

Ligand Binding and Circular Permutation Modify Residue Interaction Network in DHFR

Zengjian Hu¹, Donnell Bowen¹, William M. Southerland^{1*}, Antonio del Sol², Yongping Pan³, Ruth Nussinov^{3,4}, Buyong Ma^{3*}

1 Department of Biochemistry and Molecular Biology, Howard University College of Medicine, Washington, District of Columbia, United States of America, **2** Bioinformatics Research Project, Research and Development Division, Fujirebio, Tokyo, Japan, **3** Basic Research Program, Science Applications International Corporation—Frederick, Inc., Center for Cancer Research Nanobiology Program, National Cancer Institute—Frederick, Frederick, Maryland, United States of America, **4** Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel

Residue interaction networks and loop motions are important for catalysis in dihydrofolate reductase (DHFR). Here, we investigate the effects of ligand binding and chain connectivity on network communication in DHFR. We carry out systematic network analysis and molecular dynamics simulations of the native DHFR and 19 of its circularly permuted variants by breaking the chain connections in ten folding element regions and in nine nonfolding element regions as observed by experiment. Our studies suggest that chain cleavage in folding element areas may deactivate DHFR due to large perturbations in the network properties near the active site. The protein active site is near or coincides with residues through which the shortest paths in the residue interaction network tend to go. Further, our network analysis reveals that ligand binding has “network-bridging effects” on the DHFR structure. Our results suggest that ligand binding leads to a modification, with most of the interaction networks now passing through the cofactor, shortening the average shortest path. Ligand binding at the active site has profound effects on the network centrality, especially the closeness.

Citation: Hu Z, Bowen D, Southerland WM, del Sol A, Pan Y, et al. (2007) Ligand binding and circular permutation modify residue interaction network in DHFR. *PLoS Comput Biol* 3(6): e117. doi:10.1371/journal.pcbi.0030117

Introduction

Extensive experimental studies of dihydrofolate reductase (DHFR) have provided rich data toward the structure–function relationship in proteins. *Escherichia coli* DHFR is a 159–amino acid, monomeric, two-domain protein that is well characterized in terms of structure and function. DHFR catalyzes the reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydro-folate (THF) using the reducing cofactor nicotinamide adenine dinucleotide phosphate (NADPH). DHFR is a clinically important enzyme and is the target of a number of antifolate drugs.

Experimental kinetic analysis of various DHFR permutants has identified several loop regions important for catalysis [1,2]. Figure 1 illustrates the loop locations. The Met-20 loop (residues 10 to 23) directly controls the ligand binding to DHFR. The FG loop (residues 116 to 121) is behind the Met-20 loop. There is a network of hydrogen bonds connecting the Met-20 loop and the FG loop. The third GH loop (residues 142 to 149) is in contact with both the Met-20 and the FG loops. Based largely on the conformation of the Met-20 loop [3], the three states of the enzymatic reaction process (binding and release of cofactor, substrate, and product) can be defined using available crystal structures (Figure 1). In the open state (Figure 1, green ribbon) the Met-20 loop is flipped away from the binding site. In the closed state (Figure 1, blue), the Met-20 loop packs against the cofactor and seals the active site. In the occluded state, the Met-20 loop blocks the binding of the cofactor in the pocket. Simulations of the closed state also indicate changes in the other side of the

binding pocket, in the helix region (residues 44 to 50), which binds the cofactor. Loop region 64–71, which contacts the helix, also presents large fluctuations.

The cooperative movements of these loops couple with the overall dynamics of the protein concerning the binding and dissociation of the cofactor, substrate, and product. The communications among the various parts of the DHFR can be achieved by residue interaction networks and through peptide backbone chain connections and nonbonded residue interactions. Agarwal et al. carried out genomic analysis of sequence conservation, kinetic measurements of multiple mutations, and theoretical calculations, observing that non-bonded residue interactions in DHFR form a network of coupled motions that are important for enzyme catalysis [4].

The effect of the peptide backbone chain connection on the protein dynamics is more complex, since chain con-

Editor: Saraswathi Vishveshwara, Indian Institute of Science, India

Received: August 22, 2006; **Accepted:** May 11, 2007; **Published:** June 15, 2007

A previous version of this article appeared as an Early Online Release on May 11, 2007 (doi:10.1371/journal.pcbi.0030117.eor).

This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.

Abbreviations: DHFR, dihydrofolate reductase; MD, molecular dynamics; RMSD, root mean square deviation; RMSF, root mean square fluctuation; wild-ave, average structure of wild-type DHFR

* To whom correspondence should be addressed. E-mail: wsoutherland@howard.edu (WMS); mab@ncifcrf.gov (BM)

Author Summary

The cooperative movements within a protein concerning the binding and dissociation of the reactants and products could be important for protein function. Communication among the various parts of an enzyme can be achieved by the networks connecting amino acids through peptide backbone connections and non-bonded amino acid contact. We used dihydrofolate reductase (DHFR), a clinically important enzyme, as an example to explore the effects of amino acid communication on protein functions. We found that the peptide chain itself is an efficient “telephone wire” to transfer the communications. Breaking the telephone wire (peptide chain) at different points leads to differentiated behavior near the enzyme active site. The important points to keep the peptide chain communication are coupled with the place where protein folding occurs. On the other hand, ligand binding to the enzyme active site provides a “short cut” to the communication networks, with most of the interaction networks now passing through the added ligand and shortening the average communication path. We considered the short cuts to be “network-bridging effects” in the protein structure. The enzyme active site is the place where the short cut has the most dramatic effect in modifying protein communication networks.

nection is coupled with protein folding. Circular permutations of DHFR provide insight into chain connectivity, stability of the fold, and function. Circular permutation of a protein consists of connecting the native N- and C-termini covalently with a peptide linker and cleaving the peptide backbone at another specific site. Iwakura et al. have performed systematic circular permutation of the entire DHFR protein to investigate essential folding elements [5–8]. Other groups [9–11] have also circularly permuted the protein, selectively focusing on several permutations or cutting the backbone connection [11] to test fragment complementation. It was found that the peptide bonds in the protein could be grouped into two categories based on

the effects of breaking the backbone connectivity. While cleavage of some peptide bonds results in less active variants or affects the enzyme function only slightly, suggesting that these make only minor contributions to the ability of the protein to fold, cleavage at certain other positions leads to a complete loss of the ability of the protein to fold. When such cleavage sites occurred sequentially in the primary sequence and formed a contiguous peptide segment, the region was named a “folding element,” which is crucial for a protein to be foldable [5–8]. Folding elements distribute throughout the sequence. It was proposed that a complete set of folding elements is necessary for a protein to fold [5]. By conducting a systematic circular permutation analysis in which the original N- and C-termini of a protein are connected by an appropriate linker and new termini are created sequentially, ten folding elements have been identified in *E. coli* DHFR, each of which contains two to 14 residues [5–8] (Figure 2). It was also found that although the positions of the folding elements do not appear to correspond to the secondary structure motifs or to binding sites, almost all of the amino acid residues known to be involved in early folding events of DHFR are located within the folding elements, suggesting a close relationship between the folding elements of a protein and early folding events.

