Divergent evolution of protein conformational dynamics in dihydrofolate reductase

Gira Bhabha1,2,4, Damian C Ekiert1,2,4, Madeleine Jennewein1,2, Christian M Zmasek3, Lisa M Tuttle1,2, Gerard Kroon1,2, H Jane Dyson1,2, Adam Godzik3, Ian A Wilson1,2 & Peter E Wright1,2

Molecular evolution is driven by mutations, which may affect the fitness of an organism and are then subject to natural selection or genetic drift. Analysis of primary protein sequences and tertiary structures has yielded valuable insights into the evolution of protein function, but little is known about the evolution of functional mechanisms, protein dynamics and conformational plasticity essential for activity. We characterized the atomic-level motions across divergent members of the dihydrofolate reductase (DHFR) family. Despite structural similarity, Escherichia coli and human DHFRs use different dynamic mechanisms to perform the same function, and human DHFR cannot complement DHFR-deficient E. coli cells. Identification of the primary-sequence determinants of flexibility in DHFRs from several species allowed us to propose a likely scenario for the evolution of functionally important DHFR dynamics following a pattern of divergent evolution that is tuned by cellular environment.

RESULTS
Active site loop motions in human DHFR

Diversification of gene families and their resulting protein products through mutation, random genetic drift and natural selection has resulted in the wide spectrum of enzymes, signal transducers, cellular scaffolds and other molecular machines found in the diverse species represented in all kingdoms of life. The effects of divergent evolution on three-dimensional protein structures are addressed in many studies that provide fundamental insights into evolutionary pressures that drive diversification of protein folds1–3. However, motions and flexibility are also essential for the function of proteins and macromolecular machines, and, just as protein structures are subject to natural selection, evolutionary pressures might also be expected to tune protein dynamics to adapt proteins to new environments and facilitate the emergence of new functionalities. Indeed, comparisons between thermophilic and mesophilic enzymes reveal that their dynamics and activity are adapted to the thermal environment of the organism4,5. In principle, the adaptation of enzymes to different environments or to specialized functions may involve a radical reconfiguration of the dynamic landscape. Understanding how new dynamic modes arise would provide fundamental insight into the evolution of new functionality, and this is addressed here in the context of the enzyme DHFR.

DHFR catalyzes the NADPH-dependent reduction of dihydrofolate (DHF) to tetrahydrofolate (THF), an essential precursor for thymidylate synthesis in cells6. The evolution of DHFR is of great interest, with NADP+ and folic acid (hE–NADP+–FOL, 1.4-Å resolution) and in complex with NADP+ and 5,10-dideazatetrahydrofolate (hE–NADP+–ddTHF, 1.7-Å resolution), which model the Michaelis complex and product ternary complex, respectively. In contrast to ecDHFR, in which the Met20 loop moves from the closed conformation in the E–NADPH and E–NADP+–FOL complexes to the occluded conformation in the three product complexes (Fig. 1c)18, ecDHFR (Fig. 1a), their primary sequences are highly divergent, as reflected in subtle changes in the catalytic cycle9,10,13 with different kinetics and different rate-limiting steps under physiological concentrations of ligands (Fig. 1b). We hypothesized that ecDHFR and hDHFR may have evolved different dynamic mechanisms within the constraints of the same fold and the same key catalytic residues. To address this hypothesis, we used an integrated approach, including structural biology, mutagenesis, bioinformatic analyses and cell biology, which allowed us to uncover evolutionary aspects of the motions present in the DHFR enzyme family.
thereby facilitating ligand flux\textsuperscript{14,21–23}, hDHFR remains in the closed conformation in both ligand-bound states, without any apparent structural change in the active site loops (Fig. 1d). Thus, in hDHFR, the Met20 loop appears to be locked in place and unable to undergo this conformation change. Consistent with our findings, the active site loops adopt the closed conformation in all available crystal structures of vertebrate DHFRs, including complexes of hDHFR with small-molecule inhibitors and a substrate (folate)\textsuperscript{24}. Importantly, the closed-to-occluded conformational transition in ecDHFR can also be visualized directly in solution by comparison of the \textsuperscript{15}N HSQC spectra of the ecE–NADP\textsuperscript{+}–FOL and ecE–NADP\textsuperscript{+}–THF complexes, which differ owing to the conformational change in the Met20 loop (Fig. 1e)\textsuperscript{14,18,20}. In marked contrast to ecDHFR, the \textsuperscript{15}N HSQC spectra of the hE–NADP\textsuperscript{+}–FOL and hE–NADP\textsuperscript{+}–THF complexes are almost identical (Fig. 1f), thus showing that in solution, as well as in the crystal structures, no backbone conformational changes are observed for the human enzyme.

Active site packing and preorganization in hDHFR
The hDHFR active site cleft in the model Michaelis complex, E–NADP\textsuperscript{+}–FOL, is more tightly packed than that of ecDHFR bound to the same ligands (Fig. 2a,b) and probably has an important role in optimal positioning of the donor and acceptor atoms for catalysis, thereby contributing to its increased rate of hydride transfer\textsuperscript{8,10}. However, the tight packing of the hDHFR active site, coupled with the apparent lack of active site loop motions to facilitate ligand flux, raises a critical question: how do the ligands get into and out of the hDHFR active site?

Differences in dynamics between ecDHFR and hDHFR
In ecDHFR, millisecond-timescale fluctuations in the active site contribute to efficient ligand flux and catalysis; mutations that perturb the dynamic equilibrium between the closed and occluded conformations of the Met20 loop notably alter substrate and cofactor flux\textsuperscript{14,22,23,25,26}. Although the crystal structures and \textsuperscript{15}N HSQC spectra of hE–NADP\textsuperscript{+}–FOL and hE–NADP\textsuperscript{+}–THF (or hE–NADP\textsuperscript{+}–ddTHF) suggest that the active site loops of hDHFR are predominantly closed, we used CPMG-based \textsuperscript{15}N R\textsubscript{2} relaxation dispersion NMR experiments to assess whether transient loop fluctuations might facilitate

