^3$  distributed throughout DHFR in both RT-cryo maps shows that the disorder present at cryogenic temperature is not simply a trapped subset of the disorder present at RT. This finding is consistent with previous comparisons of RT and cryogenic structures (Fraser et al., 2011) and theoretical work on the statistical mechanics of crystal cooling (Halle, 2004). Interestingly, the most pronounced peaks in both RT-cryo maps are located near the interface between the folate substrate and NADP<sup>+</sup> cofactor in the active site, suggesting that cryocooling perturbs conformational heterogeneity that may be important for substrate interactions.

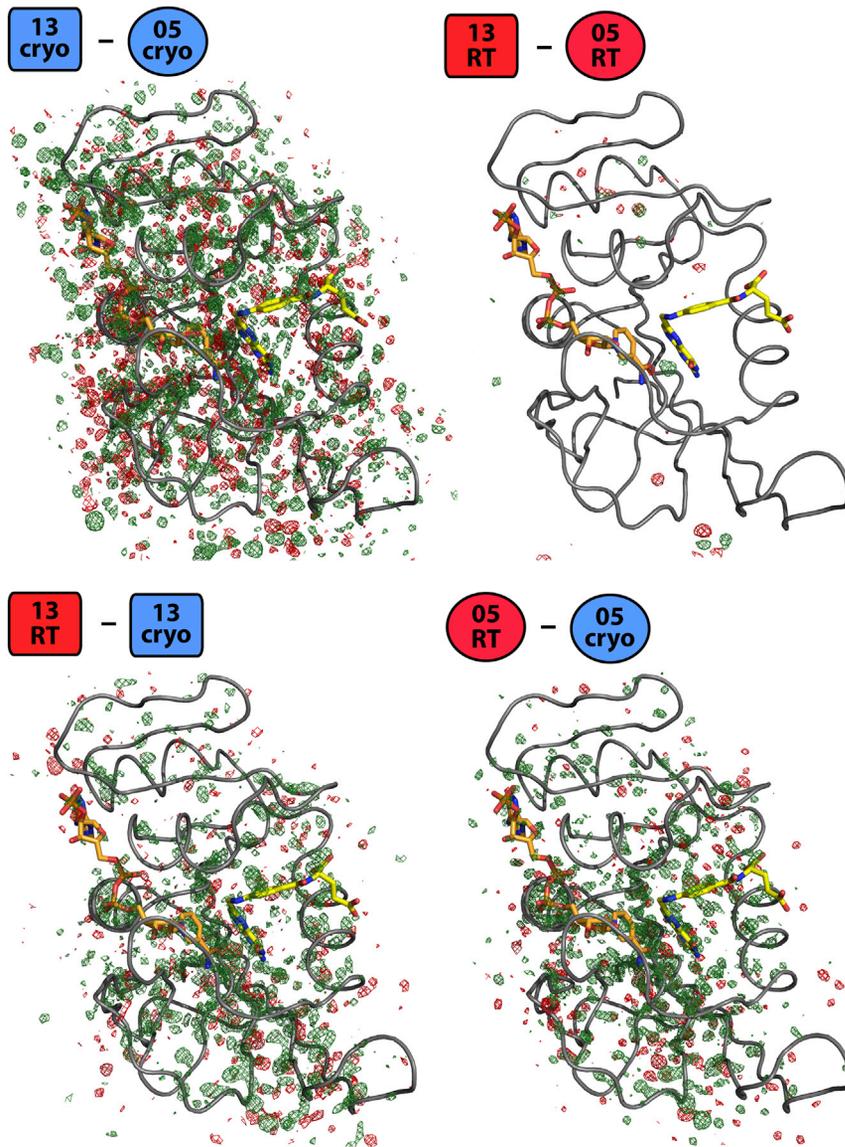
### Multiconformer Models Have Temperature-Dependent Alternate Conformations

To examine the structural basis of changes implied by the isomorphous difference maps, we compared multiconformer models built using qFit (van den Bedem et al., 2009). This approach fits multiple conformations only where locally supported by the real-space electron density distribution. The structure validation statistics for all four qFit models are similar (Table 1). In agreement with the predominance of positive peaks in the RT-cryo

## Structure

### Crystal Cryocooling Alters DHFR Conformations

CellPress



**Figure 1. Cryocooling Induces Changes in DHFR Structural Heterogeneity**

The broad distribution of strong difference peaks in the isomorphous  $F_o - F_o$  13cryo-05cryo map, in contrast to the relative flatness of the 13RT-05RT map, emphasizes that cryogenic freezing perturbed the two crystals differently, resulting in widespread and unpredictable changes to dynamics. Both the 13RT-13cryo and 05RT-05cryo difference density maps have significant peaks of both signs (green, positive; red, negative), suggesting that cryocooling idiosyncratically alters the same protein's structure and/or dynamics in different crystals at the hands of different crystallographers. All four maps are shown contoured at  $0.4 \text{ e}^-/\text{\AA}^3$  (Lang et al., 2014), which in these cases corresponds to 2.61, 3.01, 2.43, and 4.55  $\sigma$ , respectively. NADP<sup>+</sup> in orange; folate in yellow. Model shown is qFit model of first structure in  $F_o - F_o$  subtraction.

at cryo than RT, and approximately one-third of those adopt new rotamers. These results together further support the idea that cryocooling not only selects subsets of pre-existing conformational heterogeneity, which may represent static disorder (Fenwick et al., 2014), but also reshapes the free energy landscape to favor a different set of conformations.

To explore the extent to which cryocooling consistently perturbed the same regions of DHFR, we examined the colocalization of residues with altered conformational heterogeneity. Many of the same residues have altered rotamer ensembles in both the 2013 and 2005 RT/cryo pairs (Figure 2). The overlaps in which residues are consistently polymorphic by temperature are statistically significant (Experimental Procedures), which indicates that cryocooling consistently affects specific regions in a reproducible manner.

difference maps (Figure 1), the RT models have slightly higher average B-factors, indicating higher thermal mobility. The 05RT structure (0.85 Å) has the greatest number of residues modeled in multiple conformations. Collectively, these results confirm that the protein is more flexible at RT than cryo and that this flexibility is less discretely observable at slightly lower resolution.

To investigate the basis of these temperature-dependent changes to side-chain ensembles in more detail, we compared all-atom side-chain conformations represented in two ways: in dihedral space by assigning each conformation to a rotamer (Lovell et al., 2000) and in Cartesian space by calculating the maximum rmsd across all alternate conformations for each residue. Both methods reveal that cryocooling affects side-chain conformations at a global level in the 2005 and 2013 crystals similarly (Table S1 available online). For both structure pairs, approximately two-thirds of the residues in DHFR are more disordered at RT than cryo, and ~25% adopt entirely new rotamers. Furthermore, ~20% of residues are more disordered

tently affects specific regions in a reproducible manner. Despite these consistencies, some of the detailed effects of cryocooling differ between the two experiments. The differences in side-chain and main-chain rmsd between the 2013 RT and cryo structures correlate poorly with the corresponding differences in the 2005 structures (Figure 3). This result indicates that cryocooling perturbs the local conformational heterogeneity of many regions idiosyncratically, an interpretation that is further supported by the presence of significant peaks in the cryogenic minus cryogenic isomorphous electron density map as compared to the RT minus RT map (Figure 1).

#### **Time-Averaged Ensembles Show that Loop Heterogeneity Is Suppressed by Cryocooling**

The multiconformer models from qFit may fail to capture larger-scale conformational disorder present in the crystal, such as helix winding/unwinding and loop opening/closing. To determine if such larger-scale features are present and perturbed by

**Table 1. Statistics for Crystallographic Data Sets and qFit Models**

	13RT	13cryo	05RT	05cryo
Resolution range (Å)	41.34–1.35 (1.39–1.35)	40.77–1.15 (1.19–1.15)	49.36–1.05 (1.09–1.05)	40.78–0.85 (0.88–0.85)
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>			
Unit cell	34.3, 45.5, 98.9, 90, 90, 90	34.0, 44.8, 98.2, 90, 90, 90	34.4, 45.5, 98.7, 90, 90, 90	34.0, 44.8, 98.3, 90, 90, 90
Unique reflections	31,891 (1,671)	51,212 (3,410)	71,709 (6,949)	130,114 (12,526)
Completeness (%)	91.58 (48.73)	94.42 (64.25)	98.24 (96.57)	98.05 (95.74)
Mean I/σ(I)	21.7 (2.4)	22.6 (5.8)	23.3 (5.1)	16.8 (2.5)
Wilson B-factor	12.07	7.00	10.52	7.68
$R_{sym}$	0.067 (0.451)	0.046 (0.281)	0.045 (0.458)	0.049 (0.550)
$R_{work}$	0.1268 (0.2044)	0.1366 (0.1359)	0.1211 (0.1271)	0.1546 (0.2839)
$R_{free}$	0.1659 (0.3121)	0.1573 (0.1848)	0.1419 (0.1490)	0.1666 (0.2820)
$R_{gap}$ ( $R_{free} - R_{work}$ )	0.0391 (0.1077)	0.0207 (0.0489)	0.0208 (0.0219)	0.0120 (–0.0019)
Number of non-hydrogen atoms	3,204	3,184	3,698	3,273
Macromolecules	2,875	2,679	3,337	2,764
Ligands	162	160	162	162
Water	167	345	199	347
Rmsd(bonds)	0.017	0.018	0.016	0.017
Rmsd(angles)	1.54	1.69	1.61	1.71
Ramachandran favored (%)	98	98	98	98
Ramachandran outliers (%)	0	0	0	0
Clashscore	6.47	13.04	4.07	5.63
MolProbity score	1.463	1.625	1.300	1.414
Average B-factor	12.4	8.0	13.3	9.8
Macromolecules	11.5	6.5	12.2	8.1
Ligands	11.2	9.6	11.5	9.6
Water	29.3	19.4	33.4	23.3
Number of protein residues	159	159	159	159
Number of protein residues with alternate conformations	141	139	141	142
Average number of alternate conformations per protein residue	2.8	2.7	3.2	2.7
Average number of alternate conformations per multiconformer protein residue	3.0	3.0	3.5	2.9
qFit model PDB code	4KJK	4KJJ	4PST	4PSS
Ensemble PDB code	4P3Q	4P3R	4PTJ	4PTH

Numbers in parentheses are for highest-resolution 10% of reflections.