In order to delineate the complex relationship among chain connectivity, protein folding, coupled networks, and catalysis by DHFR, we have carried out a systematic network analysis and molecular dynamics (MD) simulations of the native (closed state) DHFR and 19 of its circularly permuted variants. We first obtained average protein structures from MD simulations of the native DHFR and of the circularly permuted mutants. We investigated the relationship between small-world network behavior and chain connectivity using the crystal and MD average protein structures. Small-world network analysis of the protein structure uses graph theory to explore the bonded and nonbonded amino acid residue

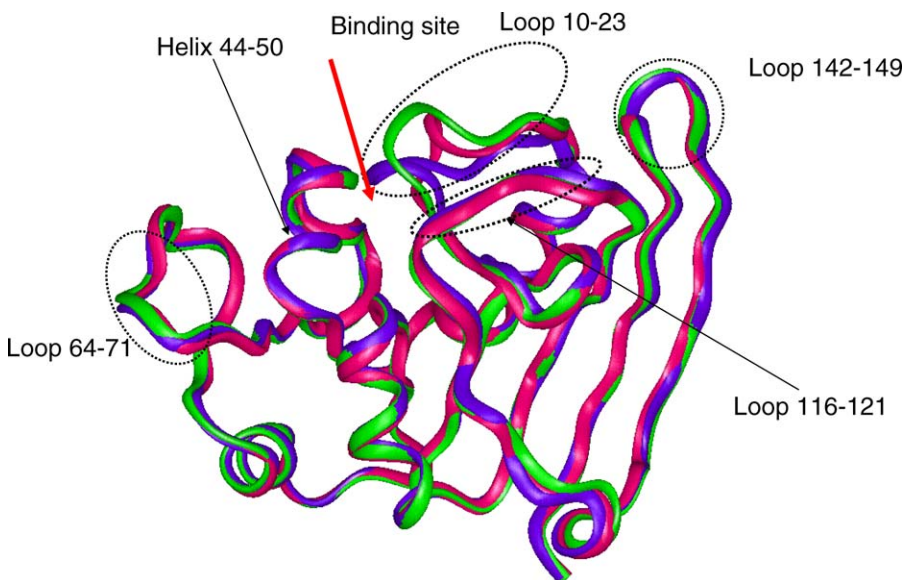


Figure 1. Illustration and Superposition of the Three States of DHFR

Red indicates occluded; green, open; and blue, closed. Five regions (loop10–23, helix 40–45, loop 64–71, loop116–121, and loop 142–149) near the active sites are highlighted.

doi:10.1371/journal.pcbi.0030117.g001

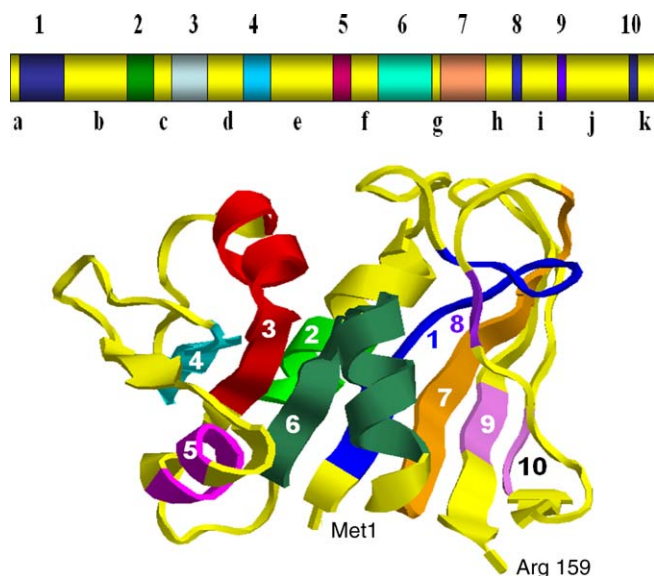


Figure 2. A Drawing of the Tertiary Structure of *E. coli* DHFR Taken from the Protein Data Bank and the Locations of the Proposed Folding Elements [5]

Four α -helices and eight β -strands are shown as ribbons. Proposed folding elements are light blue (1: Ser3~Ile14), green (2: Trp30~Leu36), red (3: Val40~Ser49), cyan (4: Arg57~Ser63), magenta (5: Glu80~Ala83), greenish-blue (6: Ile91~Leu104), orange (7: Ala107~Glu118), purple (8: His124~Phe125), and light purple (9: Val136~Ser138; and 10: Glu154~Ile155), respectively. The regions outside the folding elements form amino acid segments a–k (in yellow). The positions of the N- and C-termini are also indicated. Also see Table 1. doi:10.1371/journal.pcbi.0030117.g002

network. It was first used to identify key residues in protein folding, as these residues have high connectivity (betweenness) with respect to all possible network connections in the transition states of the protein structures [12]. The concept has been extended to the protein-folding process [13,14], the protein-protein interface [15], protein structure [16] and stability [17], protein dynamics [18], and key functional residues in enzymes [19]. By comparing the centrality of the residue interaction network, we found that ligand binding at the functional site has profound effects on the global network connections. Using the network connectivity index to distinguish between bonded and nonbonded connections, we found that breaking the chain connection in folding elements has a greater effect on active site loops than does breaking the chain in nonfolding regions. This leads us to conclude that the native sequence was selected to maximize the coupling between the protein fold and its functional dynamics. A folding-function interrelationship might particularly be the case for a fold like the DHFR, which currently has only been observed in the DHFR family, fulfilling a single function.

Results/Discussion

MD Simulations of Native DHFR and Circular Permutation Variants

Conformational fluctuations of native DHFR and circular permutation variants. We first investigated the overall structural change of the native (closed state) protein as observed in the MD simulations by computing the root mean

square deviations (RMSDs) between the initial structure and snapshots along the trajectory. The profile of the C_{α} RMSD trajectory of the native DHFR (Figure 3A; blue) resembles profiles obtained for stable protein structures. Initially, the C_{α} RMSD of the native DHFR increases rapidly to an average value of 1.7 Å, followed by a gradual rise to about 2 Å at 5 ns and remaining nearly constant in the next 5 ns (Figure 3B). In order to test the effect of prolonged MD simulations on the structural network analysis, we extended the simulations for ten selected circularly permuted variants (five from the folding group and five from the nonfolding group) for additional 5 ns (Figure 3B). There is no significant structural change for additional 5-ns simulation times for the permuted variants. The overall simulation time is 145 ns for the 19 tested circular permutants (Table 1); among these, in ten variants the peptide bond is broken in folding element regions, and in nine the peptide bond is broken outside the folding elements.

The all- C_{α} RMSDs averages (with respect to the starting geometries) of the folding element and nonfolding element groups calculated from the MD trajectories are given in Figure 3A. Comparison of these suggests limited differences between the ten circularly permuted variants of the folding element group and the nine variants of the nonfolding element group during the simulations. Additional MD simulations up to 10 ns show that the permuted variants of the nonfolding element group have slightly higher fluctuation than the folding element group (Figure 3B). It appears that the break of the peptide bond outside the folding element regions leads to relatively high flexibility, whereas breaks occurring in the folding elements can be stabilized by the surrounding residues.

The comparison of the root mean squared fluctuation (RMSF) for each residue in the folding element group versus the nonfolding element group is shown in Figure 3C. The locations of the folding regions (labeled by numbers) and nonfolding regions (labeled by letters) on the structure are given in Figure 2 and are enumerated in Table 1. These regions are taken from the experimental study [5]. In the simulations, we observe that relatively smaller fluctuations occur in folding regions 1–7. Only regions 8–10 have relatively larger fluctuations during the simulation. The large fluctuations occur mostly outside the folding element regions, in regions b, d, e, f, h, and i. This is also observed in the comparison of the RMSF for the folding and nonfolding element groups (Fig 3C). While similar to the native DHFR, most of the large fluctuations occurred outside the folding element regions; in most cases, the RMSF values of the folding element group were smaller than the nonfolding element group. Again, the tightly packed environment of the folding elements stabilized the local region.