### Table 1 Data collection and refinement statistics

|                  | hE–NADPH | hE–NADP\textsuperscript{+}–FOL | hE–NADP\textsuperscript{+}–ddTHF |
|------------------|----------|---------------------------------|----------------------------------|
| **Data collection** |          |                                 |                                  |
| Cell dimensions  |          |                                 |                                  |
| a (Å)            | 39.0     | 65.5                            | 152.3                            |
| b (Å)            | 68.2     | 68.2                            | 160.2                            |
| c (Å)            | 120.0    | 90.0                            | 90.0                             |
| Resolution (Å)   | 50–1.20  | 50–1.40                         | 50–1.70                          |
| R\textsubscript{sym} | 0.11     | (0.54)                          | (1.42–1.40)                      |
| Completeness (%) | 96.8     | (76.9)                          | (92.3)                           |
| Redundancy       | 6.5 (2.9)| 19.3 (8.7)                      | 18.3 (8.4)                       |
| **Refinement**   |          |                                 |                                  |
| Resolution (Å)   | 50–1.20  | 50–1.40                         | 50–1.70                          |
| No. reflections  | 59,264   | 44,348                          | 24,628                           |
| R\textsubscript{work} / R\textsubscript{free} | 16.3 / 18.5 | 14.5 / 18.4 | 19.2 / 24.0 |
| No. atoms        | Protein  | 1,639                           | 1,635                            |
|                  | Ligand/on| 48                               | 86                               |
|                  | Water    | 196                             | 206                              |
| B factors        | Protein  | 24.0                            | 20.2                             |
|                  | Ligand/on| 20.8                            | 19.4                             |
|                  | Water    | 38.8                            | 32.7                             |
| r.m.s. deviations| Bond lengths (Å) | 0.009 | 0.009 | 0.007 |
|                  | Bond angles (°) | 1.48 | 1.41 | 1.25 |

\textsuperscript{a}Values in parentheses are for highest-resolution shell.
In stark contrast to ecDHFR, fluctuations on the millisecond timescale are not observed in the hE–NADPH, hE–NADP+–FOL, hE–NADP+–THF, hE–FOL, E–THF or hE–NADPH complexes, thus suggesting that the human enzyme uses a different mechanism or different motions to mediate ligand flux. Indeed, $^{15}N$ R$_{1p}$ relaxation dispersion experiments revealed pervasive motions on a faster microsecond timescale in the hE–NADP+–FOL complex (Supplementary Fig. 3a–c). These rapid conformational fluctuations (at rates ranging from $\sim$15,000 to 30,000 s$^{-1}$) occur in many regions of the enzyme, including regions that line one edge of the active site, and may have a role in ligand binding and release. Remarkably, despite the structural similarity between human and E. coli DHFRs, both the nature and timescale of the ground-state conformational fluctuations have diverged considerably, thus bolstering the hypothesis that the dynamic mechanisms of ecDHFR and hDHFR are fundamentally different.

**Figure 2** Active site packing and hinge motions in hDHFR.

(a,b) Surface rendition of hDHFR–NADP$^+$–FOL (a) and ecDHFR:NADP$^+$–FOL (b) generated with only ambient occlusion, a three-dimensional light-attenuation calculation in which deep pockets render dark and exposed surfaces render light$^{35}$. c,d Superposition of crystal structures, aligned on the loop subdomain (gray), of hE–NADPH and hE–NADP$^+$–FOL (c) and ecE–NADPH (PDB 1RX1 (ref. 18)) and ecE–NADP+–FOL (PDB 1RX2 (ref. 18)) (d). The adenosine-binding subdomain is colored green for hE–NADPH and pink for hE–NADP$^+$–FOL. Ligands are shown as sticks, with NADPH in green, NADP$^+$ in magenta and folate in yellow. e–h Surface representations of hE–NADPH (e), ecE–NADPH (f), hE–NADP$^+$–FOL (g) and ecE–NADP$^+$–FOL (h). Residues highlighting the opening and closing of the active site cleft are colored in red. (i,j) Difference distance matrix for hE–NADPH and hE–NADP$^+$–FOL (i) and ecE–NADPH and ecE–NADP$^+$–FOL (j), showing the magnitude and character of the conformational changes associated with the hinge motions.

**Exaggerated hinge movements in hDHFR**

To gain further insights into how hDHFR exchanges substrate in the absence of flexible active site loops, we determined the crystal structure of the hE–NADPH binary complex at 1.2-Å resolution. Notably, the active site cleft in the hE–NADPH structure is more open than that in the hE–NADP$^+$–FOL complex (Fig. 2c.e.g). Opening of the active site is accomplished not by the motion of flexible loops, as in ecDHFR, but by a rigid-body, twisting-hinge motion that rotates the adenosine-binding subdomain (defined as Thr40–Gly129) ~10° away from the loop subdomain (Gly2–Thr39 and His130–Asp186). These subdomains are connected by two hinges: hinge 1 (Thr39–Leu49) and hinge 2 (Gly129–Leu131). In the hE–NADPH complex, the hinges stabilize the active site cleft in an open (hinge-open) conformation that probably facilitates substrate entry (Fig. 2e). Upon binding of substrate (folate), the adenosine-binding subdomain rotates inward to tightly close the active site (Fig. 2g, Supplementary Fig. 2a–c and Supplementary Fig. 2d–f).

**Figure 3** Primary-sequence features related to flexibility and conformational change in E. coli and human DHFR. (a) Sequence alignment of ecDHFR and hDHFR showing three regions of the sequence related to dynamic mechanism. The anchor residues for sequence alignment are shown in red. (b) Structure of regions highlighted in a, with anchor residues shown as spheres. Purple, ecDHFR; orange, hDHFR. Regions A, B and C correspond to the Met20 loop, hinge 1 and hinge 2, respectively. The following anchor residues were chosen for sequence alignments (E. coli numbering): Region A, Pro21 and Asp27; region B, Phe31 and Met42; region C, Tyr100 and Thr113.
Supplementary Table 1. Changes in chemical shifts between the NMR spectra of the hE–NADPH and hE–NADP+–FOL complexes (Supplementary Fig. 3d,e) suggest that this hinge movement also occurs in solution. In both hinge-open and hinge-closed conformations, an extensive hydrogen-bonding network stabilizes the hinge-1 structure and anchors it to the adenosine-binding subdomain and to the C-terminal end of the αβ helix (Supplementary Note, Supplementary Fig. 2d and Supplementary Table 2). Hinge 1 moves as a rigid body to accommodate the subdomain rotation, and hinge 2 allows the sliding motion of helix αF that opens and closes the active site cleft.