cryocooling, we turned to the recently described PHENIX ensemble refinement method (Burnley et al., 2012). We began with a grid search over the  $\rho_{TLS}$ ,  $w_{X-ray}$ , and  $\tau_x$  parameters, which balance the scope of the translation/libration/screw (TLS) model, the weighting of crystallographic data relative to the MD force field, and the degree of “memory” for calculating rolling-average structure factors during time-resolved MD, respectively (Burnley et al., 2012). The  $R_{free}$  values were relatively similar across the 75 trial ensembles for each data set: more than half are within 0.02 of the lowest  $R_{free}$  for each data set (Table 2). To test whether cryocooling affected the ensembles and the multiconformer models similarly, we plotted  $C_\alpha$  root-mean-square fluctuation (rmsf)

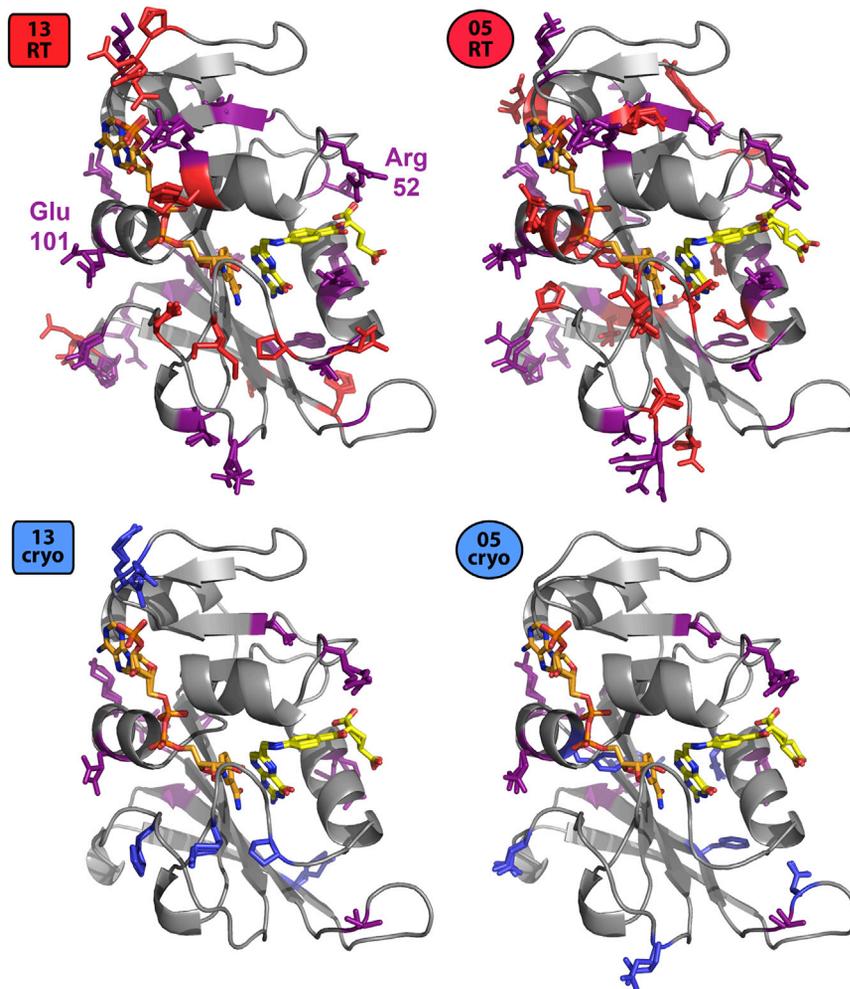
against sequence for the lowest- $R_{free}$  ensemble for each data set (Figure S1). However, because many ensembles fit the data nearly equally well for each data set, we also plotted the average  $C_\alpha$  rmsf against sequence for the subset of ensembles with  $R_{free}$  within 0.02 of the lowest  $R_{free}$  for each data set (Figure 4).

Our analysis cautions against relying solely on the conformational heterogeneity present in any single refined ensemble. For example, the temperature-related difference for the active-site Met20 loop in the lowest- $R_{free}$  ensembles (Figure S1) is provocative because this loop interconverts between several well-characterized states (Sawaya and Kraut, 1997) in the catalytic cycle of DHFR (Bhabha et al., 2011; Boehr et al., 2006). In the ensemble

## Structure

### Crystal Cryocooling Alters DHFR Conformations

CellPress



**Figure 2. Cryocooling Alters Side-Chain Rotamer Heterogeneity Preferentially in Certain Protein Regions**

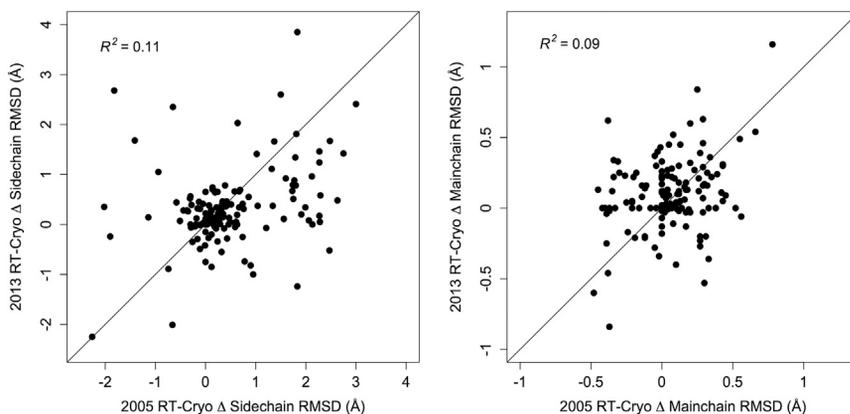
Many residues have at least one rotamer in the 13RT or 05RT structure that is missing from the corresponding cryo structure (red+purple side chains, top row). Smaller, but still substantial, sets of residues have at least one rotamer in the 13cryo or 05cryo structure that is missing from the corresponding RT structure (blue+purple side chains, bottom row). Notably, many of these residues have altered heterogeneity consistently in both the 2013 and 2005 structure pairs (purple side chains), as opposed to just one or the other. Glu101 (Figure S2) and Arg52 (Figure 6) are labeled in the top right image.

with the lowest  $R_{free}$  value for each of the four data sets, the Met20 loop fluctuates around the closed state observed previously in other ternary enzyme:folate:NADP<sup>+</sup> ternary complexes (Sawaya and Kraut, 1997; Figure S1). While the 13RT ensemble hints at a minor population similar to the occluded state seen in enzyme:folate binary complexes (Sawaya and Kraut, 1997), this is likely due to the lowest- $R_{free}$  ensemble using a combination of input parameters that down-weights the contribution of crystallo-

graphic data (relatively low  $w_{X-ray}$  and  $\tau_x$ ; see Table 2). In contrast to the Met20 loop, the differences induced by temperature change in the Arg52 and Tyr128 loops persist in the averaged traces (Figure S1; Figure 4), confirming that they are robust to parameter choice in the ensemble refinements. Taken as a whole, the time-averaged ensembles support the idea that cryocooling reduces protein flexibility, especially at specific surface regions.

### Unit Cell Shrinking and Water Freezing Dampen Flexibility at the Protein Surface

The above comparisons of multi-conformer models and time-averaged ensembles strongly suggest that cryo-



**Figure 3. Cryocooling Affects Side Chain and Main Chain Conformational Heterogeneity Idiosyncratically at the Detailed Level**

Changes in the maximum side-chain (left) and main-chain (right) rmsd across all alternate-conformation combinations at each residue from the RT to the cryogenic qFit model are poorly correlated between the 2013 and 2005 pairs.  $R^2$  correlation coefficients are from linear least-squares fits. Diagonal lines are for visual comparison.

**Table 2. Multiconformer qFit Models Are More Parsimonious than Ensemble Models and Are Affected by Cryocooling-Induced Lattice Shrinking**

	13RT	13cryo	05RT	05cryo
Ensemble $p_{TLS}$	0.55	0.775	0.775	0.55
Ensemble $w_{X-ray}$	4.375	2.5	10.0	10.0
Ensemble $\tau_x$	1.0	2.0	1.0	2.0
Ensemble $R_{free}$	0.1534	0.1362	0.1655	0.1445
qFit model $R_{free}$ (from Table 1)	0.1659	0.1573	0.1419	0.1666
Ensembles with $R_{free}$ within 0.02 of lowest $R_{free}$	51/75 (68%)	61/75 (81%)	38/75 (51%)	53/75 (71%)
Ensembles with $R_{free}$ within 0.01 of lowest $R_{free}$	44/75 (59%)	31/75 (41%)	6/75 (8%)	35/75 (47%)
Number of unique reflections	31,891	51,212	71,709	130,114
Estimated ensemble parameters	720,198	1,152,459	545,874	1,152,390
Estimated qFit model parameters	27,996	26,931	32,287	27,722
Ensemble parameters per observation	22.6	22.5	7.6	8.9
qFit model parameters per observation	0.9	0.5	0.5	0.2
Number of ensemble water molecules	188.6	87.5	184.5	103.6
Crystal contact surface area ( $\text{\AA}^2$ )	1,685.8	2,628.5 (+56%)	2,224.9	2,840.4 (+28%)
Number of protein-water contacts	20,746	35,473 (+71%)	28,954	39,369 (+36%)
Number of protein-water H-bonds	12,333	20,785 (+69%)	17,019	21,090 (+24%)

For qFit models, estimated number of parameters = number of isotropically refined atom positions \* 4 (for  $x$ ,  $y$ ,  $z$ , and one parameter for  $B$ ) + number of anisotropically refined atom positions \* 9 (for  $x$ ,  $y$ ,  $z$ , and six parameters for  $B$ ). For ensemble models, estimated number of parameters = number of models \* number of atom positions per model \* 3 (for  $x$ ,  $y$ ,  $z$ ). Parameters per observation = number of unique reflections from Table 1 divided by estimated number of parameters. All-atom contacts were calculated using Probe (Word et al., 1999) for all four qFit models. For defining crystal contact surface area, the NADP<sup>+</sup> and folate were grouped with the protein but the waters were grouped with the crystal lattice. Protein-water contacts (van der Waals interactions and/or hydrogen bonds) include contacts between the protein and symmetry-related waters.

the 2005 qFit structures (Table 2). Both cryo structures have approximately twice as many modeled waters as the RT structures (Table 1, Figure 5). A large proportion of these additional waters contacts with the protein surface, often via hydrogen bonds (Table 2), and contributes to lattice contacts. This ordering of surface-associated waters is driven by cryocooling, not increased resolution, as evidenced by the lower-resolution (1.15  $\text{\AA}$ ) 13cryo structure having ~70% more ordered waters than the higher-resolution (1.05  $\text{\AA}$ ) 05RT structure (Table 1). Average B-factors for cryogenic waters without a corresponding RT water are higher (22.7 and 28.1  $\text{\AA}^2$  for 13cryo and 05cryo) than for those with a corresponding RT water (1.5  $\text{\AA}$  or closer; 14.4 and 18.8  $\text{\AA}^2$ ); thus, waters ordered uniquely by cryocooling may be somewhat more mobile (Movies S1, S2, and S3).