We did not observe unfolding in permutants with cuts in folded regions. Assuming that the experimental observations of enzymatic deactivations are caused by unfolding, there are two possibilities for our inability to capture the unfolding events. (1) None of the variants with cuts in folding elements is indeed able to fold into a native-like state. In such a case, our short 5-ns simulations of permutation variants in the folding elements are most likely trapped since we started the simulation with the native-like folded state. The additional 5-ns simulations are unable to change the scenario. (2) Some variants with cuts in the folding elements are still able to fold

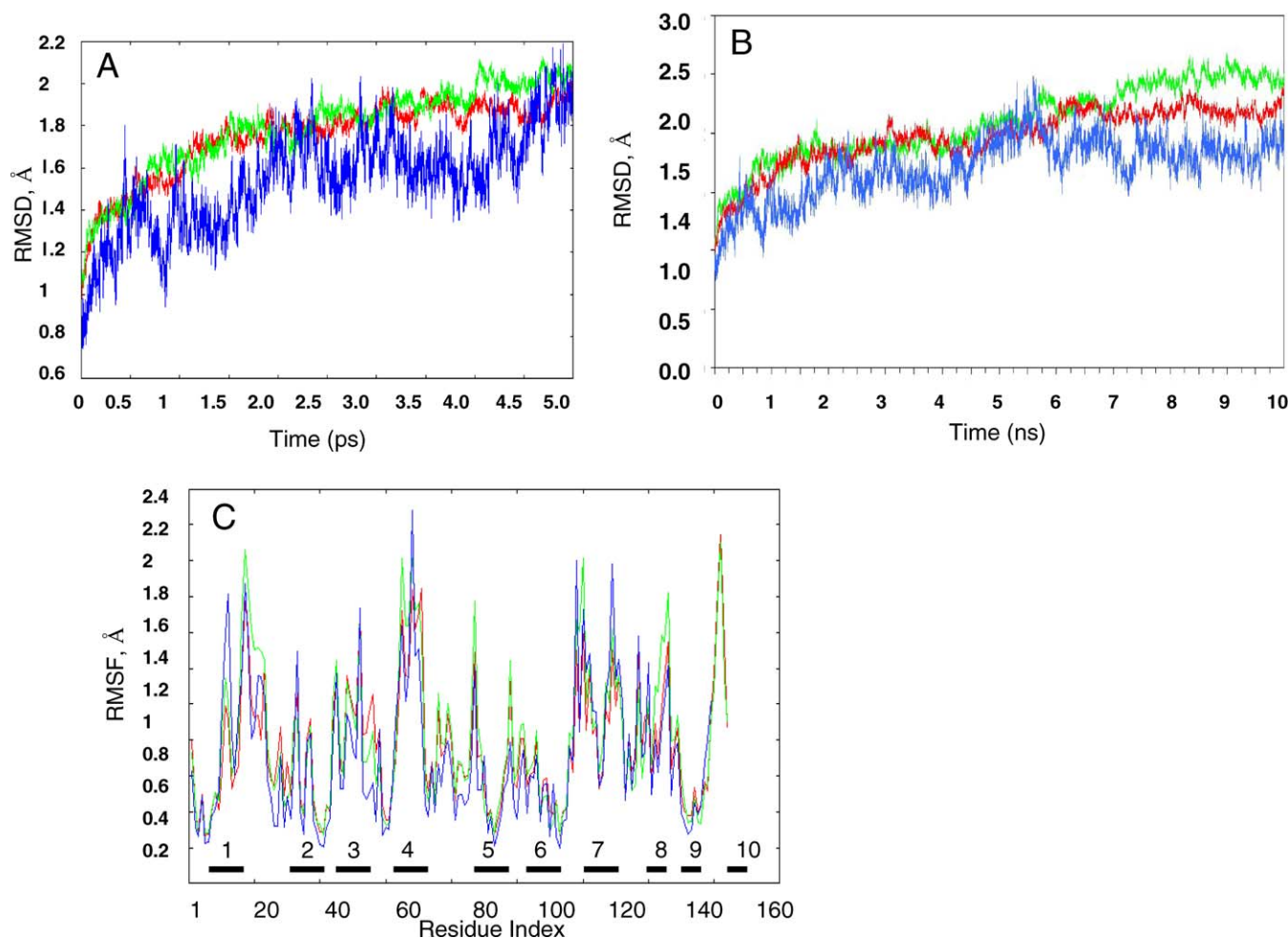


Figure 3. MD Trajectories

(A) Comparison of the all- C_{α} atom RMSDs calculated from the MD trajectories for the folding element group, nonfolding element group, and the native DHFR at 300 °K during the 5-ns simulation (red: folding elements, green: nonfolding elements, blue: native DHFR).

(B) Comparison of the all- C_{α} atom RMSDs calculated from the MD trajectories for five folding element variants (a01, a02, a08, a08, and a10), five nonfolding element variants (b01, b04, b05, b08, and b09), and native DHFR at 300 °K during the 10-ns simulations (red: folding elements, green: nonfolding elements, blue: native DHFR).

(C) The comparison of RMSF for the folding element group, nonfolding element group, and the native DHFR averaged over 2.5- to 5.0-ns simulations (red: folding elements group, green: nonfolding elements group, blue: native DHFR). The blocks and numbers along the x-axis indicate the folding element region; see Figure 2 and Table 1 for reference.

doi:10.1371/journal.pcbi.0030117.g003

into native-like folds if helped by a chaperone, but without native enzymatic function due to effects on the active sites. It has been found that there is a close relationship between folding elements and the early folding site [5]. The systematic permutation experiments coupled folding and function, since a folding element was assigned to a continuous region when the cleavage abolished DHFR enzymatic function [5]. Therefore, it is possible that some permutation mutants, even with a folded native-like structure, may nevertheless lose function due to other reasons.

We stress that our simulations of the circularly permuted folding element group do not assess the folding ability of the systems with respect to either of the above two scenarios; rather, under the assumption that the permutants are able to fold, our simulations address effects of chain connection on the overall dynamics of protein. As seen in Figure 3C, comparison of the overall structural fluctuations from the MD simulations illustrates that we are unable to distinguish

between variants with cuts in folding or in nonfolding elements. In the next section, we compare the structural changes, focusing on loop movements.

Comparison of three states of native DHFR and average structure of circular permutation variants. Experimental kinetic analysis of various DHFR permutants has identified several loop regions important for catalysis [1,2], such as the Met-20 loop, the FG loop (residues 116 to 132), and the third GH loop (residues 142 to 149).