Interestingly, subtle hinge movements have also been reported for ecDHFR. However, unlike that of the human enzyme, the subdomain rotation in ecDHFR displays a much smaller range of motion of the active site cleft (Fig. 2f,h). At the widest point, the hDHFR active site cleft opens by ~3 Å (comparison of hE–NADPH and hE–NADP+–FOL; Fig. 2e,g), whereas ecDHFR opens ~0.5 Å (comparison of PDB 1RX1 (ref. 18) and 1RX2 (ref. 18); Fig. 2f,h). We further quantified the difference between binary and ternary structures for hDHFR and ecDHFR by calculating the distance difference matrix (Fig. 2i, j), which shows clearly that the extent of motion in the human enzyme is greater at this step of the catalytic cycle than that observed for the E. coli enzyme. The length of the hinges is seemingly a key factor in determining the magnitude of the hinge bending motions. The ecDHFR hinge 1 and 2 regions are tightly packed and may not be long enough to provide the structural framework for the larger rigid-body motion observed in hDHFR (Fig. 3). In particular, hinge 1 in ecDHFR is ‘embedded’ in a tight groove where any large movement would be highly constrained by clashes with both the adenosine-binding and loop subdomains. The observation that a subtle subdomain rotation is possible in the E. coli enzyme provides a framework for understanding how alternative dynamic mechanisms may arise and change during the course of evolution. Just as new functionalities are most frequently derived from existing protein folds, new dynamic mechanisms probably derive from existing modes of protein motions. The subtle hinge motion in ecDHFR may represent an ancestral motion that has been accentuated in hDHFR through the course of evolution, increasing the amplitude of domain movement to permit ligand flux and thus eliminating the need for the closed-occluded transition of the Met20 loop used by ecDHFR.

Sequence determinants and evolution of dynamic mechanisms

To further explore the evolution of dynamics in the DHFR family, we conducted a comprehensive comparative analysis of all available DHFR sequences (~1,800 sequences), including eukaryotic DHFRs from all available fully sequenced genomes (172 sequences) (Supplementary Table 3). In particular, we focused on three regions that we hypothesized would best account for the mechanistic differences between ecDHFR and hDHFR: region A, at the end of the Met20 loop, and regions B and C, which contain the hinges implicated in subdomain rotation (Fig. 3). The analysis reveals a clear pattern in sequence and length of regions A, B and C that we link to flexibility, thus allowing us to propose that these regions of the DHFR sequence determine, at least in part, the protein motions that mediate the flux of substrate, product and cofactor at various stages of the catalytic cycle.

Region A loop length influences conformational flexibility

Region A contains seven residues in ecDHFR and eight residues in hDHFR. This region became proline rich in recent evolutionary times. Incorporation of the PWPP motif from region A of the human enzyme into ecDHFR, through the mutation N23PP (which changes 21-PWNL-24 to 21-PWPL-24) alters the motions in the Met20 loop and active site. To determine whether this proline motif in region A affects flexibility in ecDHFR, or whether the loop length itself contributes to differences in conformational sampling, we characterized an ecDHFR mutant with an alanine insertion (21-PWNAL-24). 15N HSQC spectra of the complexes before and after hydride transfer (modeled by E–NADP+–FOL and E–NADP+–THF) show that insertion of a single residue (alanine) is sufficient to prevent the transition to the occluded conformation (Fig. 4a–c), and it severely dampens millisecond-timescale motions in the active site (Supplementary Fig. 4). Thus, it is primarily the length of region A and not its proline content that influences the conformational flexibility of the Met20 loop.

The majority of prokaryotic DHFRs have seven residues in region A, and we hypothesized that they would have conformational flexibility similar to that of ecDHFR. Region A in eukaryotic DHFRs is more heterogeneous, with sequences containing seven residues (E. coli like), eight residues (human like) or more than eight residues (Fig. 5). We chose representative DHFRs from several species (Supplementary Fig. 5a–d), in which region A contains seven, eight or ten residues, and investigated their flexibility, using NMR spectroscopy to obtain insights into ligand-dependent conformational sampling. We compared 15N HSQC spectra of the DHFRs from these species bound either to NADP+ and FOL (model Michaelis complex) or to NADP+ and THF (product ternary complex) to identify conformational changes across the hydride-transfer step. For four DHFRs containing...
Figure 5 Overview of patterns in length of Met20 loop and hinges. A, B and C refer to regions A (Met20 loop), B (hinge) and C (hinge) of the DHFR primary sequence (as described in the text and Fig. 3). For A, open squares indicate seven residues, and filled squares indicate eight or ten residues in region A. Enzymes with seven residues in region A undergo conformational changes across the hydride-transfer step. An increase in the length of region A (to eight or ten residues) is associated with limited flexibility in the active site loops and the absence of conformational change upon formation of product. For B, open squares indicate a short hinge (<15 residues in region B), and filled squares indicate a long hinge (≥15 residues in region B). For C, open squares indicate a short hinge (12 residues in region C), and filled squares indicate a long hinge (≥14 residues in region C). Long hinges facilitate the exaggerated hinge-twisting motion observed in hDHFR. Whereas in some groups (for example, fungi) more than one combination of features can be found, their distributions within the group do not follow any well-established phylogenetic divisions.

We also acquired HSQC spectra for the NADP+–FOL and NADP+–THF complexes of several vertebrate DHFRs that vary in length and sequence of region A. For Sus scrofa (pig) and Bos taurus (cow) DHFR, region A is identical to that of humans, and the NMR experiments confirm that no conformational changes occur across the hydride-transfer step (Supplementary Fig. 5c–d). Rattus norvegicus (rat) DHFR (rDHFR) also contains eight residues, but Pro26 of hDHFR is replaced by leucine. Danio rerio (zebrafish) DHFR (zDHFR) contains a two-residue insertion in region A, making it ten residues in length, with histidine in place of Pro25 of hDHFR (Supplementary Fig. 5). Again, the almost-identical 15N HSQC spectra of complexes of rDHFR and zDHFR before and after hydride transfer confirm that the enzymes do not undergo conformational changes across the chemical step (Fig. 4f and Supplementary Fig. 6a–d). We therefore conclude that proline at position 25 or 26 (hDHFR numbering) is not essential for stabilization of the closed conformation and that the length of loop A is the primary determinant of conformational change in the active site loops between complexes before and after hydride transfer.

The Caenorhabditis elegans DHFR (ceDHFR) sequence is an example of a eukaryotic DHFR in which region A is seven residues (as in bacterial enzymes), but the hinges have similar lengths to that of hDHFR. We observed several chemical-shift differences in the 15N HSQC spectra between complexes of ceDHFR before and after hydride transfer, results indicating similar conformational flexibility to bacterial DHFRs (Fig. 4e). ceDHFR is prototypical of a large subset of eukaryotic DHFRs, in which the region-A length resembles that of ecDHFR, whereas regions B and C are similar to those of hDHFR.