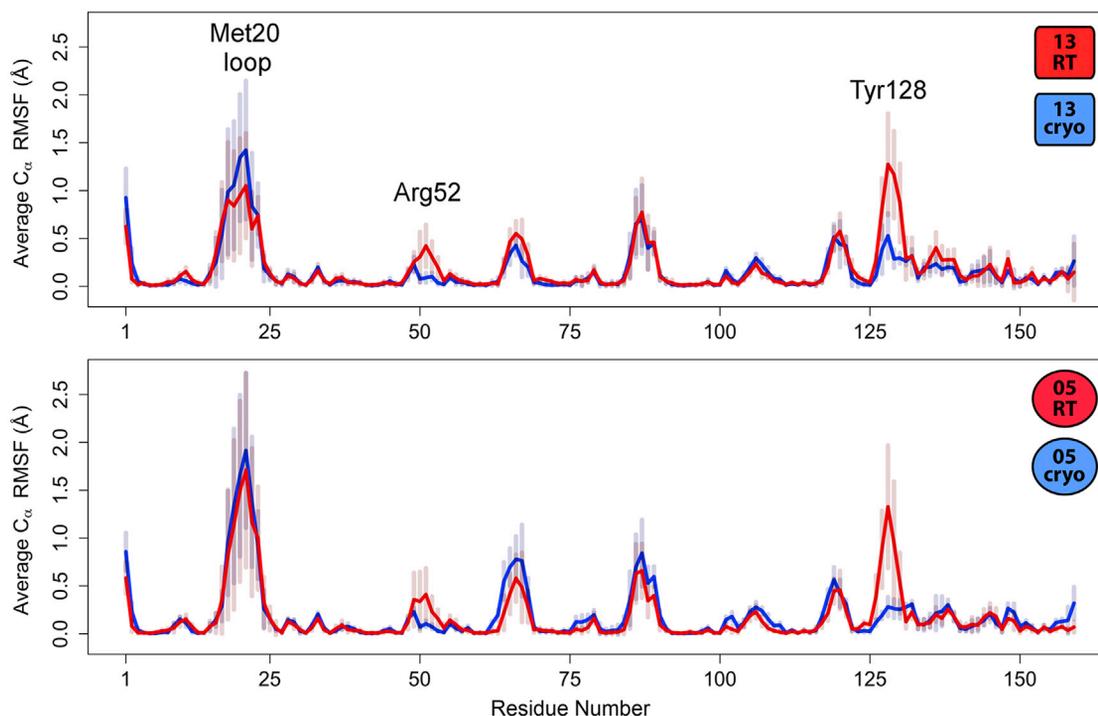
Unit cell contraction from crystal cryocooling conformationally restricts many waters in the narrowing channels between neighboring protein molecules in the lattice, which interestingly precludes some previously accessible alternate protein side-chain conformations. For example, Glu101 adopts  $\chi_1 t$  and  $m$  rotamers in both RT structures, but only  $\chi_1 m$  rotamers in both cryogenic structures; this is due to the presence of a frozen water that sterically prevents any  $\chi_1 t$  rotamers (Figure S2). Another example is Arg52 (labeled in Figure 4), which is positioned next to both the folate substrate and a crystal lattice interface (Figure 6). Cryocooling causes a series of alterations to the local energy landscape: Asp87\* switches rotamers to point away from Arg52, allowing Arg52 to move into this vacated space with a new rotamer (*ptp85*) and form previously nonexistent hydrogen bonds with Asp70\*. The rotamer that placed Arg52 near the folate substrate in the RT structures (*ptt85/ptt-85*) is abandoned in the cryo-

genic structures; a frozen water substitutes for any electrostatic interactions between the guanidinium moiety and the substrate.

In addition to these side-chain changes, we observe several main-chain alterations. For example, the 126-130 loop is more flexible in both RT time-averaged ensembles (Figure 7). The temperature dependence appears to be due to the appearance in both cryogenic ensembles of new ordered water molecules that bridge the polar side-chain moieties of Asp127 and Glu129 via a small local hydrogen-bond network, thereby stabilizing a more rigid loop conformation. In the RT ensembles, this region is unconstrained by these waters, and samples conformations that explore the adjacent solvent-exposed channel in the crystal (Figure 7, upper left).

### Networks of Coupled Conformational Heterogeneity Are Quenched at Cryogenic Temperature

The results above suggest that cryocooling both suppresses preexisting conformations and introduces new ones for many residues, particularly at the protein surface. To investigate whether these localized effects also affect collective intramolecular conformational heterogeneity, we used CONTACT, which propagates steric clash-relief pairs between alternate conformations (van den Bedem et al., 2013). Pathways of residues that participate in these clash-relief pairs are then connected into networks of residues that may be mutually conformationally coupled. This method previously identified a central dynamic contact network linking the two subdomains of DHFR via the active-site NADP<sup>+</sup> cofactor, which was corroborated by NMR chemical shift changes induced by a mutation that impairs enzyme function (G121V; Boehr et al., 2013).



**Figure 4. Room Temperature Ensembles Are Consistently More Flexible than Cryogenic Ensembles at Several Surface Loops**

$C_{\alpha}$  rmsf values averaged across the subset of the 75 total ensemble variants for each data set with  $R_{free}$  within 0.02 of the lowest  $R_{free}$  (Table 2) are higher for each RT ensemble (red lines) than the corresponding cryogenic ensemble (blue lines) at Arg52 and Tyr128 (labeled), although SDs are substantial (semitransparent bars). These regions are explored further below (Figures 6 and 7). The Met20 loop (labeled) is flexible, similarly so, in all four averaged traces.

The higher-resolution 05RT qFit model has more abundant and longer pathways than the lower-resolution 13RT qFit model (Table S2). Of particular note is the folate substrate, which is more flexible in the higher-resolution 05RT model ( $0.62 > 0.31$  Å rmsd between heavy-atom positions in the two modeled alternate conformations). As a result, CONTACT detects a clash-relief event involving the folate and places the folate in the aforementioned central network in the 05RT, but not the 13RT, model (Figure 8). This important network is not evident in either cryogenic model (Figure 8).

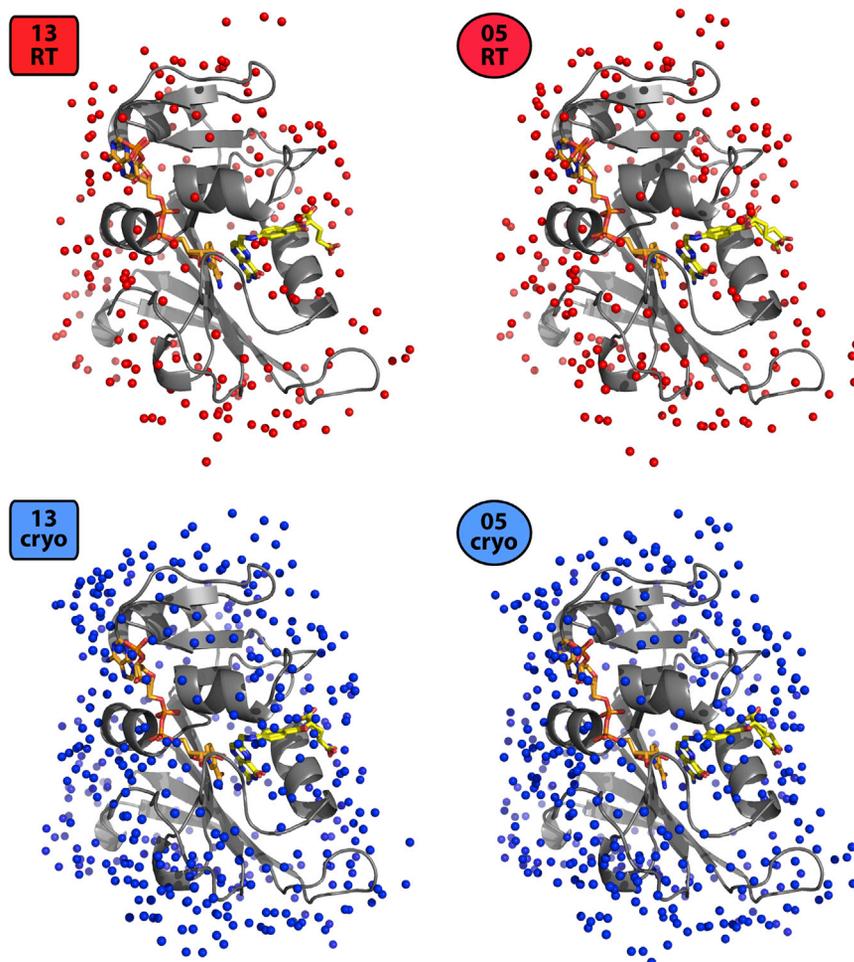
Changes in network statistics for the two RT/cryo pairs are quite similar (Table S2). For example, both RT structures have more than 130 unique pathways, but both cryo structures are reduced to  $\sim 40$  pathways. Similarly, the pathways shrink on average in the cryo structures, with 21% and 39% decreases in average pathway length. These results imply that cryocooling perturbs DHFR's energy landscape similarly in both crystals, fragmenting collective conformational heterogeneity—including known functional networks—into smaller pockets.

## DISCUSSION

The paired RT and cryo DHFR data sets presented here provide a unique opportunity to explore the effects of cryocooling on protein crystals. In large part, the conformational heterogeneity in each cryogenic structure is a subset of that in the corresponding RT structure, as shown by multiconformer rotamer analysis and ensemble rmsf traces. However, both the rotamer analysis and

the RT-cryo isomorphous electron density maps also show that some conformations are unique to the cryogenic structures. Unit cell shrinking and the dramatic increase in ordered water suggest how at least some of these new conformations are induced: crystal contacts and/or hydrogen bonds with new waters stabilize new surface side-chain conformations in lieu of previously populated conformations with more intramolecular interactions (Tyka et al., 2011). These results are in line with MD simulations with the solvent at a different temperature from the protein, which showed that solvent mobility is the dominant factor in determining differences in protein surface flexibility across the glass transition (Vitkup et al., 2000).

