Adaptive Landscapes of *Plasmodium Falciparum* Dihydrofolate Reductase Reveal Pathways to Antifolate Resistance

Authors

Danny Muzata[1,2], Dwipanjan Sanyal[3], Deeptanshu Pandey[4], Soumyananda Chakraborti[4], Vladimir N. Uversky[5], Prajna Upadhyay[2], and Sourav Chowdhury[4]*

Affiliations

[1]One Health Trust, Bangalore, India
[2]Department of Computer Science and Information Systems, Birla Institute of Technology and Science, Pilani, Hyderabad Campus, Jawahar Nagar, Kapra Mandal, Medchal District, Telangana 500078, India
[3]Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, Kolkata, India
[4]Department of Biological Sciences, Birla Institute of Technology and Science-Pilani, Hyderabad, India.
[5]USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA

***Corresponding Author:**
Sourav Chowdhury
Department of Biological Sciences,
Birla Institute of Technology and Science, Hyderabad Campus,
Hyderabad, Telangana, India, 500078
Email: sourav.chowdhury@hyderabad.bits-pilani.ac.in

Abstract

Antifolate resistance in *Plasmodium falciparum* dihydrofolate reductase (pfDHFR) remains a major challenge for malaria control. To understand how this enzyme maintains function under antifolate selection, we developed PfPATH, a computational framework that integrates mutational fitness measurements, evolutionary interactions, and structural information to map adaptive trajectories. PfPATH adaptive walks reveal that evolution is constrained to a narrow ridge in sequence space defined by a small subset of residues. Most resistance mutations are individually deleterious and become viable only when supported by stabilizing changes, restricting adaptation to a few high-resistance pathways. Structure network analyses show that these mutations do not disrupt the global enzyme architecture but instead reorganize internal communication to preserve catalytic function. Consistent with this, molecular dynamics and free-energy analyses indicate that resistant variants remain stable while sampling a broader ensemble of low-energy conformational substates. Together, these results reveal a narrow and predictable route to antifolate escape.

Keywords: malaria, plasmodium falciparum, dihydrofolate reductase, antifolate resistance, protein fitness

Introduction

Malaria remains one of the most challenging infectious diseases to control, partly because of the parasite's ability to rapidly evolve resistance to frontline drugs (1). The mutation of *Plasmodium falciparum* dihydrofolate reductase (*Pf*DHFR) is one of the most challenging resistance mechanisms that minimize the efficacy of antifolate treatments (2). Antifolate drugs have long been used not only against malaria but also against bacterial infections and cancer to target enzymes in the folate biosynthesis pathway and block tetrahydrofolate production which is a necessary cofactor in nucleotide synthesis and cell growth (3, 4). In *P. falciparum*, dihydrofolate reductase plays a central role in folate metabolism by enabling the synthesis of DNA precursors essential for parasite proliferation and survival (5–7). This central role makes it an attractive target for antifolate drugs. This enzyme plays a vital process in the production of purines, thymidylate, and some amino acids. The effectiveness of antifolate medications, such as pyrimethamine and cycloguanil, basically hinges on this critical function (3). However, resistance is often acquired through point mutations in the *pfDHFR* gene that reduces drug binding without interfering with enzymatic function (5, 8). Mutations at residue positions 16, 50, 51, 59, 108, and 164 have been shown to reduce antifolate binding affinity (9, 10). Furthermore, combinations of these point mutations confer high levels of resistance. The frequently observed N51I/C59R/S108N triple-mutant haplotype results in substantial pyrimethamine resistance and is often accompanied by additional substitutions, such as I164L, giving rise to quadruple-mutant variants with even greater resistance (2, 11). Therefore, resistance in *Pf*DHFR does not arise solely from single substitutions, but from an evolved network of amino acid changes that collectively regulate protein stability, catalytic kinetics, and inhibitor binding (10, 12). These stepwise adaptations allow the parasite to navigate fitness trade-offs between maintaining function and evading drug pressure. Such complexities emphasize the remarkable adaptive plasticity of *P. falciparum*.

The evolutionary potential of *Pf*DHFR mutations highlights the threat and the ongoing risk of antifolate-based interventions in the treatment and prophylaxis of vulnerable populations (13, 14). While resistance in *P. falciparum* may involve mechanisms beyond mutation alone, understanding how mutational constraints preserve enzyme activity is a critical step toward

understanding the resistance evolution. Although *Pf*DHFR-mediated resistance has been widely studied, the effects of resistance-associated mutations on enzyme structure, conformational dynamics and adaptive capacity remain poorly understood. A better way to frame this problem is to view the protein as a map of sequence space that encompasses all possible amino-acid variants. High-throughput mutational scanning and evolutionary studies have enabled the identification of both strongly constrained regions and more permissive, adaptable ones (15). This provides us with information on the evolutionary possibilities of the protein and its ability to respond to adaptive mutations. However, sequence-based approaches alone cannot fully explain the conformational plasticity and structural dynamics that are important for understanding adaptive protein fitness. Mutations can alter loop flexibility, active-site geometry, or long-range motions in ways that are not apparent from static sequence space analysis or crystallographic structures (16, 17). Hence, we combine evolutionary sequence analysis with all-atom molecular dynamics simulations to study proteins behavior at the atomic level. Molecular dynamics simulations provide a time-resolved perspective of the protein motions and reveal how enzymes search across protein conformations and how mutations influence those dynamics (18). In this study, we combine evolutionary sequence and structural features with all-atom molecular dynamics simulations of the apo wild-type *Pf*DHFR and known resistance mutants to construct a mechanistically informed fitness landscape. These complementary analyses form the basis of the *Pf*PATH framework, which links mutational fitness with conformational stability and dynamic constraints. Our approach provides unified evolutionary and biophysical perspective on antifolate resistance. We reveal how *Pf*DHFR adapts through constrained yet predictable mutational routes that may be interfered in the design of next-generation antifolate therapies.