**Properties of the hinges**

Hinge 1 is located in region B. Bacterial DHFR sequences predominantly contain short region-B sequences, as in ecDHFR (12 residues), whereas in eukaryotic DHFRs, with very few exceptions found in unicellular species (for example, some amoebozoans and stramenopiles), region B is ≥19 residues (Fig. 5 and Supplementary Fig. 5). In hDHFR, Asn48 hydrogen-bonds to the backbone of Thr38 and Thr40 in the hinge and to Met111 in the adenosine-binding domain, thus maintaining the hinge structure during subdomain rotation (Supplementary Fig. 2 and Supplementary Table 2). Interestingly, this asparagine is invariant in all eukaryotic 19-residue region-B sequences (Supplementary Table 4) and is likely to be a key mediator of the hinge motion.

Hinge 2, within region C, is formed by His127–Leu131 for hDHFR and Pro105–Ala107 for ecDHFR. Region C contains 14 residues in E. coli DHFR and 16 residues in hDHFR (Fig. 3). In hDHFR, helix αF slides 2.5 Å toward the active site in the hinge-open conformation, which is stabilized by hydrogen bonds between NADP and Ser119 at the N terminus of helix αF that are not formed in the hinge-closed conformation. In ecDHFR, the shorter hinge forms hydrogen bonds involving exclusively backbone atoms, thus limiting flexibility and restricting αF movement.

Although the detailed molecular mechanism is not yet fully understood, the long hinges in hDHFR appear to provide flexibility that allows the adenosine-binding and loop subdomains to rotate away from each other in the hE–NADPH complex, thereby opening the active site cleft and facilitating substrate access. Subdomain rotation is limited in DHFRs with shorter hinges, such as ecDHFR, in which ligand flux is facilitated by flexible active site loops. Mutations that perturb the conformational fluctuations of the active site loops in ecDHFR result in altered ligand flux, thus suggesting that either flexible loops or long hinges are required for optimal function. Consistent with this finding, the sequence patterns in regions A, B and C indicate that, whereas some DHFRs contain both flexible loops (seven residues in region A) and long hinges, a few archaeal enzymes contain both a rigidified active site (eight residues in region A) and short hinges (Fig. 5); these enzymes are predicted to exchange ligands inefficiently.

**Enzymes with intermediate dynamic mechanisms**

Although bacterial and human DHFRs have similar three-dimensional structures, their intrinsic dynamic properties are quite different and are associated with important differences in their catalytic mechanisms. We can only speculate as to the mechanism of an ancestral DHFR, but it is natural to expect that enzymes with intermediate mechanisms existed at some point in the evolution of the DHFR family, either as the ancestral enzyme or as intermediates between the ancestral enzyme and the other extant DHFR. DHFRs with seven residues in region A (E. coli like) and long, hDHFR-like hinges are found in various invertebrate animals, fungi and members of the flagellate unicellular eukaryotes Euglenozoa (Fig. 5), and these potentially represent enzymes with such intermediate dynamic and mechanistic characteristics. For example, ceDHFR displays dynamic characteristics that are a mix of hDHFRs and ecDHFRs. Like ecDHFR, the C. elegans protein samples different ground-state conformations across the hydride-transfer step (Fig. 4), but 15N relaxation dispersion experiments show that millisecond-timescale motions are absent, as for hDHFR. Unfortunately, ceDHFR structures are not...
Figure 6 Human DHFR cannot complement DHFR-knockout E. coli cells and is more sensitive to product inhibition than is E. coli DHFR. (a) Differential interference contrast micrographs of MG1655 ΔfolA and DHFR knock-in strains temporarily grown with or without thymidine, after initial growth in media supplemented with thymidine. Scale bars, 10 μm. Images were obtained with the open-source microscopy software μManager. (b) Relative plating efficiency of MG1655 ΔfolA and DHFR knock-in strains on LB medium with or without 100 μg/ml thymidine. Plating efficiency for each strain on LB with thymidine is normalized to 1. The mean plating efficiency is reported here, with error bars indicating the s.d. (n = 3 biological replicates) (c) Initial kinetic rates for ecDHFR (black), hDHFR (red) and E. coli N23PP S148A mutant (blue) enzyme activity, plotted as a function of increasing NADP⁺ concentrations. The IC₅₀ for human DHFR is 948 μM, for ecDHFR, 6518 μM and for the mutant N23PP S148A ecDHFR 1,274 μM. The experiment was carried out in duplicate, and values for the mean initial rates and range are shown.

DHFRs may be tuned to cellular ligand concentrations

Why have DHFRs evolved different dynamics? We hypothesized that the divergent dynamics of ecDHFR and hDHFR may represent important functional adaptations to the disparate intracellular conditions in bacterial versus mammalian cells. To test this hypothesis, we assessed the ability of hDHFR to complement a folA-null allele in the wild-type E. coli strain MG1655 (Online Methods). We replaced the folA open reading frame with that of hDHFR and found that expression of soluble hDHFR fails to support cell proliferation. Growth of the resulting ΔfolA::DHFR strain was strictly dependent upon the addition of exogenous thymidine to the culture medium, similar to that of a complete DHFR knockout (ΔfolA) (Fig. 6a,b). Subsequent culture of the ΔfolA::hDHFR strain in the absence of supplemental thymidine (after initial culture with thymidine) resulted in the arrest of cytokinesis and rapid filamentation (Fig. 6a), results reminiscent of ‘thymineless death’ in thymidylate synthase (thyA) mutants. One possible explanation is that the highly divergent human DHFR (28% sequence identity with ecDHFR) is unable to make essential interactions with other E. coli proteins. However, S. aureus DHFR (only 35% identity with ecDHFR) fully complements the ecDHFR knockout (ΔfolA::saDHFR) (Fig. 6a,b), thus suggesting that differences intrinsic to hDHFR limit its efficiency in a bacterial cell, even though its hydride-transfer rate is faster than that of ecDHFR.