We systematically compared the structural variations for the three native states (i.e., the open, the occluded, and the closed), the average structure from the native (closed) state simulation, and the two average structures from the simulations of the circular permutation variants of the folding and nonfolding elements (Table 2). As shown in Figure 1, the difference between the three native states is in loop 10–23, with an average displacement of 1.7 Å between the closed and open states, 2.6 Å between the occluded and

Table 1. An Overview of the Performed Simulations

Label of Permutation	Element Region	New N-Terminal Residue
Folding elements		
a01, Pmut-12	Ser3–Ile14	12
a02, Pmut-33	Trp30–Leu36	33
a03, Pmut-45	Val40–Ser49	45
a04, Pmut-61	Arg57–Ser63	61
a05, Pmut-81	Glu80–Ala83	81
a06, Pmut-101	Ile91–Leu104	101
a07, Pmut-110	Ala107–Glu118	110
a08, Pmut-125	His124–Phe125	125
a09, Pmut-137	Val136–Ser138	137
a10, Pmut-155	Glu154–Ile155	155
Nonfolding elements		
b01, Pmut-21	Gly15–Ala29	21
b02, Pmut-38	Asp37–Pro39	38
b03, Pmut-52	Ile50–Gly56	52
b04, Pmut-70	Ser64–Asp79	70
b05, Pmut-87	Ala84–Glu90	87
b06, Pmut-106	Pro105–Lys106	106
b07, Pmut-120	Val119–Thr123	120
b08, Pmut-130	Pro126–Ser135	130
b09, Pmut-145	Glu139–Phe153	145

The MD simulations of the closed native state and permuted variants (a01, a02, a08, a09, a10, b01, b04, b05, b08, and b09) were run for 10 ns, and the simulations for the rest were run for 5 ns.

doi:10.1371/journal.pcbi.0030117.t001

closed states, and 2.1 Å between the occluded and open states. The average structure from the simulation has a large deviation from the three crystal structures (the closed, open, and occluded states), and has an additional large change in loop 64–71 (average structure of wild-type DHFR [wild-ave]; Table 2). Using the closed state as a reference, the group of permutation variants with cuts in folding elements has a slightly smaller structural variation than that of the non-folding element cutting group. The average structures of Pmut-45 (permutant with a cut at residue 45; a folding element group cut) and Pmut-87 (a nonfolding cut) have very small variations from the closed state crystal structure. When we use the open state crystal structure as a reference, we observe that Pmut-137 from the folding group and Pmut-106 from the nonfolding group have the least structural variation. When we compare with the occluded structure, we see that all average structures for permutant variants obtained from the MD simulations have large deviations in the range of 3–4 Å in the loop 10–23 region. In summary, the loop conformations in the MD average structures are more similar to the open or closed states than to the occluded structure. Examination of the geometric features of the average structures does not show any noticeable difference among the average structures for the native state, the folding region, and the nonfolding region cutting groups.

Table 2. Structural Variation of the MD Average Structures for the Native DHFR and Its Circular Permutation Variants

Structure	Variation	Loop 10–23	Helix 44–50	Loop 64–71	Loop 116–121	Loop 142–149
Reference: Closed, PDB 1rx1	Open	1.7	0.4	0.5	0.8	0.6
	Occluded	2.6	0.2	0.6	1.0	0.7
	Wild-ave	2.7	0.7	3.6	1.6	1.2
	Pmut-45	1.2	1.0	1.4	0.8	1.0
	Group of cutting in folding elements	1.9	1.4	2.8	1.3	1.1
	Pmut-87	2.2	1.5	1.1	0.8	1.3
	Group of cutting in nonfolding elements	2.3	1.2	2.1	1.1	1.2
Reference: Open, PDB 1ra9	Occluded	2.1	0.4	0.8	0.4	0.3
	Wild-ave	3.6	0.9	3.7	1.9	1.3
	Pmut-137	1.8	0.5	1.8	2.1	0.6
	Group of cutting in folding elements	2.8	1.6	2.9	1.6	1.2
	Pmut-106	1.6	0.9	2.4	0.7	0.7
	Group of cutting in nonfolding elements	2.9	1.4	1.8	1.3	1.4
Reference: Occluded, PDB1rx5	Wild-ave	4.1	0.7	4.0	1.9	1.2
	Pmut-110	3.4	0.4	1.4	0.9	1.1
	Group of cutting in folding elements	3.3	1.4	2.9	1.6	1.2
	Pmut-106	2.4	0.7	1.8	0.8	0.6
	Group of cutting in nonfolding elements	3.5	1.2	2.2	1.4	1.4
Reference: Wild-ave, averaged structure from MD simulation of wild-type	Pmut-12	1.6	1.3	0.8	1.6	0.7
	Group of cutting in folding elements	2.1	1.4	3.2	1.6	0.8
	Pmut-21	2.7	0.4	0.8	1.7	0.8
	Group of cutting in nonfolding elements	2.0	1.3	3.3	1.6	0.8

The average displacements (Å) of the five regions are compared.

doi:10.1371/journal.pcbi.0030117.t002

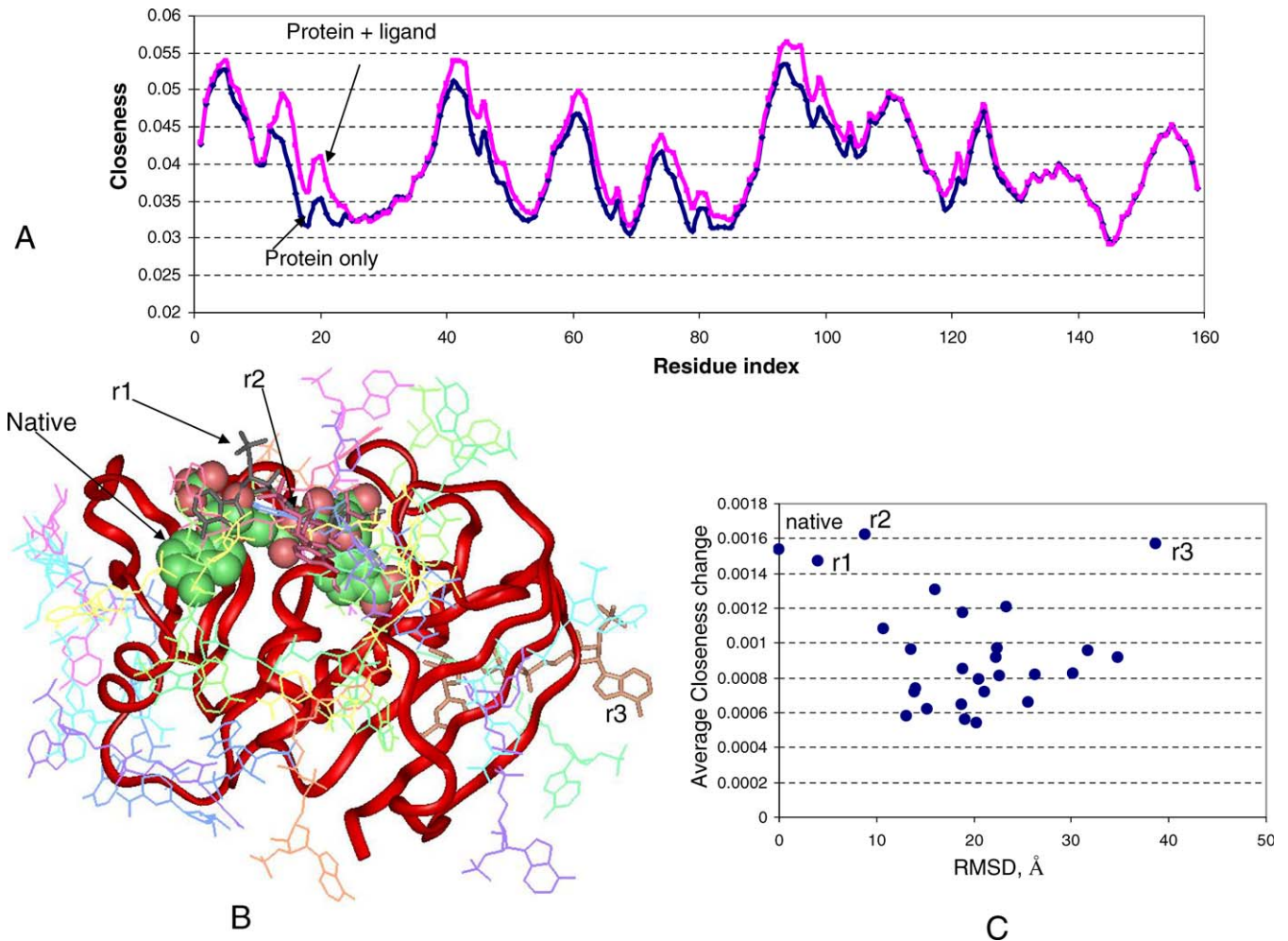


Figure 4. Network-Bridging Effect of Ligand Binding

(A) Comparison of the closeness of native DHFR with and without cofactor binding. The blue line is the plot of the closeness of the apo form of DHFR, and the pink line is the closeness for the DHFR–cofactor complex.