Results

Evolutionary and Structural Features Inferred from Sequence Space

We investigated the sequence space of *Pf*DHFR to understand how changes in amino acids affect the protein's fitness and stability. This revealed important aspects of the protein's

evolutionary history, showing that mutations at specific sites have different effects on stability, function and mutational tolerance. We also found the relationships between sequence variability and functional implications, especially at sites related to antifolate resistance. We first assessed evolutionary conservation by calculating Shannon entropy across all residue positions. With a *P. falciparum*-specific multiple sequence alignment, we identified that about 86% of the positions were conserved with most of the rest variable sites, particularly at the flexible loop regions (Fig. 1A). Among the most variable residue positions were N42, C50, S99, and I155 (Tables S1 and S2), all of which are associated with pyrimethamine resistance. To place these patterns in a broader evolutionary context, we also performed entropy analysis across diverse organisms. This comparison revealed six universally conserved positions (i.e. 30, 32, 38, 39, 91 and 104) each with zero entropy (Fig. 1A). This implies that residues at these positions are likely under strong evolutionary constraint. This suggests that substitutions at these positions could potentially disrupt structural integrity or catalytic function (Fig. 3B). Furthermore, most active-site residues exhibited low entropy suggesting limited tolerance to substitution and pointing to their importance for inhibitor binding and enzymatic activity.

We then analyzed the flexibility of the *Pf*DHFR structure through six popular disorder predictors (20): PONDR® FIT (21), PONDR® VSL2 (22), PONDR® VL3 (22), PONDR® VLXT (23), IUPred Short (24), and IUPred Long (24) and generated its disorder profile. The result of these predictors was averaged to form a mean disorder profile that was used to give a consensus view of the protein's disorder propensity. Intrinsic disorder usually is associated with a set of regions that can accommodate mutations as observed in the Shannon entropy profile. This adds to the evolutionary plasticity DHFR protein. As we found, this protein is largely ordered with most residues having disorder scores below 0.5, which points to the protein being stable. However, a few intrinsically disordered regions (≥ 0.5) were identified at positions 1–24, 60–125, 199–203, and 214–219 by the PONDR® FIT, PONDR® VSL2 and PONDR® VLXT predictors. The PONDR predictors (particularly PONDR® VSL2) are considered one of the most accurate predictors of disorder (22). Therefore, these disordered regions suggest points of potential mutational tolerance that can contribute to the protein’s adaptability (Fig. 1B). Although Shannon entropy described the general conservation trends, it failed to give us information on how individual residues may evolve in a coordinated manner. To analyze these inter-positional dependencies, we calculated Mutual Information (MI). Mutual Information measures the co-

variation in pairs of residues and helps in determining where alterations could be compensatory to maintain structure or functional integrity. The MI heatmap (in Fig. 1C) indicated that there were several strong co-evolutionary signals, denoted by the dark blue spots, at the N-terminal, central region and the C-terminal end. Some of the high-MI pairs were mapped onto the 3D structure which illustrate both short-range contacts and long-range dependencies spanning distant structural elements. These strongly co-varying residues later appear interacting with critical functional regions (like mutation sites and active-site residues) as shown in the MI networks in Fig. 2.