These changes in solvent ordering and surface residue conformations seem to propagate to the rest of the protein, as evidenced by the ablation of intramolecular dynamic CONTACT networks in both cryo qFit models relative to their RT counterparts; i.e., crystal cryocooling likely distorts functional protein conformational heterogeneity “from the outside in.” This may be in part because evolution has selected the sequence of DHFR to populate correlated states at physiological temperature (Halabi et al., 2009), but has not had an opportunity to act upon the new cryocooling-induced states of that same polypeptide chain. It remains unclear how the effects of cryocooling could be transmitted from surface side chains and waters to intramolecular networks in these crystals. For example, residues with altered rotamers at cryogenic temperature and residues omitted from the primary functional network do not colocalize more than expected. However, both of these sets of residues could



**Figure 5. Cryocooling Rigidifies Solvent Molecules Surrounding DHFR**

The 13RT and 05RT structures have dramatically fewer ordered water molecules (red spheres) than the 13cryo and 05cryo structures (blue spheres; Table 1). NADP<sup>+</sup> and folate coloring as in Figure 1.

unfolded protein to the relatively dense network of a folded protein structure (Antal et al., 2009). Therefore, percolation theory may find fruitful application to other phase-change-like phenomena in proteins, such as the rigidification of protein and water atoms due to crystal cryocooling. In particular, CONTACT analysis of correlated side-chain motion closely resembles a bond percolation model of protein disorder and may be amenable to quantitative treatment using existing tools from percolation theory.

Reducing temperature from 273 K to 100 K shifts conformational distributions to favor low-entropy, low-enthalpy states (Halle, 2004). This effect is consistent with the switch at Arg52 from multiple rotamers and disordered solvent at RT to a single hydrogen-bonded rotamer and ordered solvent at cryogenic conditions. However, because of kinetic effects, data sets based on cryocooled crystals at 100 K do not reflect the Boltzmann ensemble at 100 K; rather, different degrees of freedom correspond

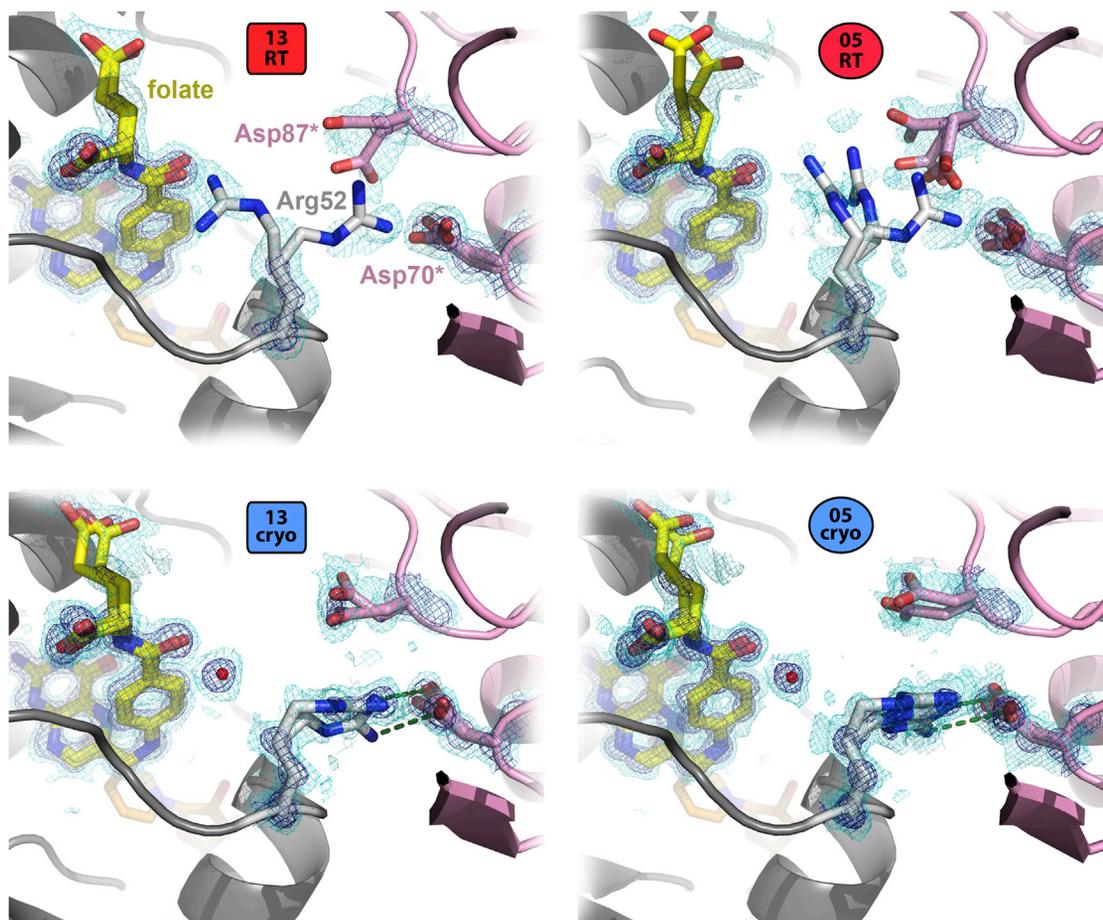
potentially be refined by methodological advances: improved side-chain and backbone sampling in qFit may reveal more alternate rotamers and remove false-positive results, and improved CONTACT analysis with ionic bonds, covalent strain, and consideration of water molecules may more accurately define conformational coupling between residues. Furthermore, independent methods that probe correlated motions, such as NMR (Sekhar and Kay, 2013; Szyperski and Mills, 2011) and MD (Brandman et al., 2012; McClendon et al., 2012), could be used as a complementary way to study these transitions across a wide temperature range.

In conjunction with empirical observations such as those described here, more powerful theoretical frameworks would be useful for elucidating the physical/material properties of protein molecules. Further parallels may be drawn between the “freezing out” of protein and water atoms and concomitant quenching of dynamic contact networks (van den Bedem et al., 2013) and conformational heterogeneity indicative of motions on the picosecond-to-nanosecond timescale (Fenwick et al., 2014). Percolation theory is widely used to describe phenomena in which connections are made between local sites until a continuous path is created that spans the entire system. Percolation models have proven useful for understanding the sudden transition from the very sparse “connection network” of an

to different effective temperatures (Halle, 2004). As a result, some of the populations we observe here could be as strongly influenced by kinetic trapping at the glass transition as by the location of free energy minima, confounding our ability to decipher their physical origin. Attempts have been made to anneal crystals to true minimum-free-energy low-temperature states with cycles of warming and cooling (Juers et al., 2007). However, it can be difficult to empirically identify a suitable cryoprotectant to cause bulk solvent to contract in lockstep with protein to reduce lattice contact remodeling (Juers and Matthews, 2001, 2004). It is also possible that the addition of cryoprotectant (both the 05cryo and 13cryo crystals were supplemented with PEG 400) is responsible for some of the observed effects in these crystals. In contrast, the RT data sets studied here are free from cryoprotectant- and cryocooling-induced distortions and show no evidence of unusual radiation damage. Continuing development of methods for predicting, detecting, and accounting for radiation damage (Diederichs et al., 2003; Garman and Weik, 2013; Southworth-Davies et al., 2007) should help facilitate future RT studies of other systems. The capabilities of X-ray free electron lasers are already allowing damage-free data collection at ambient temperatures and revealing distinct conformations and patterns of thermal motions relative to cryocooled data sets (Liu et al., 2013).

## Structure

### Crystal Cryocooling Alters DHFR Conformations

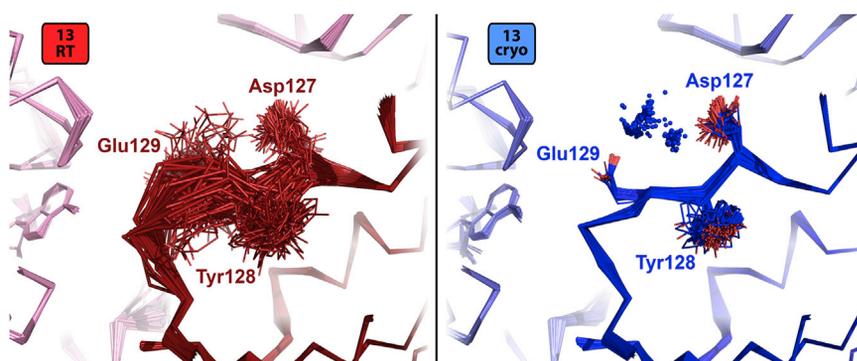


**Figure 6. Shifted Crystal Contacts in Cryogenic Structures Favor a New Arg52 Side-Chain Conformation**

In the 13RT and 05RT structures, Arg52 (gray, center) adopts rotamers that point toward folate (yellow) and the crystal lattice neighbor (pink) but are not close enough to be properly positioned for a hydrogen bond. In contrast, in both the 13cryo and 05cryo structures, the  $C_{\beta}$ - $C_{\beta}$  distance to the nearest lattice residue (Asp70\*) shrinks from 9.3 to 8.7 Å for the 2013 structures and from 9.0 to 8.7 Å for the 2005 structures. As a result, the lattice copy of Asp87 (Asp87\*) shifts upward in this view such that Arg52 can adopt a new rotamer that forms hydrogen bonds (green dotted lines) to the lattice copy of Asp70 (Asp70\*). A water molecule (red sphere) takes the place of the RT rotamer near the folate.  $2mF_o-DF_c$  electron density contoured at 1.9  $\sigma$  (dark blue) and 0.6  $\sigma$  (light blue) supports the modeled side chain and water positions in all four qFit structures.