(B) Clusters of docking solutions from patch-docking. The native cofactor is represented as colored spheres, and r1 and r2 are also located in the native binding site, but with different orientation. r3 is away from the binding site. Other docking solutions scattered around the DHFR surface in different regions.

(C) Average closeness change compared with DHFR after ligand binding into various surface patches; see Figure 3A.

doi:10.1371/journal.pcbi.0030117.g004

Changes in the Residue Interaction Network of Native DHFR by Ligand Binding and Circular Permutations

Cofactor binding and its network-bridging effects. Network analysis of protein structures has shown that residues with high closeness values are likely to be functionally important, interacting directly with or close to the binding site [19]. Closeness centrality correlates more accurately with critical residues than betweenness [20]. The above studies imply that the protein active site is near or coincides with residues through which the shortest paths in the residue interaction network tend to go. Therefore, binding of a ligand to the active site could modify the residue interaction network. As a first step in our quest for a relationship between chain connectivity and functional site interaction network, we investigate the closeness changes upon a ligand binding to the DHFR active site.

The nodes of the residue interaction network of DHFR are the C_{α} of all residues. In the DHFR–cofactor complex, the cofactor skeleton is represented by the nitrogen and oxygen

atoms because the C_{α} representation does not apply to organic molecules. We investigated the closeness change for the protein structure with and without the ligand. In Figure 4A, the blue line is the plot of the closeness of the apo form of DHFR (closed state), and the pink line represents the closeness for the DHFR–cofactor complex (closed state). As shown in Figure 4A, the cofactor binding has profound effects on the network centrality, especially the closeness of DHFR. We can see that all closeness peaks from residues 1–110 systematically increase after the cofactor binding, suggesting that most of the network interactions now pass through the cofactor, and that the average shortest path is shorter following cofactor binding. We call the effect of ligand binding on the protein residue interaction network the “network-bridging effect.”

In order to see if the network-bridging effect is unique to the protein–ligand interaction around the binding site, we examined the residue closeness change when the ligand is docked to different DHFR protein surface patches. In the

Table 3. The z-Scores of the Centrality for Key Residues

Residues	Betweenness (Crystal)	Betweenness (Ave)	Closeness (Crystal)	Closeness (Ave)
Ile14	1.62	2.98	0.56	0.66
Gly15	1.82	1.82	0.02	0.12
Phe31	-1.12	0.86	-0.95	-0.68
Met42	1.16	4.27	1.76	1.81
Ser63	-0.17	-0.34	-1.22	-1.29
Tyr100	-0.61	-0.69	1.10	0.02
Gly121	-0.5	-0.53	-0.29	-0.45
Asp122	-1.25	-1.03	-0.35	-0.78
Average	0.12	0.92	0.08	-0.07
Folding element group				
12-13	0.01	-0.04	0.75	0.11
33-34	-1.24	-1.03	0.03	-0.50
45-46	-0.24	0.0004	0.44	-0.12
61-62	1.59	1.17	1.0	1.00
81-82	-0.83	-0.75	-1.17	-0.93
101-102	-0.57	-1.03	0.7	-0.36
110-111	0.37	1.49	1.54	1.92
125-126	2.22	-0.13	0.94	-0.51
137-138	0.99	-0.81	-0.03	0.05
155-156	0.07	-0.08	0.79	1.16
Average	0.24	-0.12	0.50	0.18
Nonfolding element group				
21-22	0.0	-0.24	-1.18	-0.98
38-39	2.46	2.01	0.81	0.96
52-53	-0.65	-0.70	-1.21	-1.23
70-71	-1.11	-1.03	-1.05	-0.67
87-88	-0.41	-0.36	-0.66	-0.3
106-107	-0.65	-0.2	0.68	0.73
120-121	-0.71	-0.78	-0.53	-0.68
130-131	-1.11	-1.03	-0.69	-0.63
145-146	-1.24	-1.03	-1.67	-1.92
Average	-0.38	-0.37	-0.61	-0.52

doi:10.1371/journal.pcbi.0030117.t003

docking study, we keep the ligand structure rigid. A total of 27 clusters were generated from the patch-docking process. In Figure 4B, the native ligand is shown as spheres, and the docked solutions are shown as lines and sticks (sticks for r1, r2, r3, see below; lines for all others). As can be seen in Figure 4B, the docked solutions distribute around the native binding site, as well as at other surface patches around the DHFR. The best docking result (ranked first in the solutions) locates in the native binding site, with 4 Å RMSD from the native ligand (r1, Figure 4B). Another cluster (ranking fifth among the docking solutions) has an 8.9 Å RMSD from the native ligand (r2; Figure 4B).

Depending on the locations of the docked ligand-binding sites, the network-bridging effect changes; and, among all docked solutions, the native bound conformation is within those with the largest closeness changes. Figure 4C plots the average closeness change upon ligand binding for the native and 27 docking clusters. The x-axis is the RMSD from the native ligand when superpositioning the DHFRs. Only three docking clusters caused closeness changes comparable with the native; two of them are located at the native binding site (r1 and r2; Figure 4B and 4C). The third, r3, is located near the strands formed by residues 1-15, which constitute one of the centers of the residue network of DHFR. Most other docking solutions away from the native binding site do not

change the overall closeness pattern. Consistent with previous analysis of protein active sites that have high closeness values [19], here we show that ligand binding at the active site changes the overall closeness values significantly. This indicates that the native functional site of DHFR is sensitive to changes in the residue interaction network and to ligand binding at the active site, leading to the largest network-bridging effect.

The results point to future studies to systematically investigate ligand-binding effects in protein structure networks and the relationship to binding affinity. Our preliminary studies of 230 protein-ligand complexes confirm the general network-bridging effect, and suggest possible correlations of the change of closeness and ligand-binding affinities.

Network analysis of DHFR circular permutation variants: Chain connection and binding site closeness changes. We systematically investigated the network centrality for the native DHFR and its circular permutation variants. First, we examined the crystal structure to obtain the betweenness of locations of important network residues and the locations of the cuts in the permutations. Agarwal et al. identified several key residues in DHFR enzymatic catalysis [4]. The locations of these residues are within the folding elements (for example, Phe31 and Met42) and the nonfolding elements (like G15 and