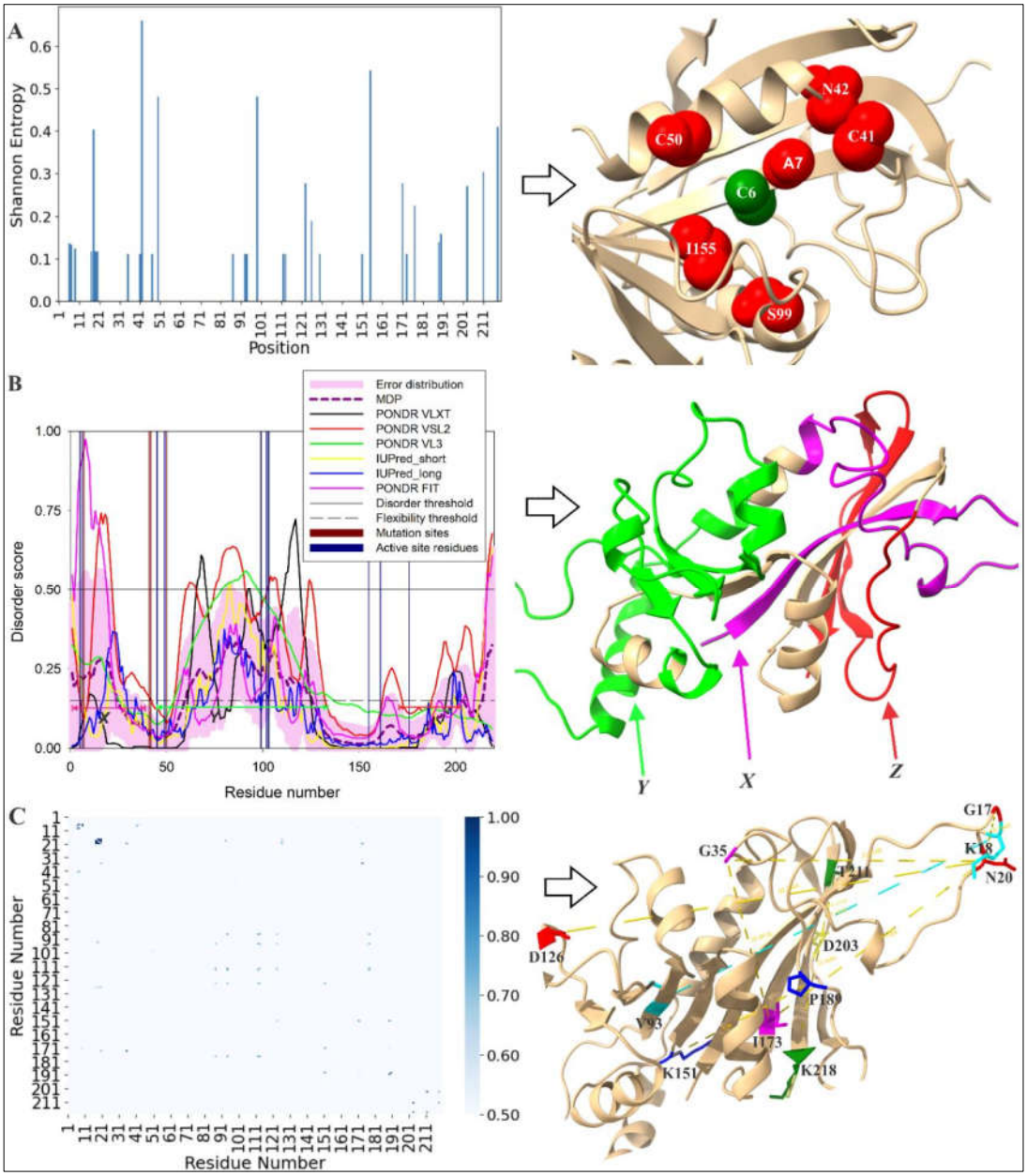

Fig. 1: Sequence-derived features of *Pf*DHFR highlight regions of variability, intrinsic flexibility, and co-evolutionary coupling. (A) Shannon entropy values calculated from the *P. falciparum*–specific alignment highlights mostly conserved regions with a few variable positions. Highly variable residues occur in the loop regions (C41 and N42). Residue positions such as 6 (an active site position), 7 and 155 map to structured element like β-strands and while 50 and 99 map to α-helices. Thus, sequence variability is not limited to flexible loops. **(B)** Per-residue intrinsic disorder was assessed using six

predictors: PONDR® FIT, PONDR® VSL2, PONDR®VL3, PONDR®VLXT, IUPred_short and IUPred_long. The mean disorder profile (MDP) captures their consensus. Residues with scores ≥0.5 were classified as disordered, whereas values between 0.15 and 0.5 indicate flexible but ordered regions. Although *Pf*DHFR is predominantly ordered, the insert highlights three segments that are relatively flexible labeled X (N-terminal region), Y (central region), and Z (C-terminal region) locally. **(C)** Mutual information (MI) analysis identified pairs of co-varying residues, with higher MI values (darker blue) indicating stronger evolutionary coupling. The insert maps represent high-MI pairs onto the 3D structure.

With the correlations seen in the MI heatmap, we further examined the local interaction networks formed around these individual mutational sites (Fig. 2). The networks revealed more pronounced and extended patterns of co-variation of these sites. They create an interaction hub with distributed residues across different regions in the protein structure. The respective 3D inserts also show that these interactions span both short-range and long-range distances. Hence, co-variation is not driven by spatial proximity. For example, the A7 interaction network (Fig. 2A) revealed co-varying partners, many of which sit in distant regions in the protein. It had the strongest correlation with residue C41, a position on which three resistance substitutions by Arginine, Tyrosine and Asparagine have been reported (8, 10). A similar interaction network was observed for R41 displaying wide-reaching interactions across both β-strand and loop regions. N42 and C50 (Fig. S4) even formed co-variation links with residues in the central catalytic core. In contrast, position I155 showed weaker but still notable interactions, suggesting a more modest degree of evolutionary coupling (Fig. S4). Table S3 summarizes the MI values and Cα–Cα distances for each network. Across the networks, mutation sites such as N42, C50 and I155 were linked to other mutation-associated positions and residues situated near the active-site region. This pattern suggests that resistance evolution in *Pf*DHFR may be partly coordinated across a distributed set of these interacting residues. The distances between co-varying residue pairs covered a wide range. This indicates that strong MI correlation does not depend on how close the residues are in 3D space, and that some residues can co-evolve via

long-range evolutionary coupling. Such residues could serve as intermediate or supporting positions in compensatory pathways.