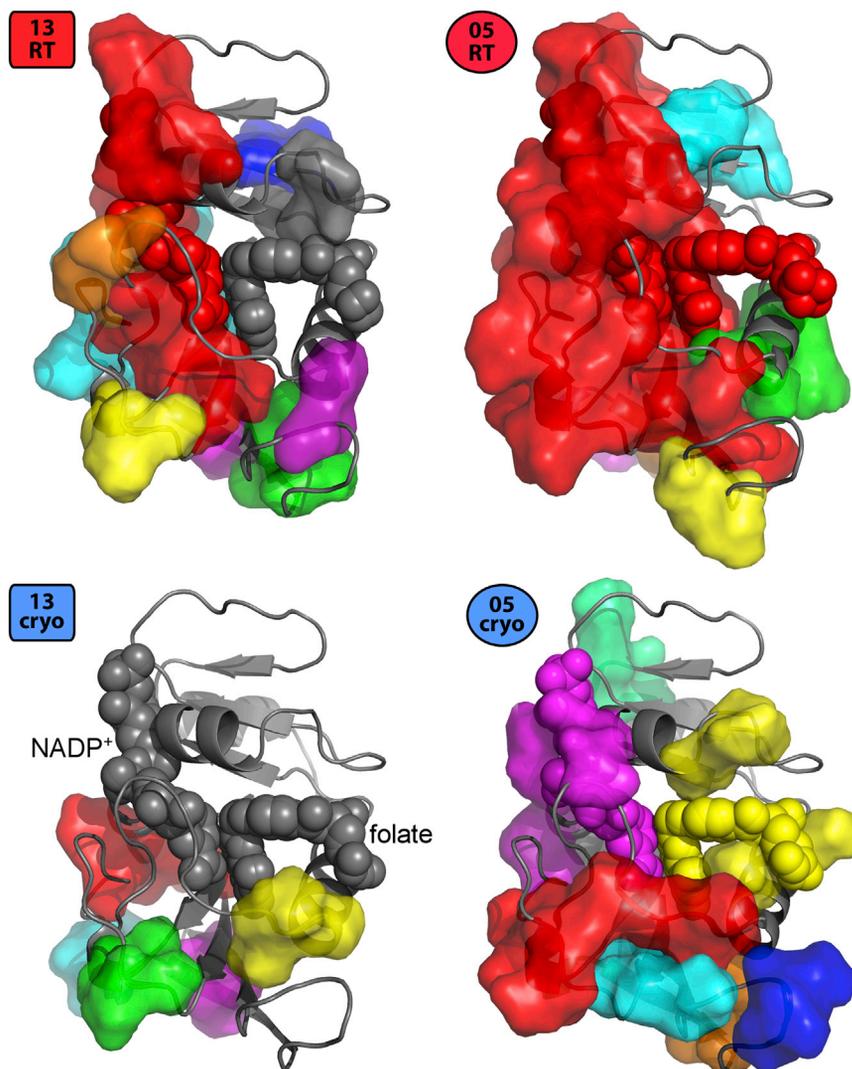
The qFit multiconformer and PHENIX ensemble models provide complementary ways to explore temperature-paired data sets. These two representations agree in suggesting that cryocooling significantly diminishes conformational heterogeneity and in pointing to water ordering as a major contributing factor.

Both methods are under active development and our results highlight several areas for future improvements. For example, qFit models fail to explore more than very subtle backbone flexibility. Ensembles, by contrast, can capture larger-scale loop flexibility, but suffer other problems: they are potentially



**Figure 7. Cryogenically Ordered Water Molecules Induce Ensemble Loop Ordering**

The surface-exposed loop encompassing Asp127, Tyr128, and Glu129 is very flexible in the lowest- $R_{free}$  13RT ensemble (red, left) because it is relatively unconstrained by contacts to crystal lattice mates (left and top in this view, light pink). In contrast, consistently modeled ordered waters (blue spheres) unique to the cryogenic ensemble (blue, right) bridge the polar Asp127 and Glu129 side chains via hydrogen bonds to stabilize a more unique loop conformation. This phenomenon is extremely similar for the 05RT and 05cryo models (not shown).



**Figure 8. Temperature Dependence of Coupled Conformational Heterogeneity Linking the Two Subdomains of DHFR**

The largest dynamic CONTACT network (red) from both the 13RT and 05RT qFit models spans the two subdomains (top versus bottom in this view) via one or both of the NADP<sup>+</sup> cofactor and folate substrate (spheres, labeled for 13cryo) in the active site (van den Bedem et al., 2013). In contrast, using the same  $T_{stress}$  values (see Experimental Procedures) as input to CONTACT, none of the networks (different colors) from the corresponding 13cryo and 05cryo models connect the two subdomains via the active site.

of biological mechanism from crystallographic data. We show here that the general trends of protein compression and water ordering are consistent, but that different crystals of the same protein can respond somewhat idiosyncratically. It may be possible to leverage our observations of these DHFR crystals to explore the feasibility of reversing cryocooling in silico for other systems by performing MD or Monte Carlo simulations with the crystal lattice present. “Warming up the PDB” in this way may shed new light on the details of temperature-dependent perturbations to first-shell solvent, protein surfaces, and lattice interfaces.

#### EXPERIMENTAL PROCEDURES

##### Protein Expression, Purification, Crystallization, Data Collection, and Initial Refinement

The 2013 data sets were taken from previous work (van den Bedem et al., 2013). Briefly, crystals for

both data sets were grown by the hanging drop method with a well solution consisting of 100 mM HEPES pH 7.5, 21% PEG 8000, and 200 mM MgCl<sub>2</sub>. Diffraction data were collected at the Advanced Light Source (ALS, Berkeley, CA), beamline 8.3.1. RT data were collected with the cryojet stream set to 273 K; there were no signs of radiation damage. Cryogenic data were collected after the addition of 10% PEG 400 to the mother liquor with the cryojet stream set to 100 K. Molecular replacement was performed with Phaser using the 1.8 Å resolution crystal structure of the *Escherichia coli* DHFR-folate-NADP<sup>+</sup> complex (PDB accession code 1RX2) determined by Sawaya and Kraut (Sawaya and Kraut, 1997); subsequent initial refinement was performed in the PHENIX GUI (Adams et al., 2010).

The 2005 data sets have not been previously reported. Briefly, *E. coli* DHFR from a previously described construct (Cameron and Benkovic, 1997) was expressed and purified (see Supplemental Experimental Procedures). Folate and NADP<sup>+</sup> were added, and crystals were grown by the hanging drop method with a well solution consisting of 17% PEG 400, 20 mM imidazole pH 7.0, and 125 mM MnCl<sub>2</sub>. For the 100 K data set, PEG 400 was added to a final concentration of 30% for cryoprotection. X-ray diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL, Palo Alto, CA), beamline 11-1. Further details are available in the Supplemental Experimental Procedures.

##### Multiconformer Model Building and Analysis

Using the initial 2013 and 2005 RT and cryo models, qFit was run to identify alternate conformations and construct all-atom, multiconformer models. For

overfit/underdetermined due to the large number of parameters relative to the smaller number of crystallographic observations, and methods are needed to elucidate the relationships between subsets of models within the ensemble (such as MutInf for longer MD trajectories; McClendon et al., 2012). Furthermore, both methods (as well as CONTACT network analysis) may be subject to the idiosyncrasies of particular combinations of input parameters. Exploring the model selection tradeoffs between the parsimony of multiconformer models and the conformational expressiveness of ensembles will be an important area of future research. Nevertheless, in this study, the multiconformer and ensemble representations in their current states of development paint a cohesive picture of how cryocooling perturbs protein energy landscapes.

Cryocooling protein crystals perturbs protein energy landscapes in ways that are difficult to predict at the detailed atomic level. In DHFR, this phenomenon alters the preferred conformations of side chains buttressing the substrate and suppresses a dynamic contact network bridging the adenine binding domain, cofactor, and active site. In other systems, cryocooling may also introduce obstacles to extracting a fundamental understanding

## Structure

### Crystal Cryocooling Alters DHFR Conformations

CellPress

the 2013 models, PHENIX version 1.8-1069 was used; for the 2005 models in this work, PHENIX 1.8.4-1496 was used. A few small adjustments were made to the models after qFit refinement. For the 13RT qFit model, one calcium atom was moved to a symmetry-related position to match a manganese position in the 2005 models. Then, for both 2013 models, all atoms (including water molecules) were relabeled chain A, and hydrogens were re-optimized using phenix.reduce. Although PHENIX by default now uses hydrogens based on electron-cloud centers instead of hydrogen nucleus positions, CONTACT (see below) is designed with the previously utilized nuclear positions, so we used nuclear hydrogen positions in our qFit models. See Table 1 for final model statistics. MolProbity 3 (Chen et al., 2010), including Probe (Word et al., 1999) with nuclear hydrogen positions (as opposed to electron-cloud-based positions), was used to calculate clashscores (Word et al., 1999), MolProbity scores (Chen et al., 2010), and crystal contact dots. Compared to the ensemble refinement, these models fit the data slightly less well (2%–3% worse  $R_{free}$ ) for three of the four cases here; the higher-resolution (2005) RT data set is the only one of the four for which the qFit model is in better fit to the diffraction data. The  $R_{free}$  values are sensitive to small changes; for example, changing the occupancies of metal ions in the 13cryo model from 100% as in the (previously published) 13RT model to 50% led to  $R_{free}$  changes on the order of 0.15–0.20. We therefore caution against overinterpretation of small  $R_{free}$  differences between our models. All qFit and ensemble models fit the data better than single-conformation models.

#### Calculation of Isomorphous Difference Density Maps

To generate  $F_o - F_c$  isomorphous difference density maps, we first used the “Find alternate-origin symmetry mate” utility in the PHENIX GUI to place the 13cryo and both 2005 models in the same frame of reference as the 13RT model. We then used the “Isomorphous difference map” utility in the PHENIX GUI to generate the  $F_o - F_c$  map coefficients in MTZ format.

#### Analysis of Changes to Flexibility of Individual Side Chains and Main-Chain Segments

Side-chain rotamers were computed using phenix.rotalyze. “Outlier” conformations that fall outside of well-populated “rotameric” regions of dihedral space often suggest difficulty in fitting a more common conformation, perhaps due to increased disorder, and may in some cases be valid conformations supported by local enthalpic interactions, so they were counted as different rotamers for the purposes of compiling these statistics. The statistical significance of colocalization of residues with altered rotamers was assessed by the probability of uncorrelated residues showing an overlap at least as large as that observed ( $p < 10^{-5}$  for both RT and cryo comparisons), as derived from a one-tailed Fisher’s exact test based on the hypergeometric distribution (Rivals et al., 2007).

Residue-level side-chain and main-chain rmsd were derived by computing an rmsd between the heavy atoms in a given residue for every pair of alternate combinations for that residue, then taking the maximum. This quantifies the degree of dissimilarity among conformations identified by qFit for a given residue, which we use here as a measure of flexibility.