Vertebrate DHFRs function at much lower cellular concentrations of NADP⁺ (~20 μM versus ~2 mM) and THF (~0.3 μM versus ~13 μM) than does the E. coli enzyme. The ratio of NADP⁺ to NADPH differs greatly between mammalian (~100:1) and E. coli (~1:1) cells. Consequently, the E. coli enzyme may have evolved a dynamic mechanism that allows it to avoid end-product inhibition due to the high concentrations of NADP⁺ in bacterial cells. In contrast, human DHFR was not subjected to the same evolutionary pressure, as the large excess of NADPH over NADP⁺ in mammalian cells favors efficient exchange of NADPH for NADP⁺. To assess whether the human and E. coli enzymes are differentially susceptible to end-product inhibition by NADP⁺, we monitored the effect of increasing concentrations of NADP⁺ on the initial catalytic rate for both ecDHFR and hDHFR. As predicted, hDHFR is ~10 times more sensitive to inhibition by NADP⁺ (half-maximal inhibitory concentration (IC₅₀) ~620 μM for hDHFR versus ~5 mM for ecDHFR) (Fig. 6c). Thus, at concentrations of NADP⁺ and NADPH similar to those found in E. coli, ecDHFR retains near-maximal activity, whereas hDHFR is strongly inhibited. Additionally, the N23PP S148A mutant in which the Met20 loop is stabilized in the closed conformation shows increased NADP⁺ inhibition compared to that of ecDHFR, almost to the same extent as for hDHFR (Fig. 6c). This confirms that the increased product inhibition of hDHFR and N23PP S148A ecDHFR can be largely attributed to the altered dynamics of the Met20 loop, which remains in the closed position.

DISCUSSION

With new insights in hand, we are poised to begin addressing why hDHFR cannot complement DHFR-deficient E. coli cells. We propose that differences in the dynamic mechanisms of the two enzymes are responsible, at least in part, for the inability of hDHFR to function efficiently in the environment of an E. coli cell. In ecDHFR, formation of product is accompanied by a shift in the ground-state conformation from a partially occluded, with concomitant expulsion of the nicotinamide ring from the active site. This process is dynamic, with conformational fluctuations between the occluded and closed states and exchange of nicotinamide out of and into the active site occurring at a rate of 1,300 s⁻¹ (refs. 16, 25). The transition to the occluded ground-state conformation in the E–NADP⁺–THF product complex of ecDHFR is intimately linked to NADP dissociation, and mutations that perturb the flexibility of the Met20 loop or alter the closed-to-occluded equilibrium perturb the NADP flux. In particular, mutations such as N23PP that ‘lock’ ecDHFR in the closed conformation and alter millisecond-timescale fluctuations of the Met20 loop greatly decrease the rate of NADP⁺ dissociation. Because vertebrate DHFRs function at much lower cellular concentrations of NADP⁺ and THF, they have evolved to bind cofactor more tightly and release it more slowly, primarily by eliminating coordinated conformational fluctuations on the millisecond timescale and by stabilizing the active site loops in the closed conformation. The permanently closed conformation comes at the expense of slower or off rates for substrates and products and suggests that the exaggerated domain twisting might compensate for reduced ligand accessibility to the active site. Whereas hDHFR has higher activity, it is unable to rescue an E. coli cell when expressed under the endogenous promoter because the concentrations of NADP⁺ and THF in E. coli...
are inhibitory. Thus, our data suggest that protein dynamics in DHFR have a major role in tuning the catalytic mechanism and ligand flux and have evolved to allow optimal enzyme function in a given cellular environment.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Coordinates and structure-factor amplitudes have been deposited in the Protein Data Bank under accession codes 4M6J (hDHFR-NADPH), 4M6K (hDHFR-NADP+-FOL) and 4M6L (hDHFR-NADP+-dTHF).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We gratefully acknowledge M. Yamout for assistance with design and preparation of DHFR expression constructs, G. Johnson for assistance with figure preparation, X. Zhang (Scripps Research Institute) for providing anti-hDHFR antibody, D. Boger (Scripps Research Institute) for kindly providing dTHF, X. Dai for assistance with crystallography data collection, M. Mettlen for assistance with microscopy and J. James and E. Jonsson for assistance with analysis of kinetic data. This work was supported by the US National Institutes of Health (NIH) grant GM75995 and the Skaggs Institute of Chemical Biology (P.E.W.). G.B. is supported as the Merck Fellow of the Damon Runyon Cancer Research Foundation (DRG-2136-12). D.C.E. is supported as a Damon Runyon Fellow by the Damon Runyon Cancer Research Foundation (DRG-2140-12). D.C.E. was supported by a predoctoral fellowship from the Achievement Rewards for College Scientists Foundation, grant GM080269 from the NIH Molecular Evolution Training Program. The Joint Center for Structural Genomics is supported by NIH National Institute of General Medical Sciences (NIGMS) (U54 GM094586). C.M.Z. was supported by NIH National Institute of General Medical Sciences (NIGMS) (U54 GM094586). C.M.Z. was supported by NIH grant GM101457. The GM/CA CAT 23-ID-D has been funded in whole or in part with federal funds from the US National Cancer Institute under contract DE-AC02-06CH11357. C.M.Z. was supported by NIH grant GM75995 and the Skaggs Institute of Chemical Biology (P.E.W.). G.B. is Supported by the US National Institutes of Health (NIH) grant GM094586. C.M.Z. was supported by NIH grant GM101457. The GM/CA CAT 23-ID-D has been funded in whole or in part with federal funds from the US National Cancer Institute under contract DE-AC02-06CH11357.

AUTHOR CONTRIBUTIONS

G.B. and P.E.W. designed the research. G.B., M.J. and G.K. collected the data. G.B., D.C.E., C.M.Z., A.G., L.M.T., H.J.D., I.A.W. and P.E.W. analyzed the data. All authors contributed to writing the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

Cloning, expression and purification of DHFR. Wild-type human DHFR (hDHFR) was cloned into a pET21a vector and transformed into E. coli BL21 (DE3) (DNAY) cells for expression. Cells were grown in M9 minimal medium containing 2 mM folic acid, 0.5 g/L [15N] ammonium sulfate, 0.5 g/L [13C, 15N] ammonium chloride, 3 g/L [13C] glucose or deuterium oxide for uniform labeling. [12C, 13N] or [15H, 13C, 15N] uniformly labeled samples were used for triple resonance experiments. Cells were harvested after ~24 h by centrifugation at 3,000g for 15 min and frozen at ~20 °C. Frozen cell pellets can be stored for several weeks at ~20 °C. hDHFR was purified, and the desired complex was prepared by refolding as previously described37.