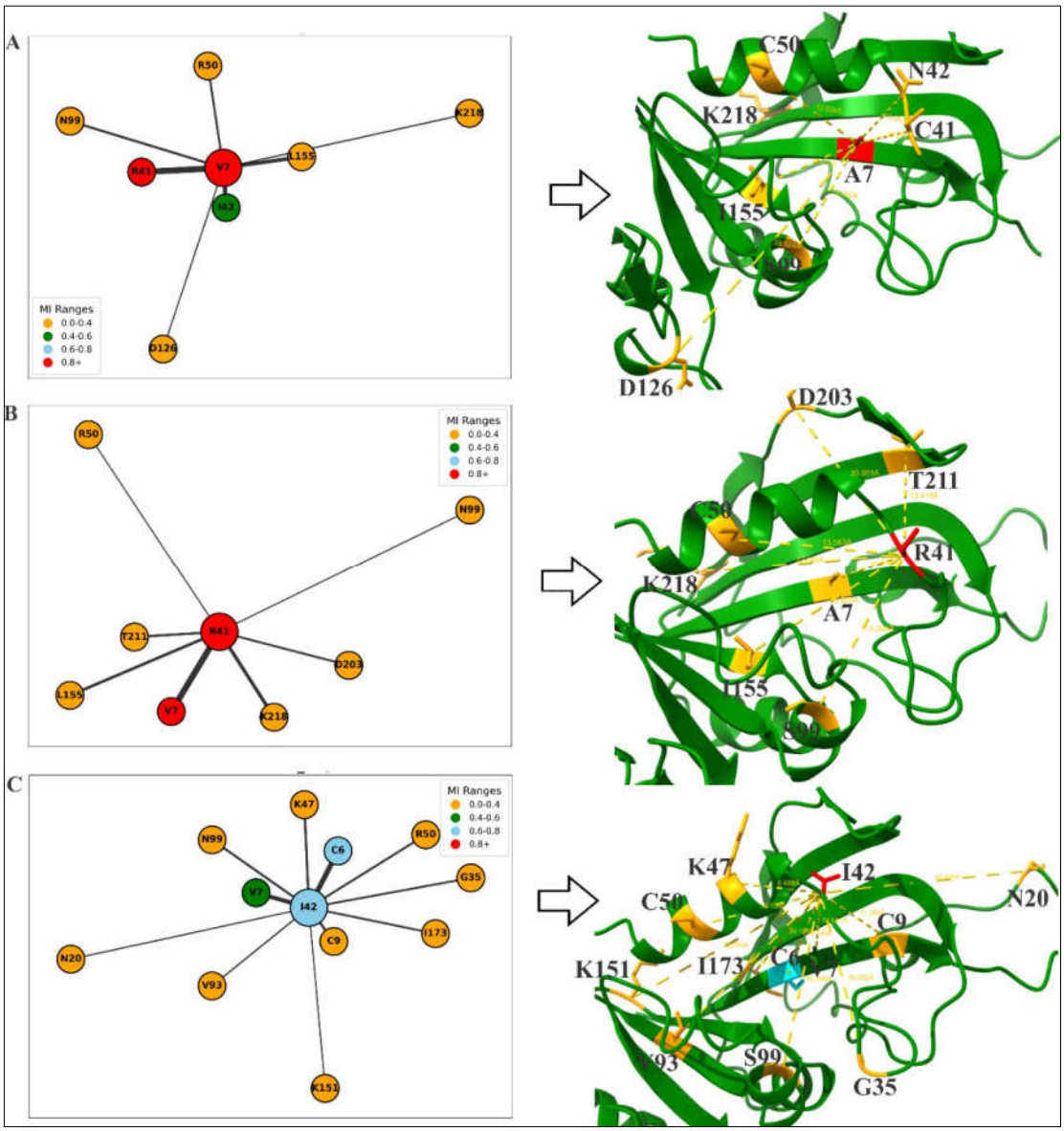

Fig. 2. Local Mutual Information networks centered on key *Pf*DHFR mutation sites reveal extensive co-variation across distant residues. (A) The network around mutation site A7 shows strong co-variation with multiple residues. The node colors reflect MI strength and the inserts on the right illustrate how these residues are distributed across β-strands, helices, and loop segments. **(B)** The R41 network revealed a wide interaction spread extending across the protein fold. Several co-varying partners overlap with the A7

network. **(C)** Mutation site N42 also exhibits moderate co-variation with residues distributed across the fold, including positions that lie close to the catalytic core. The insert shows these connections bridging distant regions of the protein.

Network-based analyses reveal compensatory hubs and shared structural modules linking active and mutation-prone residues in *Pf*DHFR

To understand how the co-variation patterns identified in the MI map are organized within the protein structure network, we analyzed the residue–residue correlation network using maximal clique analysis (25, 26) and community detection (27). These approaches describe the organization of the protein at two levels. Maximal cliques capture very tightly coupled micro-modules where all residues are connected to each other. The community detection captures larger dynamic regions that move together on average. These methods highlight which residues form local compensatory clusters and which belong to broader functional modules. In the maximal clique analysis, we identified 270 cliques across the *Pf*DHFR network. Several mutations and active-site residues were consistently located in the same dense cliques. Residues 41, 42, 50, 99 and 155, as well as active-site positions 5, 6, 161, 176, 102 and 103, belonged to the same large 8-member clique. The summary Tables S4 to S9 shows that positions 7, 99 and 155 appeared across dozens of cliques. These residues co-occur with active-site positions and with other mutation sites suggesting that they may help buffer the structural consequences of resistance-associated substitutions. For example, residues 42 and 50 co-occurred with active-site residues 45 to 49 in several cliques. These patterns support the interpretation that *Pf*DHFR leverages locally connected clusters of residues to maintain catalytic geometry while tolerating destabilizing changes. Community detection further divided the residue network into nine larger sub-structures. Mutation and active-site residues again appeared together in several communities. Residues 41, 42 and 50 are grouped with active-site residues 45, 46, 48 and 49 in Community 5 while residues 99, 102 and 103 clustered in Community 1. These communities represent broader segments of the protein that exhibit coordinated motion. All these analyses show that *Pf*DHFR resistance evolution is supported by both fine-scale and larger-scale structural coordination.

Deep mutational scanning revealed that *Pf*DHFR is globally constrained yet locally flexible, with conserved core and catalytic residues displaying strong intolerance to substitutions

To quantify the functional consequences of amino acid changes across the sequence, we carried out a Deep Mutational Scanning (DMS) (28). This technique measures the fitness impact of every possible single-residue substitution. This allows us to map the mutational landscape of the protein and distinguish regions that are highly constrained from those that tolerate changes. In the heatmap (Fig. 3C), beneficial substitutions are shown in shades of blue, whereas deleterious ones appear in red. The results show that different positions vary widely in how much mutation they can tolerate. At residue position D1 for example, most substitutions are beneficial. This means substitutions at this position can accommodate the changes without compromising function. However, substitutions at the neighboring positions were predominantly deleterious. Distinct regions of mutational intolerance were also observed across residues 30–53 and 90–196, whereas residues 54–90 (a loop–helix segment) exhibited comparatively higher mutational permissiveness. Two positions, 156 and 157, showed extreme intolerance. Nearly all substitutions at these sites were deleterious and this was consistent with our Shannon entropy analysis which identified them as highly conserved sites. Active-site and resistance-associated positions also followed expected patterns. Residue 7 was strongly deleterious across most substitutions, while residues 41 and 42 exhibited beneficial profiles. Conversely, residue 99 and residue 155 were strongly deleterious for most substitutions. The apparent contradiction between the fitness costs at these positions and natural occurrence of resistance mutations at the same sites could be explained by the compensatory or buffering mechanisms as observed in our MI network analysis (29).