#### CONTACT Analysis of Networks of Coupled Conformational Heterogeneity

We used all four qFit models as input to CONTACT (van den Bedem et al., 2013), which builds networks of coupled conformational heterogeneity. Briefly, if one conformation of residue 1 clashes with a conformation of a neighboring residue 2, a pathway is propagated from residue 1 to residue 2 if another conformation at residue 2 alleviates that clash. Pathways are continued in this way until a final clash is either relieved or cannot be relieved. Finally, these all-atom pathways are merged to create a smaller set of unique residue-level pathways; each set of residue-level pathways with at least partial overlaps of constituent residues is defined as a network. Details on parameter choices are available in the Supplemental Experimental Procedures.

#### Calculation and Analysis of PHENIX Ensembles

Time-averaged ensembles were generated with phenix.ensemble\_refinement in PHENIX version 1.8.4-1496. For the 2013 data sets, the input models were PDB codes 4KJJ (cryo) and 4KJK (RT), with alternate conformations removed. A grid search was performed by repeating the simulation with five values of

$p_{TLS}$  from 0.1 to 1.0, five values of  $w_{X-ray}$  from 2.5 to 10, and three values of  $\tau_x$  from 1.0 to 2.0; the ensemble with the lowest  $R_{free}$  for each data set was used for most analyses here.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, two tables, and three movies and can be found with this article online at <http://dx.doi.org/10.1016/j.str.2014.04.016>.

#### ACKNOWLEDGMENTS

This collaboration was catalyzed at a symposium convened by Tom Alber, doctoral student of G.A.P. and doctoral advisor to J.S.F. We dedicate this manuscript to Tom.

We thank Rahel Woldeyes and other Fraser lab members for helpful comments; Peter Wright and Gira Bhabha for discussions about DHFR; Tom Burnley for ensemble analysis scripts; James Holton, George Miegs, and Jane Tanamachi for technical support at ALS; and Nat Echols for PHENIX support. J.S.F. is supported by NIH OD009180, GM110580, and NSF STC-1231306. 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 a SLAC National Accelerator Laboratory LDRD (Laboratory Directed Research and Development) grant SLAC-LDRD-0014-13-2. D.A.S. is supported by NIH GM081879. G.A.P. and D.R. are supported by NIH GM32415. M.A.W. is supported by NIH GM092999.

Diffraction data were collected at the Advanced Light Source, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231, and at the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, which is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract no. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393).

Received: March 10, 2014

Revised: April 26, 2014

Accepted: April 29, 2014

Published: May 29, 2014

#### REFERENCES

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221.
- Antal, M.A., Böde, C., and Csermely, P. (2009). Perturbation waves in proteins and protein networks: applications of percolation and game theories in signaling and drug design. *Curr. Protein Pept. Sci.* 10, 161–172.
- Bhabha, G., Lee, J., Ekiert, D.C., Gam, J., Wilson, I.A., Dyson, H.J., Benkovic, S.J., and Wright, P.E. (2011). A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* 332, 234–238.
- Boehr, D.D., McElheny, D., Dyson, H.J., and Wright, P.E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* 313, 1638–1642.
- Boehr, D.D., Schnell, J.R., McElheny, D., Bae, S.H., Duggan, B.M., Benkovic, S.J., Dyson, H.J., and Wright, P.E. (2013). A distal mutation perturbs dynamic amino acid networks in dihydrofolate reductase. *Biochemistry* 52, 4605–4619.
- Brandman, R., Brandman, Y., and Pande, V.S. (2012). A-site residues move independently from P-site residues in all-atom molecular dynamics simulations of the 70S bacterial ribosome. *PLoS ONE* 7, e29377.
- Burnley, B.T., Afonine, P.V., Adams, P.D., and Gros, P. (2012). Modelling dynamics in protein crystal structures by ensemble refinement. *eLife* 1, e00311.

- Cameron, C.E., and Benkovic, S.J. (1997). 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.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21.
- Diederichs, K., McSweeney, S., and Ravelli, R.B. (2003). Zero-dose extrapolation as part of macromolecular synchrotron data reduction. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 903–909.
- Falzone, C.J., Wright, P.E., and Benkovic, S.J. (1994). Dynamics of a flexible loop in dihydrofolate reductase from *Escherichia coli* and its implication for catalysis. *Biochemistry* **33**, 439–442.
- Fenwick, R.B., van den Bedem, H., Fraser, J.S., and Wright, P.E. (2014). Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. *Proc. Natl. Acad. Sci. USA* **111**, E445–E454.
- Fraser, J.S., Clarkson, M.W., Degnan, S.C., Erion, R., Kern, D., and Alber, T. (2009). Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673.
- Fraser, J.S., van den Bedem, H., Samelson, A.J., Lang, P.T., Holton, J.M., Echols, N., and Alber, T. (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. USA* **108**, 16247–16252.
- Frauenfelder, H., Petsko, G.A., and Tsernoglou, D. (1979). Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* **280**, 558–563.
- Frauenfelder, H., Hartmann, H., Karplus, M., Kuntz, I.D., Jr., Kuriyan, J., Parak, F., Petsko, G.A., Ringe, D., Tilton, R.F., Jr., Connolly, M.L., et al. (1987). Thermal expansion of a protein. *Biochemistry* **26**, 254–261.
- Garman, E.F., and Weik, M. (2013). Radiation damage to biological macromolecules: some answers and more questions. *J. Synchrotron Radiat.* **20**, 1–6.
- Halabi, N., Rivoire, O., Leibler, S., and Ranganathan, R. (2009). Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138**, 774–786.
- Halle, B. (2004). Biomolecular cryocrystallography: structural changes during flash-cooling. *Proc. Natl. Acad. Sci. USA* **101**, 4793–4798.
- Juers, D.H., and Matthews, B.W. (2001). Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. *J. Mol. Biol.* **311**, 851–862.
- Juers, D.H., and Matthews, B.W. (2004). Cryo-cooling in macromolecular crystallography: advantages, disadvantages and optimization. *Q. Rev. Biophys.* **37**, 105–119.
- Juers, D.H., Lovelace, J., Bellamy, H.D., Snell, E.H., Matthews, B.W., and Borgstahl, G.E. (2007). Changes to crystals of *Escherichia coli* beta-galactosidase during room-temperature/low-temperature cycling and their relation to cryo-annealing. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 1139–1153.
- Kuriyan, J., Petsko, G.A., Levy, R.M., and Karplus, M. (1986). Effect of anisotropy and anharmonicity on protein crystallographic refinement. An evaluation by molecular dynamics. *J. Mol. Biol.* **190**, 227–254.
- Lang, P.T., Ng, H.L., Fraser, J.S., Corn, J.E., Echols, N., Sales, M., Holton, J.M., and Alber, T. (2010). Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci.* **19**, 1420–1431.
- Lang, P.T., Holton, J.M., Fraser, J.S., and Alber, T. (2014). Protein structural ensembles are revealed by redefining X-ray electron density noise. *Proc. Natl. Acad. Sci. USA* **111**, 237–242.
- Liu, W., Wacker, D., Gati, C., Han, G.W., James, D., Wang, D., Nelson, G., Weierstall, U., Katritch, V., Barty, A., et al. (2013). Serial femtosecond crystallography of G protein-coupled receptors. *Science* **342**, 1521–1524.
- Lorber, D.M., and Shoichet, B.K. (1998). Flexible ligand docking using conformational ensembles. *Protein Sci.* **7**, 938–950.
- Lovell, S.C., Word, J.M., Richardson, J.S., and Richardson, D.C. (2000). The penultimate rotamer library. *Proteins* **40**, 389–408.
- Mandell, D.J., and Kortemme, T. (2009). Backbone flexibility in computational protein design. *Curr. Opin. Biotechnol.* **20**, 420–428.
- McClendon, C.L., Hua, L., Barreiro, A., and Jacobson, M.P. (2012). Comparing conformational ensembles using the Kullback-Leibler divergence expansion. *J. Chem. Theory Comput.* **8**, 2115–2126.
- Rasmussen, B.F., Stock, A.M., Ringe, D., and Petsko, G.A. (1992). Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature* **357**, 423–424.
- Rivals, I., Personnaz, L., Taing, L., and Potier, M.C. (2007). Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics* **23**, 401–407.
- Sawaya, M.R., and Kraut, J. (1997). Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. *Biochemistry* **36**, 586–603.
- Sekhar, A., and Kay, L.E. (2013). NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl. Acad. Sci. USA* **110**, 12867–12874.
- Smith, J.L., Hendrickson, W.A., Honzatko, R.B., and Sheriff, S. (1986). Structural heterogeneity in protein crystals. *Biochemistry* **25**, 5018–5027.
- Southworth-Davies, R.J., Medina, M.A., Carmichael, I., and Garman, E.F. (2007). Observation of decreased radiation damage at higher dose rates in room temperature protein crystallography. *Structure* **15**, 1531–1541.
- Szyperki, T., and Mills, J.L. (2011). NMR-based structural biology of proteins in supercooled water. *J. Struct. Funct. Genomics* **12**, 1–7.
- Tilton, R.F., Jr., Dewan, J.C., and Petsko, G.A. (1992). Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K. *Biochemistry* **31**, 2469–2481.
- Tyka, M.D., Keedy, D.A., André, I., Dimaio, F., Song, Y., Richardson, D.C., Richardson, J.S., and Baker, D. (2011). Alternate states of proteins revealed by detailed energy landscape mapping. *J. Mol. Biol.* **405**, 607–618.
- van den Bedem, H., Dhanik, A., Latombe, J.C., and Deacon, A.M. (2009). Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1107–1117.
- van den Bedem, H., Bhabha, G., Yang, K., Wright, P.E., and Fraser, J.S. (2013). Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat. Methods* **10**, 896–902.
- Vitkup, D., Ringe, D., Petsko, G.A., and Karplus, M. (2000). Solvent mobility and the protein 'glass' transition. *Nat. Struct. Biol.* **7**, 34–38.
- Word, J.M., Lovell, S.C., LaBean, T.H., Taylor, H.C., Zalis, M.E., Presley, B.K., Richardson, J.S., and Richardson, D.C. (1999). Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* **285**, 1711–1733.