Wild-type reference sequences (as listed in the NCBI database) were synthesized by the company Mr. Gene for DHFR from the following species: B. anthracis, S. aureus, S. pneumoniae, V. cholerae, S. scrofa, B. taurus, D. rerio, C. elegans and R. norvegicus. Constructs were codon-optimized for expression in E. coli and contained an N-terminal His6 tag. Constructs were cloned into the pET15b vector and expressed in the host E. coli BL21 (DE3) (DNAY). Cells were grown at 37 °C in M9 minimal medium containing NADP+ and THF.

pET15b vector and expressed in the host E. coli BL21 (DE3) (DNAY). Cells were grown at 37 °C in M9 minimal medium containing [15N] ammonium sulfate and [15N] ammonium chloride, induced with 1 mM IPTG at an OD600 of ~0.9 at 15 °C and harvested after ~24 h by centrifugation at 3,000g and 4 °C for 15 min. Cell pellets were frozen at ~20 °C and can be stored up to several weeks.

Cell pellets from 2 l of cell culture were resuspended in 80 mL lysis buffer containing 50 mM Tris, pH 8.0, 200 mM NaCl, 1 mM folic acid, 10 mM imidazole and one EDTA-free protease inhibitor cocktail tablet (Roche) and lysed by two passes through an EmulsiFlex C-3 cell disruptor (Avestin). Affinity purification was carried out with Ni-NTA resin (Qagen), and protein was eluted with 50 mM Tris, pH 8.0, 200 mM NaCl, 1 mM folic acid, 250 mM imidazole. Fractions containing DHFR were further purified by reversed-phase HPLC and refolded as previously described37. DHFR from each species was divided into two aliquots; one was refolded with NADP+ and FOL, whereas the other was refolded with NADP+ and THF.

21-PWNL-24 cDHFR was expressed and purified by anion exchange and gel filtration as previously described37.

Preparation of NMR samples. NADP+, NADPH and folic acid were purchased from Sigma. (6S)-THF was purchased from Schircks Laboratories. Folic acid is light sensitive, and THF and NADPH are both light and oxygen sensitive and need to be treated accordingly. Buffer was extensively degassed with freeze-pump-thaw cycles on a vacuum line, after which 1 mM DTT was added. Protein was exchanged into the degassed NMR buffer (50 mM potassium phosphate, pH 6.5 or pH 8.0, 50 mM KCl, 1 mM EDTA, 1 mM DTT and 0.02% NaN3) in an inert environment in a glove box with a NAP-5 column, and fresh ligands were added in ten-fold excess. All spectra were recorded at pH 6.5 except for the hE–NADP+–FOL complex, for which data were acquired at pH 8.0. The samples were placed in amber NMR tubes with a vacuum-line adaptor, subjected to gentle vacuum, overlaid with argon and flame-sealed to prevent oxidation of ligands.

NMR experiments. The hE–NADP+–FOL and hE–NADP+–THF complexes are stable and could be concentrated to ~1 mM for NMR experiments. Owing to the low yield after refolding and instability of the complexes, hE–NADPH and hE–FOL or hE–THF were concentrated to ~300 µM and 500 µM, respectively, for NMR experiments. Standard 3D HNCA, HNCOCA, HNCCAB and HNCOCACB triple resonance experiments were used to assign the spectra of all complexes formed with either uniformly [1H, 12C, 15N]-labeled protein or [13C, 15N] uniformly labeled protein. All NMR spectra were processed with NMRPipe and analyzed with NMRView.

15N R2 relaxation dispersion for the 21-PWNL-24 mutant of cDHFR bound to NADP+ and FOL was measured at 500 MHz and 750 MHz with constant-time CPMG experiments, as described previously38,39. Dispersion measurements for hDHFR were made at 800 MHz and 599 MHz, and at 800 MHz for C. elegans DHFR. All data were collected at 300 K with a total relaxation period (TPPMD) of 40 ms. Data were fitted as previously described38,39, with an in-house fitting program, GLOVE40.

Off-resonance 15N R2 relaxation dispersion experiments for the E–NADP+–FOL complex of hDHFR were carried out at 280 K with 900 MHz and 750 MHz spectrometers. A TROSY version of the pulse sequence described in ‘scheme 2’ of Massi et al. was used41. Rp, relaxation rate constants for each residue were determined by acquiring a series of two-dimensional spectra with different relaxation delays, from 6 to 160 ms. Intensities of cross-peaks were fitted with a two-parameter exponential decay function to give relaxation rate constants. Spectra were processed and analyzed with NMRPipe, NMRView and Curvedit.

Ten different effective fields were obtained with a spin-lock field of 1,000 Hz and different resonance offsets. On average, for each residue a maximum effective field (ν2) of between 4 and 6 was obtained. R1 and R2 values were measured at 280 K with established inversion recovery and CPMG experiments42. Maximum relaxation delays of 2,000 ms and 240 ms were used for R1 and R2, respectively. Heteronuclear NOE measurements were made in triplicate, with established methods43,44. R3 values were calculated and used instead of R eff values directly according to the following relationship: R eff = R1 cos2 θ + R3 sin2 θ, where θ is the tilt angle between the static magnetic field and the effective field in the rotating frame. Dispersion curve fitting was carried out as described by Massi et al.44, with the fitting program, GLOVE40.

Crystallography. Purified hDHFR was concentrated to ~12 mg/mL, exchanged into buffer containing potassium phosphate (50 mM, pH 7.5) and DTT (1 mM), and an approximately three-fold excess of the desired ligand (NADP+, NADPH, FOL and/or ddTHF) was added. Initial crystallization trials were performed with the automated Rigaku Crystalization robotic system at the Joint Center for Structural Genomics (JCSG, http://www.jcsg.org/). Crystals were optimized in sitting drops by setting up a fine screen around the conditions of the original hit, with 0.5 µL prot + 4.5 µL precipitant per drop and varying pH and precipitant concentrations. For the E–NADP+–FOL complex, no optimization was necessary, and data were collected on a crystal obtained directly from the robotic screen. Crystals were cryoprotected in the reservoir solution supplemented with ~30% glycerol, then flash cooled and stored in liquid nitrogen until data collection. All diffraction data were collected at the Advanced Photon Source (APS) General Medical Sciences and National Cancer Institute Collaborative Access Team (GM/CA-CAT) beamlines 23ID-B or 23ID-D.

hE–NADP+–FOL was crystallized in 2 mM ammonium sulfate and 100 mM phosphate-citrate, pH 4.2 at 4 °C. The 1.4-Å-resolution data set was indexed and integrated in space group P6322. The hE–NADP+–FOL complex was crystallized in 24% PEG 6000 and 100 mM Tris, pH 9.5 at 4 °C, and data were collected to 1.2-Å resolution and indexed and integrated in space group C2221. The hE–NADP+–ddTHF complex was crystallized in 2.2 M ammonium sulfate and 100 mM sodium citrate, pH 4.0 at 4 °C. A 1.7-Å-resolution data set was indexed and integrated in space group P6522. Initial data processing was done with HKL2000 (HKL: Research) and merged with Xprep (Bruker). Data collection statistics for all structures are summarized in Table 1.