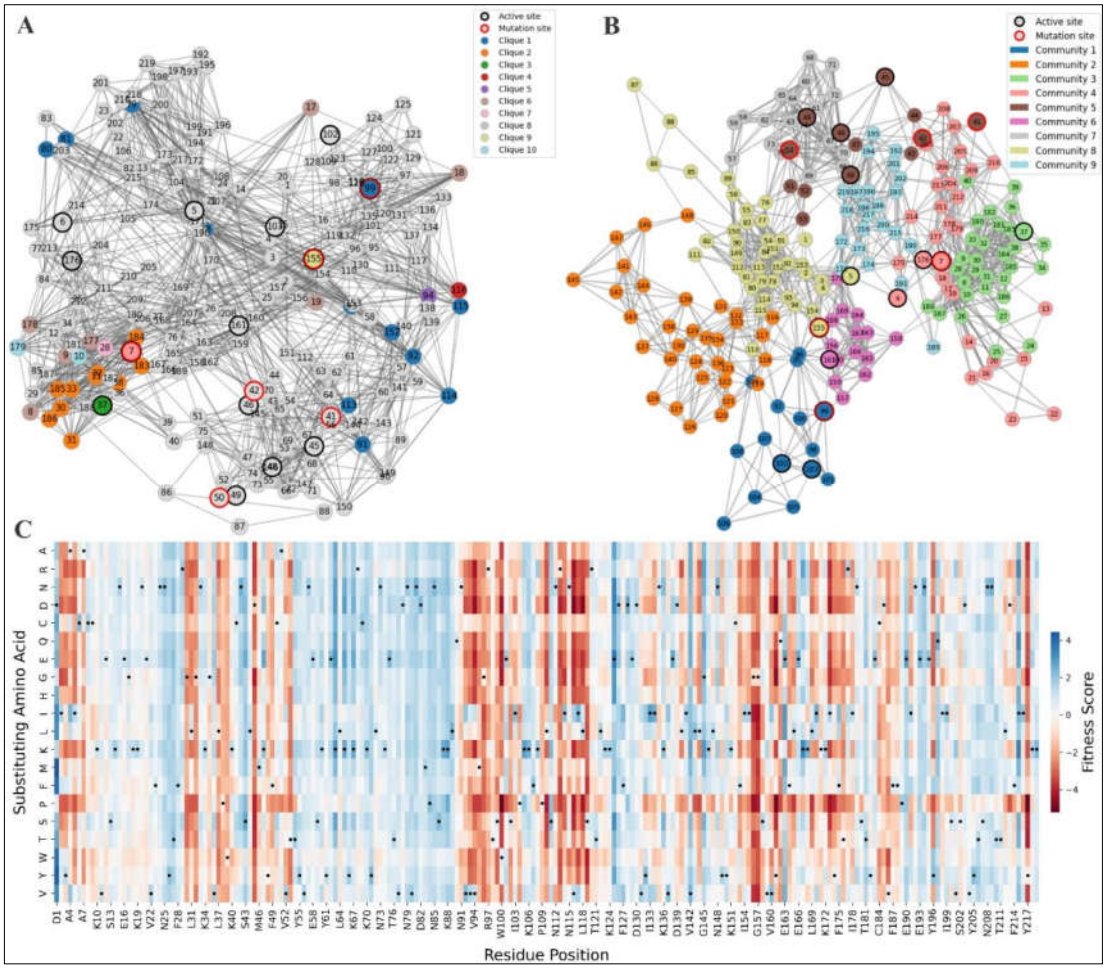

Fig. 3: **Network topology and mutational scanning highlight the structural basis of *Pf*DHFR adaptability. (A)** Maximal clique analysis shows the top ten color-coded cliques in the residue–interaction network. Active-site residues (black outlines) and mutation sites (red outlines) frequently appear within the same cliques. **(B)** Shows community detection of the DHFR interaction network. Nodes represent residues and edges represent correlated motions. Nine communities were identified. Residues that function as both catalytic and mutation hotspots are marked with both outlines and can be seen belonging to the same community in some cases. **(C)** Deep mutational scanning heatmap displays the predicted fitness effects for all single amino-acid substitutions across the protein fold. Blue shaded regions indicate tolerated or beneficial substitutions, whereas red shades indicate deleterious effects.

Structural Features Inferred from 3D Protein Structure

Our sequence-based analyses identified differences in residues that tend to change together. These patterns, as informative as they are, do not indicate how the protein copes with these substitutions within the three-dimensional structure. Specifically, pairwise evolutionary signals are unable to tell the difference between the change which was caused by local packing and that which was caused by long-range dynamic communication. To address this, we looked at the correlated motions and structuring of dynamic sub-networks based on the dynamic cross-correlation matrix (DCCM) (30). The WT contact map (Fig. 4C) exhibited a landscape with no large positive correlated regions but had a predominantly weakly correlated. These were mostly localized in the N-terminal and the C-terminal tail. Eight communities were detected and as we observed in our previous analyses, many active-site residues and mutation-prone residues belonged to the same communities and not isolated. This demonstrates that the important catalytic and mutation-sensitive residues are involved in a distributed communication scaffold. We also noted that all mutants modified the WT network differently. The C41R mutant weakened some of the contacts identified and resulted in more anticorrelated patches throughout the DHFR fold (Fig. S6). Because of this mutation, the network reorganized itself into ten communities where C41 clustered together with the active size residue C50 in community 5. The other active-site residues moved into different communities compared to what we observed in the WT. On the contrary, the N42I mutant was seen to strengthen some contacts especially in the N-terminal. The network expanded into eleven communities and the mutant remained within the 41–55 structure block with active-site residues 45–49 (Fig. S6). C50R and S99N mutations resulted in even more anticorrelations. C50R also existed in the same united structure block with N42 and C41. Although long-range couplings were weakened, new short-range contacts were obtained especially between the stretch of residue position 75 and 85. S99N clustered with the active sites 102 and 103 although they moved into a smaller community. A lot of correlations with the C-terminal were lost. The I155L mutant showed the appearance of strong positive correlations around residues 68–79 as well as prominent anticorrelation patches. It is grouped with residues sites 99, 102 and 103. A residue-level view of how these local changes reorganize dynamic communities across the trajectory is shown in Fig. S7. Mutations in all systems neither isolated residues nor broke down the network. Rather, they reshaped the communication pathways in a distinct manner. The WT enzyme has a unified, hub-based structure that has

established communication channels. Mutations either destroy or form new short or long-range stabilizing interactions. The result is a list of other communication solutions that enable the protein to trade-off between catalytic demands and conformational changes related to resistance.