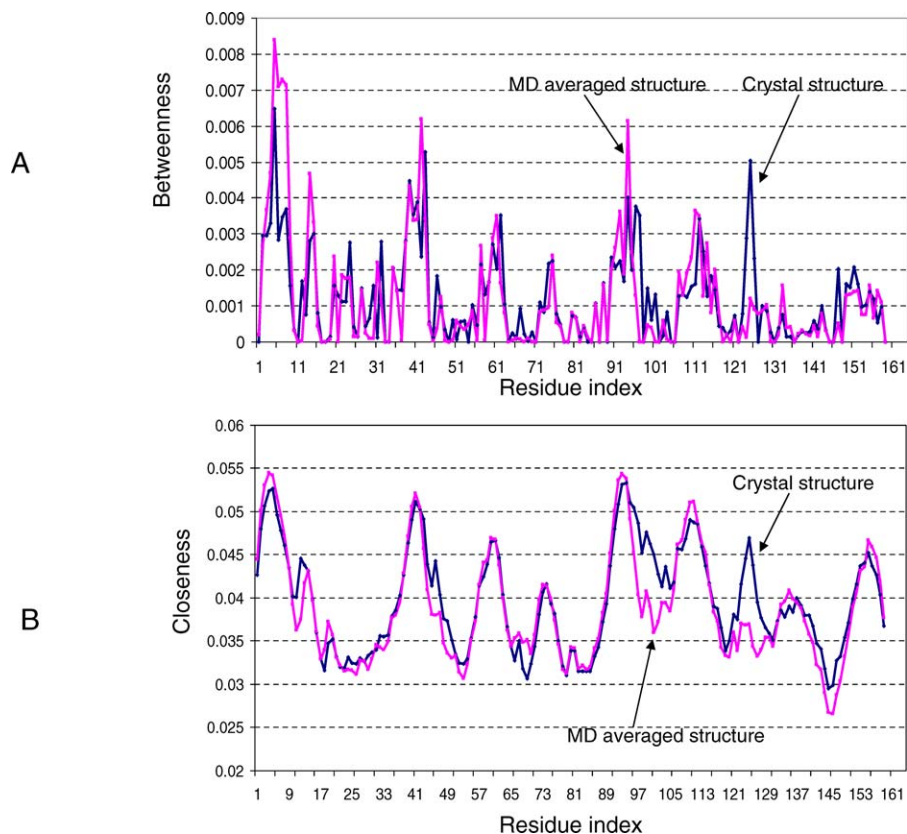


Figure 5. Comparison of Network Centrality of the Crystal Structure (Blue Line) and MD Average Structure (Pink Line) for the Native DHFR (A) Betweenness. (B) Closeness. doi:10.1371/journal.pcbi.0030117.g005

G121, and Asp122). In general, these key residues and the cutting points within the folding elements have higher betweenness than the cutting points in the nonfolding group (Table 3). Analysis of the MD average structure also shows that the z-scores of centrality are higher for the folding element group. The results confirm the importance of the folding element regions in network interaction and protein folding. However, as to individual cutting, we do not see complete agreement, and not all residues are important for folding. Several residues have very low betweenness, and no shortest path goes through residue 102.

Comparison of the betweenness of the crystal structure (closed state) with the MD average structure shows limited changes in certain areas (Figure 5A). The average structure has higher betweenness for residues 1–10, an area experimentally shown to be important in DHFR folding [5]. This area was also proposed to be important as an intramolecular chaperone to guide folding of other regions [21]. In the region around residue 125, the betweenness and closeness of the MD average structure drop sharply. This region is among the folding fragments (of residues 124–125), which also raises the question of the potential folding significance of the fragment. Early work using network analysis to identify the folding nucleus was based on a computed transition structure in protein folding [12], rather than on the crystal structure. Later, Paszkiewicz et al. published their study predicting viable circular permutants using network analysis based only on the crystal structure [22]. They found that in order to

increase the match between predictions and experimentally determined folding regions, they needed to break the DHFR into two domains. In our study, we do not focus on the prediction of possible cutting points. Rather, we are interested in the effects of breaking chain connections on network properties near the active site.

Interestingly, even though there is a large conformational change for loop 10–23, the changes of the network centrality around the loop are much smaller than in the region of residues 120–125, which has only small geometrical variation. Therefore, the analyses of both the static crystal structure and the MD average native structure have shown a consistent functional role for the Met20 loop of residues 10–23.

Comparison with the native DHFR structures (both crystal structure and MD average structure) shows that the network centrality of circular permutation variants change considerably. The betweenness changes are random-like and can draw different patterns for the folding element group and the nonfolding element group (unpublished data). When we look at the closeness centrality, we see that the cuttings in the folding element group cause larger deviation in the binding-related regions compared with the native DHFR structures. Figure 6 plots two graphs to compare the changes of the closeness for the two cutting groups. In Figure 6A, the blue line is the z-score for the MD wild-ave structure, the pink line is the average z-score for the folding element cutting group, and the green line is the average z-score for the nonfolding element cutting group. It can be seen that the three lines have

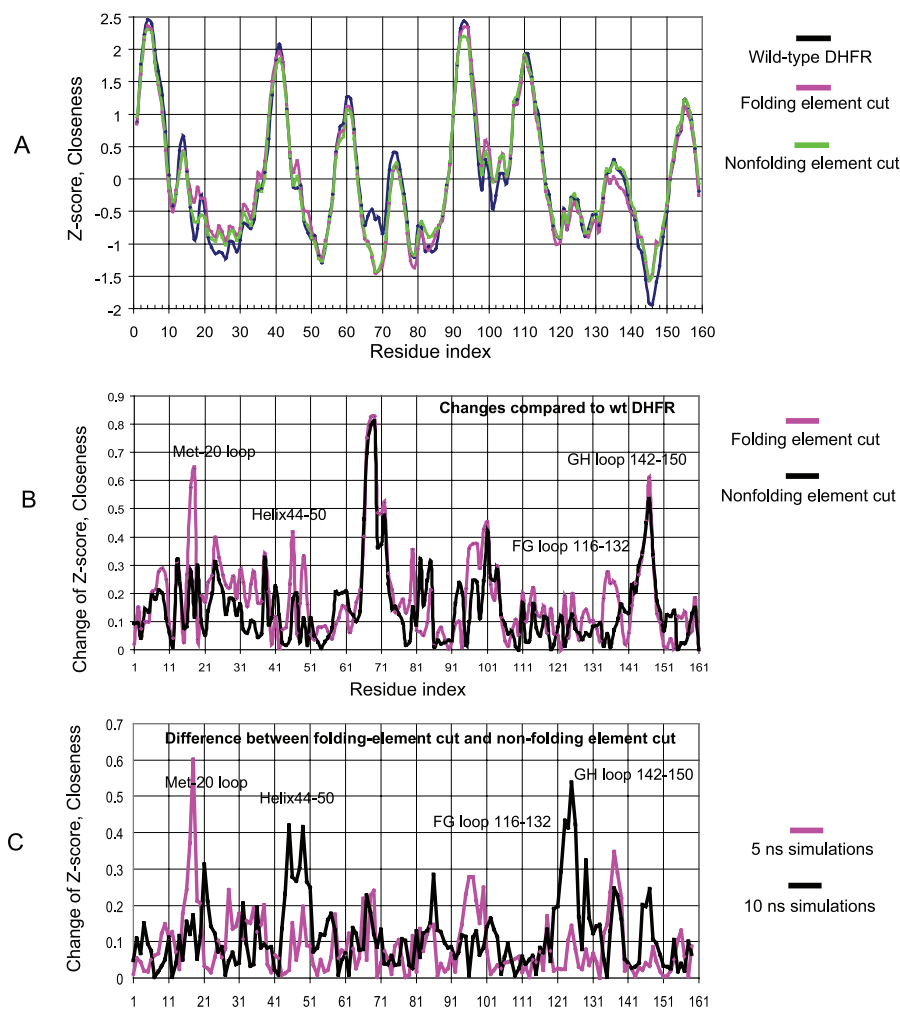


Figure 6. The Change of Closeness for MD Average Structures for Folding and Nonfolding Groups

(A) z-Score of closeness for MD average wild-type DHFR structure (blue), averaged z-scores for the folding elements cutting group (pink), and averaged z-scores for the nonfolding elements cutting group (green).

(B) The changes of the z-scores of the two cutting groups compared with the z-score of closeness for the MD average wild-type DHFR structure.