The structures were solved by molecular replacement with Phaser with PDB 1DSL45 (hE–NADPH–MTX, L22Y variant) as a search model for hE–NADP+–FOL and with the hE–NADP+–FOL structure as a search model for the other two structures. Ligands were removed from all the search models before molecular replacement. One copy of the DHFR complex was found in the asymmetric unit for all structures. Rigid-body, restrained refinement and simulated annealing were carried out in Phenix46. After manual adjustment of the model in Coot47, including ligand placement and the addition of waters, two TLS groups were defined for hE–NADP+–FOL, and hE–NADP+–ddTHF restrained TLS refinement of the structure was completed with Phenix version 1.7. Anisotropic ADP refinement was carried out for the hE–NADP+–FOL structure at 1.2-Å resolution. The structures were validated with the Quality Control Check v2.7 developed by the JCSG, which included MolProbity48 (http://smb.slc.stanford.edu/jcsg/QCC). Final refinement statistics are shown in Table 1. The wavelength of data collection was 0.980 for hE–NADPH, and 1.033 for hE–NADP+–FOL and hE–NADP+–ddTHF. Ramachandran statistics are 98.5% favored, 0.0% disallowed for hE–NADPH and hE–NADP+–FOL, and 98.9% favored, 0.0% disallowed for hE–NADP+–ddTHF.

Model building. Strong, clear electron density was observed for all ligands in each structure (Supplementary Fig. 1), with the exception of NADP+ in the hE–NADP+–ddTHF structure (Supplementary Fig. 1f). The hE–NADP+–FOL and hE–NADP+–ddTHF complexes crystallized under similar conditions and in the same space group (P6522) and diffracted to similar resolutions. However, whereas the electron density for NADP+ is clear and well defined in the hE–NADP+–FOL structure (Supplementary Fig. 1c), it is weaker and discontinuous...
in the hE–NADP+–ddTHF structure, thus indicating disorder or low occupancy of the ligand. For the comparable ecDHFR structure (PDB 1RX4 (ref. 18)), the electron density for the cofactor is well defined except for the nicotinamide ring, which resides outside the active site and for which no electron density is observed. Electron density that would place NADP+ outside the active site (as in ec–NADP+–ddTHF) was not observed in the hE–NADP+–ddTHF electron density map. In both the hE–NADP+–ddTHF and hE–NADP+–FOL structures, three-fold excess NADP+ was added to the sample. Electron density is observed in the NADP-binding site in hE–NADP+–ddTHF; however, in the 2F–F map contoured at \( \sigma = 1 \), the electron density is broken in the region of the adenosyl ribose, despite clear density for the phosphates on either side (Supplementary Fig. 1f). In order to assess whether the observed electron density represents NADP+ or molecules of free buffer components, we modeled in NADP+ and, in parallel, carried out refinement of the structure in which we modeled in two phosphates, two glycerol molecules and waters to fit the electron density as best as possible in the absence of NADP+. The density is satisfied better with NADP+ modeled into the active site, with the occupancy refined to \( -0.8 \) (Supplementary Fig. 1g,h). Given the strong density for the phosphate groups, we conclude that NADP+ is bound but is disordered in several regions. The B values of the NADP+ ligand vary greatly, ranging from 24 Å² for the more ordered phosphate to 61 Å² in the adenosyl ribose. The lower B values of the NADP+ are comparable to those in the surrounding regions of the protein, consistent with the notion that the ligand is present but not well ordered. These structures suggest that at low pH (4–4.5), NADP+ is well ordered in the ternary Michaelis model complex, hE–NADP+–FOL, but is disordered in the product ternary complex, hE–NADP+–ddTHF, for which the next step is cofactor or product release. hDHFR was cocrystallized with 5,10-dideazaterahydrofolate. Although the configuration of the C6 stereocenter was unknown in the starting material (potentially 6R, 6s or a racemic mixture), the electron density in the DHFR folate-binding site indicates that the 6S enantiomer is the predominant form bound in the crystal structure (Supplementary Fig. 1j). As our hE–NADPH crystal structure is in a different space group from the other structures and at a different pH, we supplemented the crystal structure analysis with solution NMR data (Supplementary Fig. 3d and Supplementary Note) in order to elucidate the conformational changes in the human enzyme as it progresses through the catalytic cycle. The two subdomains in hDHFR were identified and defined with DynDom49.

Bioinformatic analysis. –1,800 DHFR sequences deposited in the UniProt database50 were aligned with MUSCLE51. In all of our analyses, DHFR-thymidylate synthase (DHFR-TS) enzymes were excluded, as these enzymes represent a distinct evolutionary lineage, and the dynamics of these bifunctional enzymes have not been extensively characterized. In addition, a separate analysis was carried out in which eukaryotic DHFR sequences from 172 completely sequenced eukaryotic genomes were analyzed by alignment of their DHFR domains (as defined by the Pfam HMM model of the DHFR family, PF00186) with MAFFT52 followed by phylogenetic inference with the minimal evolution approach implemented in FastME53. The details of the analysis are as follows. Protein predictions for 172 completely sequenced eukaryotic genomes were obtained from a variety of sources (details in Supplementary Table 3). The domain repertoire for each genome was determined by hmmscan (with default options, except for an E-value cutoff of 2.0 and ‘no bias’) from the HHMNER 3.0 package (http://hmmer.janelia.org/) with hidden Markov models from the Pfam database. In a second step, the hmmscan results were filtered by the domain-specific ‘gathering’ (GA) cutoff scores provided by Pfam. From this set of analyzed proteins, we selected those that contained DHFR domains for phylogenetic analysis. The extent of the DHFR domain for the purpose of multiple alignments was defined by the Pfam 25.0 model of the DHFR domain. Multiple sequence alignments were then produced by MAFFT 6.240 (localpair, maximxate 1000)54. Multiple sequence alignment columns with a gap in more than 50% of sequences were deleted and not used in further analysis. Phylogenetic trees were then calculated with FastME 1.1 (ref. 53). All sequence, alignment and phylogeny files are available upon request.