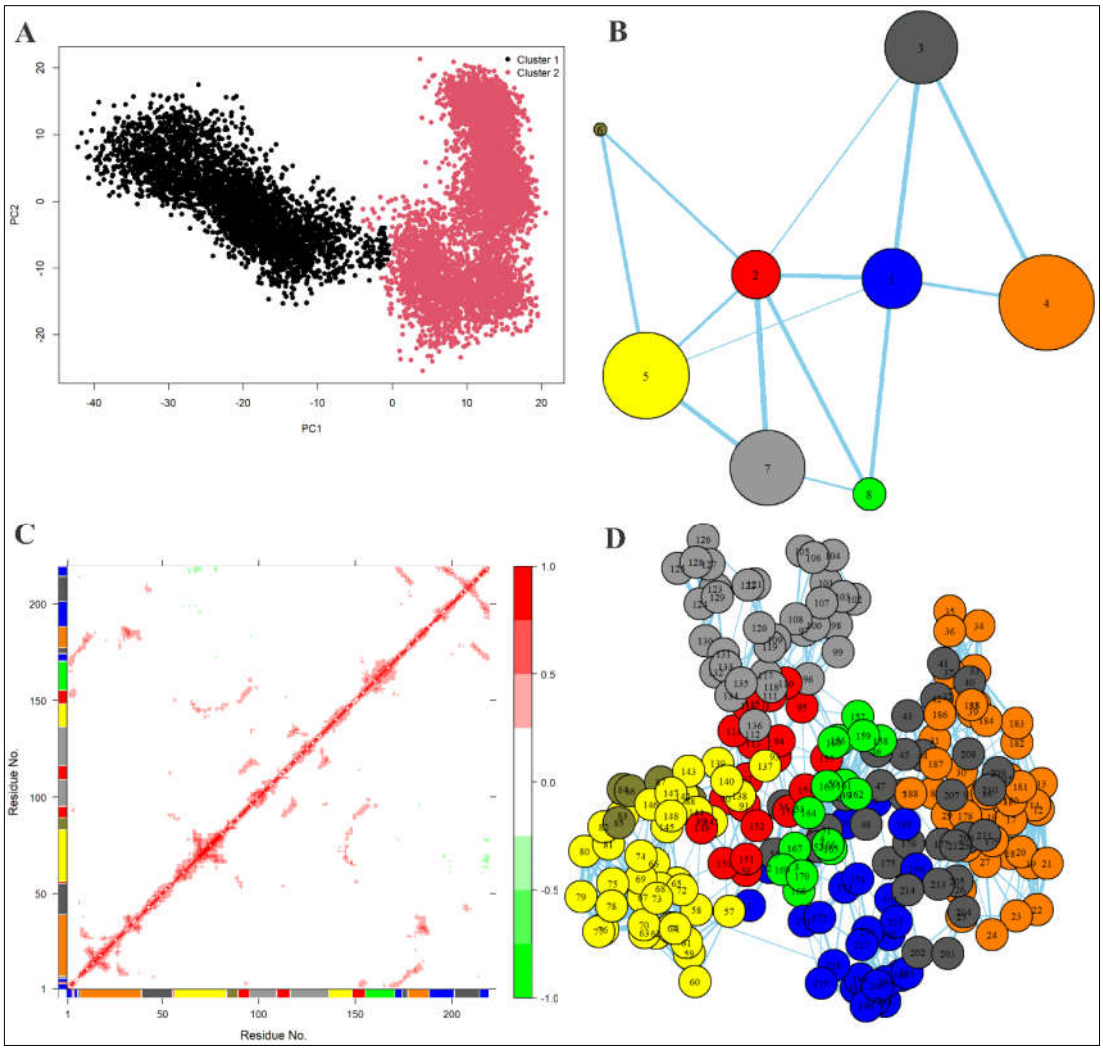

Fig. 4. Local flexibility and correlated motions in WT *Pf*DHFR. (A) Principal component analysis shows that the WT enzyme samples two dominant conformational clusters indicating transitions between long-lived substates. **(B)** Community network representation of the WT dynamic correlation map. Each node represents a community of residues that move together, with node size proportional to community size and edge thickness indicating the strength of inter-community coupling. **(C)** The contact map of WT shows positive correlations (red) and negative correlations (green) which reveal

patterns of cooperative and opposing motions across the fold. The color bars along the axes indicate community assignments for each residue. **(D)** All-residue community network in which each node corresponds to a single residue and is colored according to its community membership.

Molecular Dynamics Reveal a Stable Wild Type Fold and Mutations Increase Local Flexibility