**Structure, Volume 22**

**Supplemental Information**

**Crystal Cryocooling Distorts**

**Conformational Heterogeneity**

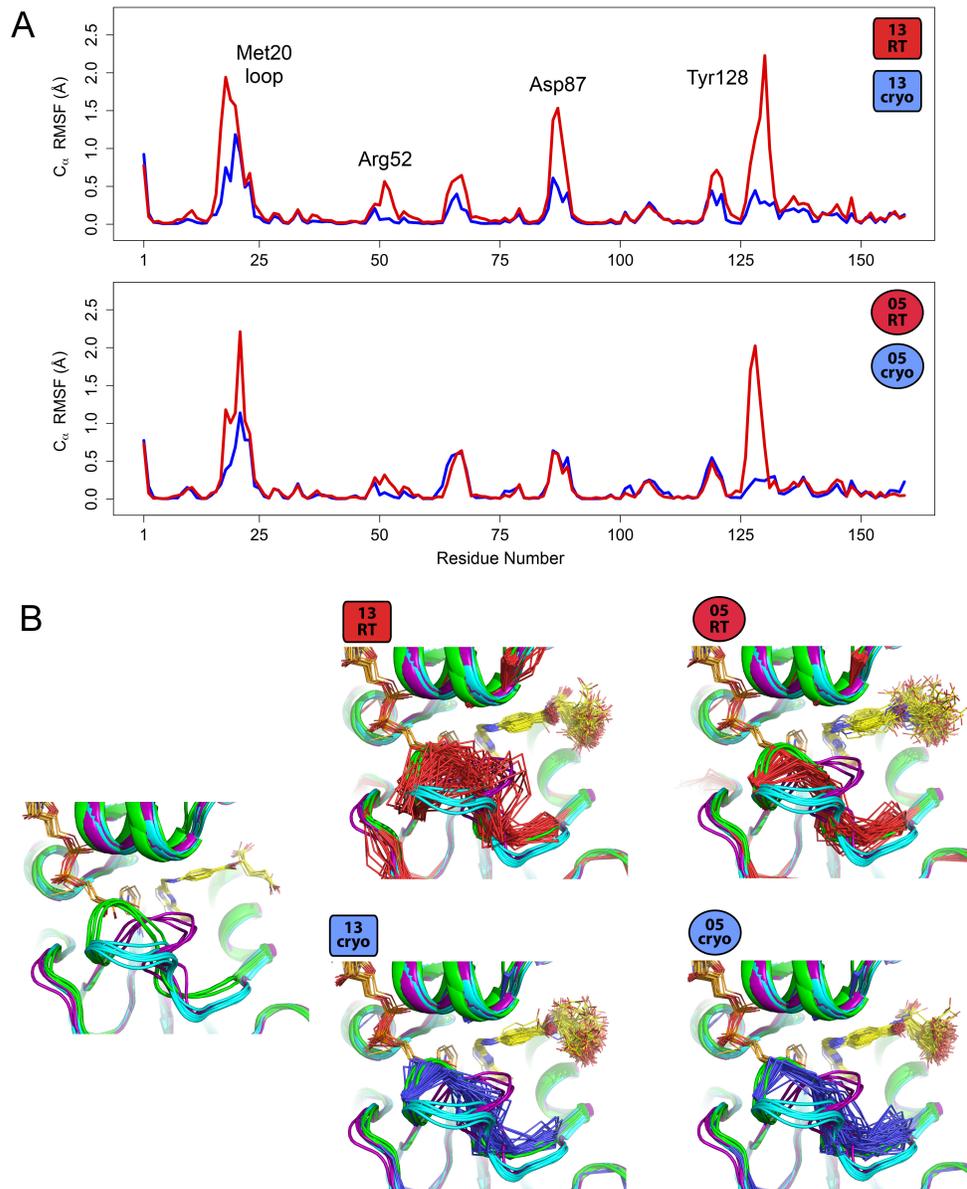
**in a Model Michaelis Complex of DHFR**

**Daniel A. Keedy, Henry van den Bedem, David A. Sivak, Gregory A. Petsko, Dagmar Ringe, Mark A. Wilson, and James S. Fraser**

## Supplemental Data

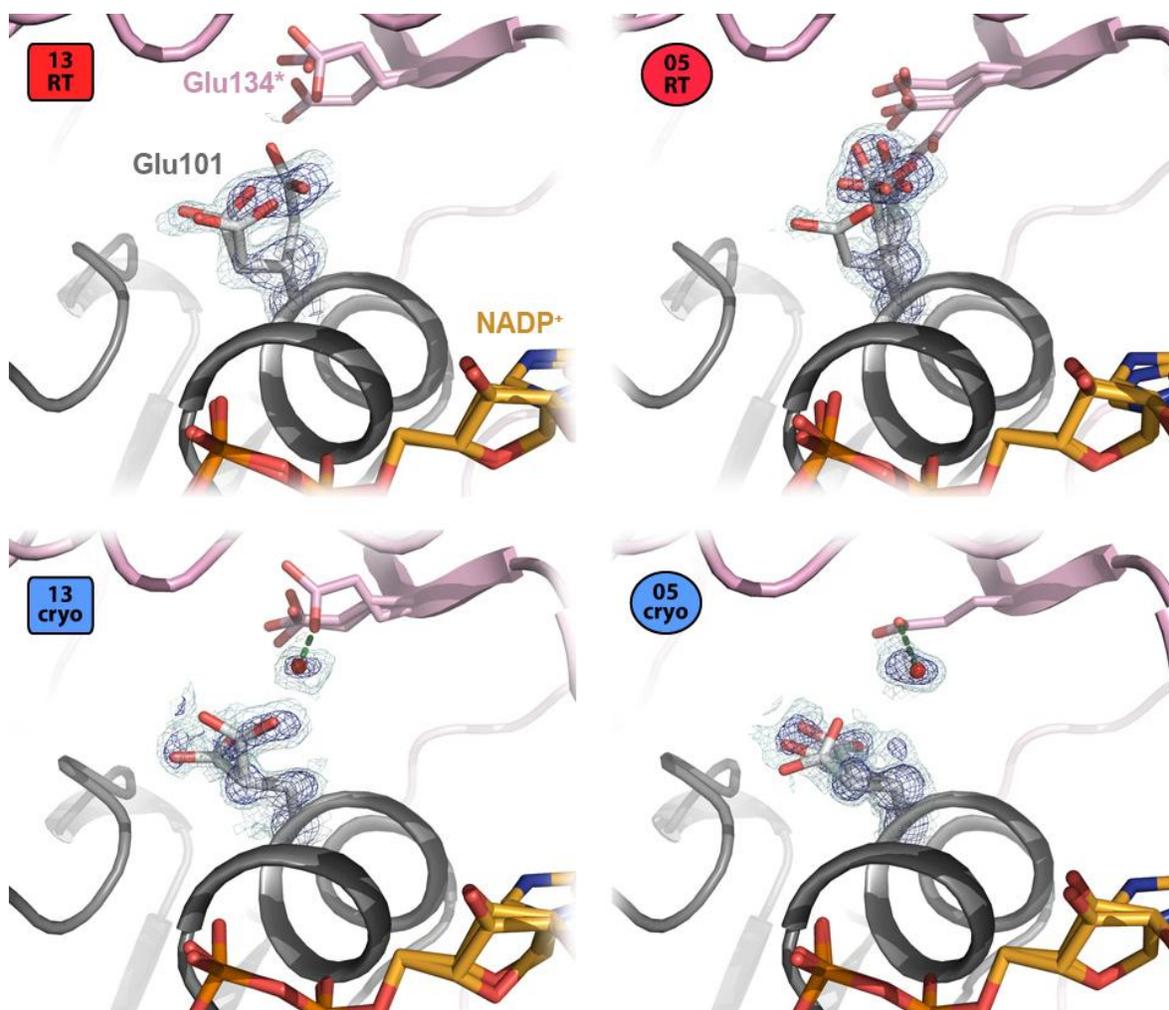
“Crystal cryocooling distorts conformational heterogeneity in a model Michaelis complex of DHFR”

Daniel A. Keedy, Henry van den Bedem, David A. Sivak, Gregory A. Petsko, Dagmar Ringe, Mark A. Wilson, James S. Fraser



**Figure S1, related to Figure 4: (A)** Room-temperature ensembles are consistently more flexible than cryogenic ensembles at several surface loops.  $C_{\alpha}$  RMSF values for the lowest- $R_{free}$  ensemble for each dataset are higher for each room-temperature ensemble (red lines) than the corresponding cryogenic ensemble (blue lines) at several regions (labeled). For example, the 13RT ensemble appears additionally flexible relative to the 05RT ensemble around Asp87. However, Asp87 has one of the most protruding

sidechains in DHFR, and sits amidst a loop that contacts multiple nearby loops in crystal contacts – including a lattice copy of the Met20 loop, which is known to transition between distinct conformations to enable catalysis (Sawaya and Kraut, 1997) (**Figure S2**). The spatial proximity of these variable loops in the crystal seems to make it difficult to determine a consensus ensemble of conformations for the Asp87 loop; the dramatically increased flexibility in the 13RT ensemble in particular may be peculiar to the set of input parameters that led to the lowest  $R_{free}$  for this dataset (**Table 2**). Interestingly, the Asp87 13RT-13cryo difference is diminished and becomes more in line with the 05RT-05cryo difference in the RMSF traces averaged over multiple ensembles with different input parameters (**Figure 4**). This example highlights the balance between selecting a single “best” ensemble and incorporating some amount of information about the contribution of simulation input parameters to the contents of output ensembles, which should be a topic of future study. (**B**) The Met20 loop stays near the closed (green) state, not the occluded (purple) or open (cyan) states, in all four lowest- $R_{free}$  ensembles. Examples of closed, occluded, and cyan states (top) are taken from a thorough crystallographic study (Sawaya and Kraut, 1997). Coloring for NADP<sup>+</sup> and folate as in **Figure 1**.



**Figure S2, related to Figure 6:** Frozen water in cryogenic structures diminishes Glu101 sidechain conformational heterogeneity. Both the 13RT and 05RT structures adopt a  $\chi_1 t$  rotamer, which places the Glu101 sidechain (gray) pointing directly upward in this view. In the respective cryogenic structures, however, a new ordered water molecule (red sphere; B-factors less than  $30 \text{ \AA}^2$  in both cryogenic structures) fills the volume previously occupied by that rotamer, and Glu101 instead adopts exclusively  $\chi_1 m$  rotamers. The new water position is stabilized by a hydrogen bond (green dotted line) that is made possible by encroachment by Glu134\* of a crystal lattice mate (pink). Glu101 does not make direct contact to NADP<sup>+</sup> (orange), but it is located nearby, on the helix that buttresses the cofactor in the active site.  $2mF_o - DF_c$  electron density contoured at  $1.2 \sigma$  (dark blue) and  $0.5 \sigma$  (light blue) supports the modeled sidechain and water positions in all four qFit structures.