(C) The differences of the z-scores of the folding element cutting group (a01, a02, a08, a08, and a10) and nonfolding element variants (b01, b04, b05, b08, and b09) averaged from 5-ns MD simulations and 10-ns simulations. The dynamics effects are in the functional loop regions.

doi:10.1371/journal.pcbi.0030117.g006

very similar shapes, indicating overall consistency in the network analysis. In Figure 6B, we plot the changes of the z-scores for the two cutting groups, comparing with the z-score of the closeness for the MD wild-ave structure. The larger the z-score change, the larger the perturbation from the permutation cutting.

At a first glance, we can see that there is a larger change around residues 10–50 for the folding element cutting group. This region has the Met-20 loop, and the helix 44–50, which is also in the binding site. In another catalytically important loop region (FG loop 116–132), we also see a larger change in the folding element cutting group. Both cutting groups have similar perturbations in the GH loop (140–150).

The changes in closeness for these loop regions are dynamic. Figure 6C shows the difference between the closeness of the folding element cutting group and nonfolding element cutting group, averaged from the 2.5- to 5.0-ns simulations and the 5.0- to 10.0-ns simulations. It can be seen that while the difference for the Met-20 loop is large during the 2.5- to 5.0-ns time span, the 5.0- to 10.0-ns dynamics show

differences in the regions of helix 44–50 and the FG loop 116–132.

In summary, the closeness centrality in the network analysis indicates the differing behavior of the functional loop regions upon breaking the chain connection in folding element regions versus in nonfolding element regions.

Conclusions

We have carried out a systematic network analysis and MD simulations of the native DHFR and 19 of its circularly permuted variants. Starting with the crystal structure of the native DHFR, we constructed circular permutations by linking the N-termini and C-termini with five glycine residues, and, following the experiment, introduced breaks in the chain connections in ten folding element regions and in nine nonfolding element regions. The 5-ns MD simulations provided average structures for the native DHFR and for the circularly permuted mutants that are able to fold according to experimental data [5]. For ten circularly permuted variants (five from the folding element group and five from the

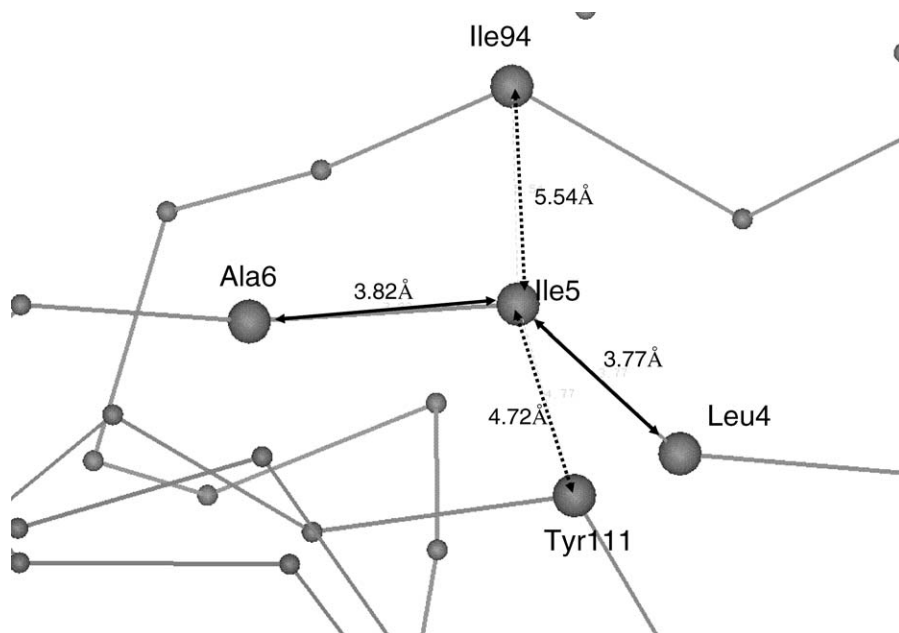


Figure 7. An Illustration of the Difference between Backbone-Connected Contact and Nonbonded Contact. The peptide backbone-linked connections are shown as solid arrows, and the nonbonded packing is shown as dashed arrows. doi:10.1371/journal.pcbi.0030117.g007

nonfolding element group), additional 5-ns simulations provided dynamics effects reflecting the changes in the closeness in the functional loop regions. Analysis of the structural variations illustrated that there are no distinctive structural features indicating whether particular circularly permuted mutants with cuttings within the folding elements will be able to fold. Simulations of experimentally determined permuted mutants that are unable to fold sample only the local minimum on the protein-folding energy landscape and do not reveal the true folding properties of such circularly permuted variants. Nevertheless, our combined MD and network studies with breaks in the chain connection in the folding element or nonfolding element regions present different patterns of perturbation of the network properties near the active site.

Our network analysis further leads us to propose that ligand binding induces network-bridging effects in the protein structure. We observed that substrate binding has profound effects on the DHFR network centrality, especially on closeness. The protein active site is near or coincides with residues through which the shortest paths in the residue interaction network tend to go. Our results suggest that cofactor binding leads to a modification of the interaction network, with most of the interactions now passing through the cofactor. The average shortest path is shorter with cofactor binding. Our findings are consistent with experimental observations that substrate binding increases DHFR folding stability [23,24].

In conclusion, our analysis demonstrates that active site dynamics of DHFR are communicated to the whole protein via both the peptide backbone and the nonbonded residue contacts. Such communication is indicated by the closeness change accompanying ligand binding, by breaking chain connections in the folding element region, and by breaking chain connections in the nonfolding element region. Even though the native and the circularly permuted proteins have

similar overall folds, breaking the chain connections at different regions and ligand binding can change the properties of the network.

Materials and Methods

MD simulations. The structure of DHFR from *E. coli* (closed state) was taken from the Protein Data Bank (<http://www.rcsb.org/pdb>). Both crystallographic waters and substrates were deleted. Circular permutation of the protein consists of connecting the native N- and C-termini covalently with a peptide linker and cleaving the peptide backbone at one specific site. Because a five-glycine peptide was shown to be the most favorable linker in the circularly permuted DHFR [5], this peptide linker was used in all variants in our study. The MOE software (Chemical Computing Group, <http://www.chemcomp.com>) was used to obtain all the circularly permuted variants of the native DHFR, which now has 164 residues. A total of 19 variants were selected for simulation: ten in the proposed folding element region and nine in the nonfolding element region (Table 1).

The CHARMM program [25] (version 30b1) was used for all computations with the CHARMM force field version 22 using all atom representation [26]. The native DHFR and its circularly permuted variants were simulated in a $60 \times 60 \times 60 \text{ \AA}^3$ explicitly solvated periodic box. TIP3P water molecules were introduced. The simulations were carried out with a distance cutoff of 13 Å and a constant dielectric constant of 1. Each simulation was initialized with adopted basis Newton-Raphson (ABNR) minimization followed by 3 ps system heating and 17 ps system equilibration. A production simulation run was carried out for each of the protein structures described above with a 1-fs time step. The coordinates were saved at 1-ps time intervals. Each simulation was run at 300 °K for 5 ns. The average structures from the last 2.5-ns simulations were used for network analysis.

Molecular docking of DHFR ligand complex. Patchdock [27], a geometry-based molecular docking algorithm, was used to generate clusters for the DHFR-ligand interactions. We kept those clusters with at least four docked conformations with RMSDs within 3.5 Å.

Analysis of the protein residue interaction network: Chain connection and nonbonded packing. The amino acid network is defined by all residues within a contact distance. Residue interaction network analysis often uses uniformed distance as long as the contacting residues are within a cutoff distance. Such an approach does not distinguish between chain connection and nonbonded interaction. A recent network permutation analysis of DHFR based on such a definition led to two separate regions [22].