Molecular dynamics (MD) simulations of the apo form of WT *Pf*DHFR were performed to complement the sequence-based analyses and to understand how the protein behaves in 3D over time. Although entropy and co-variation describe mutational tolerance and coupling, molecular dynamics help show how the protein fold is able to accommodate these sequence-level changes by intrinsic motions. Across the one-microsecond trajectory, the WT enzyme remained globally stable. After initial equilibration (about 100–350ns), the structure settled into a stable conformation by roughly 550ns and maintained this state for the rest of the run. The RMSD plateaued around 4Å with only a single transient peak near 5Å. This indicated no unfolding or large-scale destabilization (Fig. 5A). The radius of gyration (Fig. S9A) dropped slightly from approximately 1.98 to 1.85nm suggesting a slight compaction and the number of intramolecular hydrogen bonds remained steady at roughly 145–150 throughout the simulation (Fig. 5C). Per-residue fluctuations showed a clear division between a rigid catalytic core and highly dynamic loop regions. Loop 1 (residues ~20–35), Loop 2 (~85–100) and the C-terminal tail (~180–200) (Fig. 5B) displayed significant motions in the mutants while the catalytic core remained tightly constrained. Solvent Accessible Surface Area (SASA) was consistent with the small compaction depicted in the radius of gyration (Fig. 5D). Similarly, in all the resistance mutations, the global fold of *Pf*DHFR remained stable. But the local disruptions in the structure, particularly in the flexible loops and the adjoining segments were introduced by these mutations. In the WT simulation, loop motions reached peaks of around 6Å, whereas all mutants showed stronger fluctuations, often exceeding 7–8Å. The N42I mutant produced the strongest increase in flexibility, especially across Loop 2, where it displayed the highest RMSF values of all variants. This means the mutation can potentially cause destabilization of adjacent segments but the protein balances it by redistributing the movement on the surrounding loops (10). I155L showed a similar pattern. The core itself was stable, yet surface-exposed loops were more mobile and

the radius of gyration (Fig. S10) was a little bit higher which suits its purpose of being a permissible mutation in multi-mutation resistance backgrounds (10). S99N exhibited a low degree of local flexibility in the vicinity of the binding pocket that is in line with previous results that have associated this mutation with pocket adjustments and a modified interaction with ligands (31). The other resistance mutations also enhanced local flexibility about their respective substitution sites without adversely affecting the structural integrity. Despite the specific patterns of fluctuations being varied in them, they all had increased loop mobility compared to the WT with no apparent destabilization. These findings demonstrate that the structure of *Pf*DHFR is robust, and mutations that confer resistance always reshape local dynamics. They do not break the fold but instead destabilize their immediate neighborhoods and redistribute flexibility across surface loops that surround or communicate with the catalytic region. This observation is consistent with our co-variation analyses which showed that mutation sites and active-site residues form compensatory hubs.

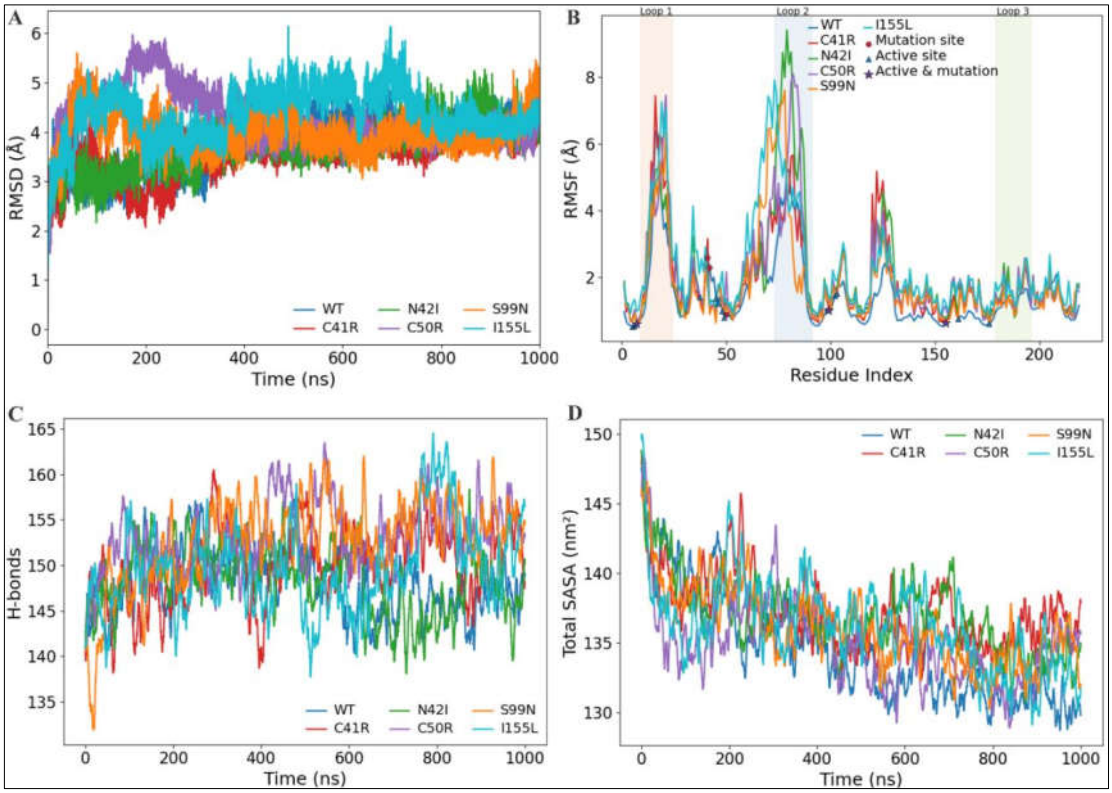

Fig. 5. MD comparison of WT and mutant *Pf*DHFR shows stable global folds with mutation-specific increases in loop flexibility. (A) All variants equilibrated within the

first ~350 ns and remain folded throughout the simulation. RMSD differences reflect conformational shifts rather than unfolding. **(B)** RMSF profiles showed that the same three loops are flexible in all systems, but the mutants consistently display higher fluctuations than WT. **(C)** Hydrogen-bond counts remain within a similar range for all variants indicating preserved overall packing. **(D)** Total SASA stabilizes between ~130–140 nm² across systems with only minor differences in compactness.