	2013	2005
Total protein residues	159	159
Residues with same single rotamer (or Gly/Ala)	112	110
Residues with same set of multiple rotamers	11	8
Residues with rotamer(s) unique to room temperature structure	30	38
Residues with rotamer(s) unique to cryogenic structure	15	16
Residues with more total rotamers in room-temperature structure	22	27
Residues with more total rotamers in cryogenic structure	7	5
Residues with higher sidechain RMSD in room-temperature structure	102	98
Residues with higher sidechain RMSD in cryogenic structure	30	35

**Table S1, related to Figure 2:** Overall differences in local sidechain conformational heterogeneity between room-temperature and cryogenic structures.

	13RT	13cryo	05RT	05cryo
$T_{stress}$ used for CONTACT	0.28	0.28	0.33	0.33
Number of pathways	131	40	375	39
Average pathway length	3.4	2.7	5.1	3.1

**Table S2, related to Figure 8:** Statistics from CONTACT network analysis of qFit models. “Number of pathways” is the count of unique residue-level pathways. “Average pathway length” is the average length of all identified atom-level pathways.

**Movie S1, related to Figure 5:** Movie illustrating unit cell shrinking and water ordering for the 2005 dataset pair in atomic detail.

**Movie S2, related to Figure 5:** Movie illustrating unit cell shrinking and water ordering for the 2013 dataset pair in atomic detail.

**Movie S3, related to Figure 5:** Movie in cartoon depiction illustrating the prevalence of lower-entropy, lower-enthalpy states at lower temperatures vs. higher-entropy, higher-enthalpy states at higher temperatures.

## Supplemental Experimental Procedures

### *Protein Expression, Purification, Crystallization, Data Collection, and Initial Refinement*

Recombinant *E. coli* DHFR was overexpressed in the *E. coli* strain BL21(DE3) (Novagen) using a pET22b construct described previously (Cameron and Benkovic, 1997). Cells were grown in LB broth supplemented with 100 µg/mL ampicillin with shaking at 37°C to an OD<sub>600</sub> of 0.7 and then induced by the addition of IPTG to a final concentration of 0.4 mM. Induced cells were incubated at with shaking at room temperature overnight and then harvested by centrifugation, snap-frozen on liquid nitrogen, and stored at -80°C.

Frozen bacterial cell pellets were resuspended on ice in 25 mM HEPES pH 7.5, 1 mM EDTA, 1mM DTT (stabilization buffer) and lysed by the addition of lysozyme to 1 mg/mL followed by brief sonication. Cell debris was removed by centrifugation at 10,000xg and nucleic acid was precipitated from the supernatant by the slow addition of streptomycin sulfate to a final concentration of 10 mg/mL. Precipitated nucleic acid was removed by centrifugation at 10,000xg for 30 minutes and the supernatant was dialysed overnight against stabilization buffer at 4°C. The clarified supernatant was applied to a column packed with methotrexate-agarose DHFR affinity resin (Sigma), and the column was washed with 50 column volumes of stabilization buffer, 50 column volumes stabilization buffer supplemented with 1 M NaCl, and a final wash with 50 column volumes of stabilization buffer. Bound DHFR was eluted from the affinity column with 200 mM sodium borate pH 9.0, 3 mM folic acid, 0.5 M NaCl, 1 mM EDTA, 1mM DTT (elution buffer). At this stage, the DHFR is highly purified (i.e. there are no detectable protein contaminants on overloaded Coomassie blue-stained SDS-PAGE); however, the DHFR is bound both to folic acid and also to any remaining endogenous nucleotide cofactor. To remove these low-molecular weight anionic contaminants, the sample was dialyzed overnight against stabilization buffer, applied to a DEAE-sephacel anion exchange column, and eluted in a 20-column volume linear gradient from 0-0.5 M KCl in 10 mM Tris-HCl, 1 mM EDTA, 1 mM DTT. Fractions containing purified DHFR were pooled and protein concentration was determined by absorbance at 280 nm ( $\epsilon = 33,690 \text{ M}^{-1} \text{ cm}^{-1}$ ). Protein was bound to folic acid by overnight dialysis against crystallization buffer (20 mM imidazole pH 7.0) supplemented with 1 mM folic acid at 4°C and protected from light.

The folate-bound protein was centrifugally concentrated to 35 mg/mL for crystallization and stored at 4°C.

For the 2005 datasets, immediately prior to crystallization, freshly dissolved NADP<sup>+</sup> was added to aliquots of the DHFR-folate complex to a final concentration of 6 mM. The sample was centrifuged at 16,000xg and the supernatant was used to grow crystals using the hanging drop vapor diffusion method at 4°C by mixing 2 µL of DHFR-folate-NADP<sup>+</sup> with 2 µL of a reservoir solution containing 17% PEG 400, 20 mM imidazole pH 7.0, 125 mM MnCl<sub>2</sub>. In order to maintain the integrity of the light-sensitive folic acid, the crystallization experiments were set up in semi-dark and protected from light during crystal growth. In addition, all crystal manipulation was performed at 4°C and in the semi-dark. Blade-shaped yellow crystals measuring 0.8 mm x 0.4 mm x 0.4 mm appeared after 1-2 weeks. For the 100 K dataset, a crystal was removed from the drop with a nylon loop and serially transferred through increasing concentrations of PEG 400 to a final concentration of 30% in 20 mM imidazole pH 7.0, 125 mM MnCl<sub>2</sub>. The crystal was allowed to equilibrate in each condition for 1 minute and was then cooled by direct immersion into liquid nitrogen after equilibration for five minutes in the final condition. For the 277 K dataset, several crystals were removed from the mother liquor and placed into reservoir solution (17% PEG 400, 20 mM imidazole pH 7.0, 125 mM MnCl<sub>2</sub>) from which they were mounted into 0.7-1.0 mm quartz capillary tubes (Hampton Research). A plug of reservoir solution was included in the capillary in order to maintain vapor equilibrium with the crystal and the tubes were sealed with mineral oil and beeswax.

X-ray diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL, Palo Alto, CA), beamline 11-1. For the 100 K dataset, the diffraction data were measured from a single crystal maintained at 100 K and illuminated with 13.776 KeV incident radiation. Because these rod-shaped crystals tend to align with the b-axis parallel to the spindle axis, thereby creating a significant blind region, data were collected with two crystal orientations using a 90° arc goniometer attachment in order to record reflections that would otherwise have been lost in the blind region. Data were collected in both high- (2.0-0.85 Å) and low-resolution (75-1.5 Å) passes with different exposure times, crystal-to-detector distances, and oscillation ranges in order to accurately measure both the weak high-resolution and the intense low-resolution data. For the 277 K dataset, capillary-mounted crystals were maintained at 277 K using a cryostream and data were collected in 45° wedges with 0.5° oscillation/frame from multiple crystals in order to minimize radiation damage. For the largest crystals, multiple wedges of data were collected from the same sample by translating the rod-shaped crystal in the beam such that a fresh volume of the crystal was irradiated on subsequent exposures. Separate high- and low-resolution passes were collected from a total of three crystals at 277 K. All diffraction data were indexed, integrated, and scaled in HKL2000 with data statistics provided in Table 1. As with the 2013 datasets (van den Bedem et al., 2013), the  $R_{\text{merge}}$  values indicate no unusual radiation damage.

For the 2005 datasets, 1RX2 was also used as a starting point in the refinement of the *E*.

*coli* DHFR-folate-NADP<sup>+</sup> complex. Initial rigid body refinement using data to 2.5 Å resolution was performed in SHELX-97 (Sheldrick and Schneider, 1997) using the AFIX 6 instruction for all protein, folate, and NADP<sup>+</sup> atoms. Additional rounds of stereochemically restrained refinement of coordinates and individual atomic isotropic displacement parameters against an intensity-based target were performed with a stepwise improvement in resolution (STIR) to 1.5 Å resolution. Automated water picking was performed using SHELXWAT and supplemented by manual editing of the solvent model. An additional stepwise improvement in resolution instruction was used to gradually include all data (0.85 Å resolution for the 100 K dataset, 1.0 Å resolution for the 277 K dataset) during further cycles of refinement. Stereochemical restraints were used for all protein and cofactor atoms at every stage of the refinement. Upon convergence of the isotropic atomic displacement parameter (ADP) model, individual anisotropic ADPs were refined, decreasing the  $R$  and  $R_{free}$  values by 4%. The anisotropic ADPs were restrained to have similar magnitudes for bonded atoms (SIMU) and minimum differences in their projections onto the bond joining them (DELU). For solvent atoms, an isotropy restraint (ISOR) was also used. Subsequent cycles of refinement included the modeling of residues in alternate conformations and the introduction of an ambiguously oxidized cysteine at residue 152 that was modeled as a cysteine sulfinic acid in dual conformations. In the later cycles of refinement, riding hydrogen atoms were added except for the O atom of tyrosine, the Nε1 and Nδ2 atoms of histidine, the Oy atom of serine, and the Oy2 atom of threonine, because the placement of hydrogens for these atoms is ambiguous when based solely on geometric considerations. In the final cycles of refinement, the test set (5% of the reflections that had been sequestered for the calculation of  $R_{free}$ ) and the working set were combined and the model refined for an additional 10 cycles.

### *CONTACT Analysis of Networks of Coupled Conformational Heterogeneity*

Clashes in CONTACT are defined by the  $T_{stress}$  parameter, which is expressed as a percentile of all van der Waals overlap in a given model; any overlap worse than this threshold is considered a clash. Previous work found  $T_{stress}$  of 30% to be a useful consensus choice, but suggested the optimal  $T_{stress}$  for exploring network properties in a given protein may vary (van den Bedem et al., 2013). We therefore varied  $T_{stress}$  from 25% (top 25% of clashes) to 35% in 1% increments in this work. The lowest  $T_{stress}$  values at which the known inter-subdomain network was formed were 28% for the 13RT model and 33% for the 05RT model; we used these values to analyze the 13cryo and 05cryo models, respectively. Reliefs of clashes are defined by the relief value, which equals the fraction of the van der Waals overlap for each clash that must be alleviated for that clash to be considered relieved. Here we used the standard relief value from previous work (0.90), though the effect of varying this parameter could also be explored in future studies (van den Bedem et al., 2013).

## **Supplemental References**

Cameron, C.E., and Benkovic, S.J. (1997). 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.

Sheldrick, G.M., and Schneider, T.R. (1997). SHELXL: high-resolution refinement. *Methods in enzymology* 277, 319-343.

van den Bedem, H., Dhanik, A., Latombe, J.C., and Deacon, A.M. (2009). Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. *Acta crystallographica Section D, Biological crystallography* 65, 1107-1117.