Complementation assays. A ΔfolA MG1655 strain was generated with recombineering, essentially as previously described34. The folA region was replaced with a Kanamycin-resistance cassette. In other strains, the folA gene was replaced with genes expressing wild-type E. coli DHFR (folA, control), human DHFR or S. aureus DHFR. Wild-type MG1655 E. coli cells and ΔfolA, ΔfolA:ecDHFR, ΔfolA:hdDHFR or ΔfolA:saDHFR strains were grown in media supplemented with 100 µg/mL thymidine, gently pelleted, resuspended in LB medium without thymidine and plated either on plates containing LB plus 100 µg/mL thymidine or LB only. Plates were incubated at 37 °C overnight, and colonies were counted the following morning to determine plating efficiency. Each experiment was performed in triplicate. For ΔfolA:chDHFR and ΔfolA strains, zero colonies grew in the absence of thymidine; therefore, errors could not be determined for these samples. For filamentation assays, strains were grown in LB medium supplemented with 100 µg/mL thymidine (Sigma) until an OD600 of ~0.6 was reached. Cells were gently pelleted (~2,000 r.p.m., 5 min) and washed twice in LB medium to remove traces of thymidine. Cells were diluted to an OD600 ~0.1, and grown at 37 °C, 250 r.p.m. for 4 h. Cells were fixed directly in the culture medium by the addition of paraformaldehyde to a final concentration of 4% and incubated at room temperature for 30 min. 10 µL of each cell suspension was spotted on polysine-coated coverslips and mounted in 80% glycerol. Samples were imaged on an inverted fluorescence microscope (Olympus model IX71) with a PlanApo N 60X (NA 1.42; Olympus) objective, equipped with a Hamamatsu Photons ORCA-ER camera (model C4742-80-12AG). The open-source microscopy software µManager55 was used to control image acquisition.

To assess the solubility of human DHFR, we used western blotting. Endogenous DHFR is expressed at very low levels, and hDHFR expression is nearly undetectable in human cell lysates unless the cells are transfected with a hDHFR-expressing plasmid. Consequently, before western blotting with an anti-hDHFR antibody (Abnova, cat. no. H0001719-M01; validation provided on the manufacturer’s website), soluble fractions of MG1655 ΔfolA:chDHFR and MG1655 ΔfolA lysates were enriched in His-tagged proteins by binding to Ni-NTA resin (Qiagen). After elution, the Ni-enriched soluble fractions were subjected to western blotting with standard methods.

Activity assays. All kinetic measurements were made at pH 7.0 in buffer containing 50 mM potassium phosphate, 100 mM NaCl and 2 mM DTT at room temperature. The enzyme at 400 nM was pre-equilibrated with 500 µM NADPH in a 100-µL reaction volume for 5 min in a 96-well plate. The reaction was initiated by addition of DTF to a final concentration of 200 µM. Initial rates were calculated from the change in absorbance at 340 nm, which was monitored immediately after addition of DTF for 120 s with a cycle time of 10 s. Reaction rates were measured in the presence of 0, 5, 50, 500, 5,000 or 50,000 µM NADP+ to assess product inhibition. All experiments were done in duplicate and were interleaved to minimize the effect of potential degradation of the unstable ligands, which were incubated on ice for the duration of the experiment. Control experiments were carried out in which each protein or ligand was omitted from the reaction, and these resulted in rates of ~0. Data were fit with MATLAB.

37. Bhabha, G., Tuttle, L., Martinez-Yamout, M.A. & Wright, P.E. Identification of endogenous ligands bound to bacterially expressed human and E. coli dihydrofolate reductase by 2D NMR. FEBS Lett. 585, 3528–3532 (2011).
38. Loria, J.P., Rance, M. & Palmer, A.G. III. A relaxation-compensated Carr–Puercell–Meiboom–Gill sequence for characterizing chemical exchange by NMR spectroscopy. J. Am. Chem. Soc. 121, 2331–2332 (1999).
39. McElheny, D., Schnell, J.R., Lansing, J.C., Dyson, H.J. & Wright, P.E. Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. Proc. Natl. Acad. Sci. USA 102, 5032–5037 (2005).
40. Sugase, K., Konuma, T., Lansing, J.C. & Wright, P.E. Fast and accurate fitting of relaxation dispersion data using the flexible software package GLOVE. J. Biomol. NMR 56, 275–283 (2013).
41. Massi, F., Johnson, E., Wang, C., Rance, M. & Palmer, A.G. III. Microsecond timescale backbone conformational dynamics in dihydrofolate reductase studied by 15N NMR relaxation. Proc. Natl. Acad. Sci. USA 102, 727–734 (1994).
42. Farrow, N.A. et al. Backbone dynamics of a free and a phosphohexose-precipitated S. cerevisiae folate 2 domain studied by 13N NMR relaxation. Biochemistry 33, 5984–6003 (1994).
43. Farrow, N.A., Zhang, O., Forman-Kay, J.D. & Kay, L.E. A heteronuclear correlation experiment for simultaneous determination of 1H and 13C exchange rates of systems in slow equilibrium. J. Biomol. NMR 4, 727–734 (1994).
44. Massi, F., Grey, M.J. & Palmer, A.G. III. Microsecond timescale backbone conformational dynamics in ubiquitin studied with NMR relaxation experiments. Protein Sci. 14, 735–742 (2005).
45. Lewis, W.S. et al. Methotrexate-resistant variants of human dihydrofolate reductase with substitutions of leucine 22: kinetics, crystallography, and potential as selectable markers. J. Biol. Chem. 270, 5057–5064 (1995).
46. Adams, P.D. et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. 58, 1948–1954 (2002).
47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132 (2004).
48. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21 (2010).
49. Hayward, S. & Berendsen, H.J. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins 30, 144–154 (1998).
50. Punta, M. et al. The Pfam protein families database. Nucleic Acids Res. 40, D290–D301 (2012).
51. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797 (2004).
52. Katoh, K. & Toh, H. Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics 26, 1899–1900 (2010).
53. Desper, R. & Gascuel, O. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. J. Comput. Biol. 9, 687–705 (2002).
54. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. USA 97, 6640–6645 (2000).