The nature of the chemical bond argues for a strong communication when two residues are sequentially linked by a peptide bond. In Figure 7, we show that the most connected residue Ile5 has two bonded connections with Leu4 and Ala6 and the two closely packed residues Ile94 and Tyr111. Since residues connected by a peptide bond have shorter C_{α} distances, a weighting of distance can distinguish between chain connection and nonbonded interaction. Thus, we define contact distances based on the C_{α} of all residues within 6.0 Å and use the integer of the distance as weight. Therefore, 6 is coded for a distance of 6.0 Å, 5 for a distance between 5.0 and 6.0 Å, 4 for a distance between 4.0 and 5.0 Å, and 3 for a distance between 3.0 and 4.0 Å. The chain connection can have a distance index of 3 or 4, and a nonbonded contact can have a distance index of 5 or 6. This definition can effectively reflect the effects of chain cleavage on the network properties. The algorithm by Pape [28] was used to calculate the shortest path lengths between nodes. Two network properties are computed to characterize network properties of a given protein structure.

The betweenness [29] is one of the standard measures of node centrality. The betweenness bi of a node i is defined as

$$bi = \sum_{j,k \in N, j \neq k} \frac{n_{jk}(i)}{n_{jk}} \quad (1)$$

where n_{jk} is the number of shortest paths connecting j and k , while $n_{jk}(i)$ is the number of shortest paths connecting j and k and passing through i .

The closeness centrality of node x is the inverse of the average distance between x and other nodes:

$$C(x) = 1/\text{ave}(d) \quad (2)$$

The z-score of the closeness is calculated by $z\text{-score} = (C(x) - \mu) / \sigma$, where μ is the average value of closeness and σ is the standard deviation.

References

- McElheny D, Schnell JR, Lansing JC, Dyson HJ, Wright PE (2005) Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc Natl Acad Sci U S A* 102: 5032–5037.
- Venkitakrishnan RP, Zaborowski E, McElheny D, Benkovic SJ, Dyson HJ, et al. (2004) Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle. *Biochemistry* 43: 16046–16055.
- Sawaya MR, Kraut J (1997) Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: Crystallographic evidence. *Biochemistry* 36: 586–603.
- Agarwal PK, Billeter SR, Rajagopalan PT, Benkovic SJ, Hammes-Schiffer S (2002) Network of coupled promoting motions in enzyme catalysis. *Proc Natl Acad Sci U S A* 99: 2794–2799.
- Iwakura M, Nakamura T, Yamane C, Maki K (2000) Systematic circular permutation of an entire protein reveals essential folding elements. *Nat Struct Biol* 7: 580–585.
- Nakamura T, Iwakura M (1999) Circular permutation analysis as a method for distinction of functional elements in the M20 loop of *Escherichia coli* dihydrofolate reductase. *J Biol Chem* 274: 19041–19047.
- Arai M, Iwakura M (2005) Probing the interactions between the folding elements early in the folding of *Escherichia coli* dihydrofolate reductase by systematic sequence perturbation analysis. *J Mol Biol* 347: 337–353.
- Arai M, Iwakura M (2006) Peptide fragment studies on the folding elements of dihydrofolate reductase from *Escherichia coli*. *Proteins* 62: 399–410.
- Svensson AK, Zitzewitz JA, Matthews CR, Smith VF (2006) The relationship between chain connectivity and domain stability in the equilibrium and kinetic folding mechanisms of dihydrofolate reductase from *E. coli*. *Protein Eng Des Sel* 19: 175–185.
- Smith VF, Matthews CR (2001) Testing the role of chain connectivity on the stability and structure of dihydrofolate reductase from *E. coli*: Fragment complementation and circular permutation reveal stable, alternatively folded forms. *Protein Sci* 10: 116–128.
- Pelletier JN, Arndt KM, Pluckthun A, Michnick SW (1999) An in vivo library-versus-library selection of optimized protein-protein interactions. *Nat Biotechnol* 17: 683–690.
- Vendruscolo M, Dokholyan NV, Paci E, Karplus M (2002) Small-world view of the amino acids that play a key role in protein folding. *Phys Rev E Stat Nonlin Soft Matter Phys* 65: 061910.
- Rao F, Caflisch A (2004) The protein folding network. *J Mol Biol* 342: 299–306.

Supporting Information

Accession Numbers

The Protein Data Bank (<http://www.rcsb.org/pdb>) accession numbers for the structures discussed in this paper are open state (1ra9); closed state (1rx1); and occluded state (1rx5).

Acknowledgments

Computations were conducted using the resources in the Pittsburgh Supercomputing Center (<http://www.psc.edu>).

Author contributions. BM conceived and designed the experiments. ZH, DB, WMS, and YP performed the experiments. ZH and BM analyzed the data. AdS, RN, and BM wrote the paper.

Funding. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health (NIH), under contract number NO1-CO-12400. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research and by the US Army Medical Research Acquisition Activity under grant W81XWH-05-1-0002. The work at Howard University is supported by grant RCMI-NIH 2G12RR03048 from the Research Centers in Minority Institutions (RCMI) Program (Division of Research Infrastructure, National Center for Research Resources, NIH), and grant MCB030037P from the Pittsburgh Supercomputing Center. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Competing interests. The authors have declared that no competing interests exist.

- Dokholyan NV, Li L, Ding F, Shakhnovich EI (2002) Topological determinants of protein folding. *Proc Natl Acad Sci U S A* 99: 8637–8641.
- del Sol A, O'Meara P (2005) Small-world network approach to identify key residues in protein-protein interaction. *Proteins* 58: 672–682.
- Greene LH, Higman VA (2003) Uncovering network systems within protein structures. *J Mol Biol* 334: 781–791.
- Brinda KV, Vishveshwara S (2005) A network representation of protein structures: Implications for protein stability. *Biophys J* 89: 4159–4170.
- Atigan AR, Akan P, Baysal C (2004) Small-world communication of residues and significance for protein dynamics. *Biophys J* 86: 85–91.
- Amitai G, Shemesh A, Sitbon E, Shklar M, Netanel D, et al. (2004) Network analysis of protein structures identifies functional residues. *J Mol Biol* 344: 1135–1146.
- Thibert B, Bredesen DE, del Rio G (2005) Improved prediction of critical residues for protein function based on network and phylogenetic analyses. *BMC Bioinformatics* 6: 213.
- Ma B, Tsai CJ, Nussinov R (2000) Binding and folding: In search of intramolecular chaperone-like building block fragments. *Protein Eng* 13: 617–627.
- Paszkiwicz KH, Sternberg MJ, Lappe M (2006) Prediction of viable circular permutants using a graph theoretic approach. *Bioinformatics* 22: 1353–1358.
- Ainavarapu SR, Li L, Badilla CL, Fernandez JM (2005) Ligand binding modulates the mechanical stability of dihydrofolate reductase. *Biophys J* 89: 3337–3344.
- Junker JP, Hell K, Schlierf M, Neupert W, Rief M (2005) Influence of substrate binding on the mechanical stability of mouse dihydrofolate reductase. *Biophys J* 89: L46–L48.
- Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4: 187–217.
- MacKerell AD Jr, Bashford D, Bellott RL, Dunbrack RL Jr, Evanseck JD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102: 3586–3616.
- Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res* 33: W363–W367.
- Pape U (1980) Algorithm 562—Shortest-path lengths [H]. *ACM TOMS* 6: 450–455.
- Boccaletti S, Pecora LM (2006) Introduction: Stability and pattern formation in networks of dynamical systems. *Chaos* 16: 15101.