Free-energy landscapes reveal a rugged WT funnel and flattened, multi-basin mutant ensembles

The MD simulations demonstrated that *Pf*DHFR has a stable core with very flexible loop regions. However, these motions could only describe the movement of the protein, but not the energy cost and occupancy of each state. To capture this thermodynamic dimension, we constructed free-energy landscapes (FELs) by projecting all trajectories onto the same principal component plane so that the same collective motions apply to both the WT and all the variants. Free energy was calculated from the probability distribution using $\Delta G = -kBT \ln p$, where kBT is the thermal energy (10). The 3D surfaces (Fig. 6) and corresponding 2D contours (Fig. S11) map how often each conformation is visited. The valleys represent low-energy, frequently sampled states, while elevated regions correspond to high-energy conformations rarely accessed. The WT enzyme forms a rugged but focused funnel with most sampling concentrated around a single dominant low-energy basin. About 35% of frames fall within 2 kJ/mol of the minimum, 94% within 5 kJ/mol and nearly all within 10 kJ/mol, which indicates a well compacted native ensemble. All the five mutants span a similar overall energetic range (≈31–32 kJ/mol). However, they differ markedly in the shape of their low-energy regions. Each mutant shows a flatter minimum which indicates a redistribution of conformational probability into broader, shallower basins rather than an expansion of the total energy range. This effect was most pronounced in S99N, with approximately 71% of frames lying within 2 kJ/mol of the minimum, followed by C50R (~67%) and N42I (~61%). I155L showed intermediate broadening, while C41R remained most similar to the WT. Such differences show that resistance-associated mutations restructure the organization of accessible conformations without altering global stability. The wider and less deep basins reduce the energetic barriers between the nearby substates. This enables the mutants to be able to access a broader range of

low-energy conformations. This enhanced accessibility offers a thermodynamic pathway through which *Pf*DHFR can retain catalytic activity but be less vulnerable to antifolate inhibition.

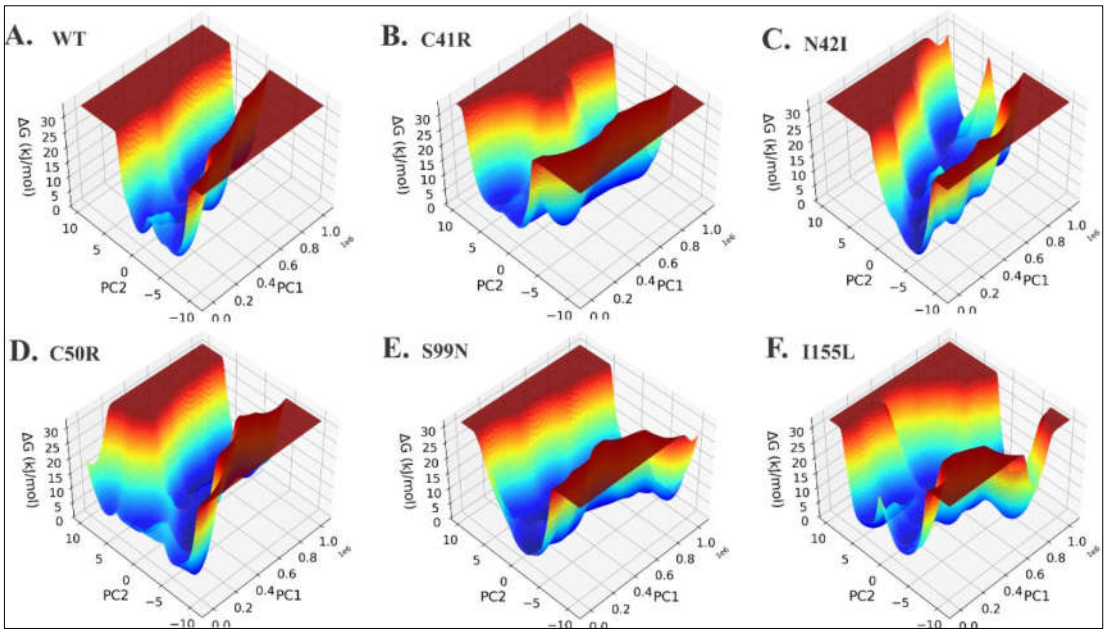

Fig. 6: Free-energy landscapes of *Pf*DHFR reveal that mutations flatten the WT funnel, leading to more flexible and adaptive conformations. **(A)** WT shows a steep, well-defined funnel with deep minima which is characteristic of a rigid folded ensemble. **(B)** C41R displayed a broader, saddle-like surface with multiple shallow dimples **(C)** N42I presented an uneven landscape with connected minima, indicating increased substate diversity. **(D)** C50R formed a flattened, near-degenerate basin allowing easy movement between similar conformations **(E)** S99N had the flattest surface with multiple equivalent minima linked to loop flexibility. **(F)** I155L showed a wide main basin with shallow side basins separated by low ridges. The shifts between nearby conformations are easier than in WT.

Adaptive-walk simulations reveal a latent compensatory ridge of *Pf*DHFR adaptability

To understand how *Pf*DHFR navigates the rugged fitness landscape, we developed the *Plasmodium* Fitness Pathway Analysis of Trajectories in Heterogeneous Landscapes (*Pf*PATH) model. This model simulates stochastic adaptive walks through a composite fitness landscape, which combines evolutionary sequence constraints, co-evolutionary coupling, protein stability,

and experimentally derived biochemical and kinetic measurements. These adaptive walks induce an initial mutation at the starting position and the algorithm then explores more fit neighbors by using the composite fitness function. Each bundle initially explored distinct mutational solutions depending on the first-step substitution. Every climb quickly narrowed down to a common high-fitness plateau, which indicates a segment of sequence space that is very accessible and resilient to the selection process. Importantly, many of the same compensatory substitutions recurred across the independent walks and across different starting residues. This repeated convergence highlights what we term as "latent compensatory ridge", a narrow set of positions that repeatedly serve as stabilizing, path-enabling intermediates regardless of where adaptation begins. Table 1 below shows this set of recurring residues. Positions such as V36, R29, K30, L31, C84, F187, R177, P38, L212, N209, T54 and neighboring residues repeatedly appeared in high-fitness trajectories. These sites occupy dynamic nodes that enhance stability, flexibility or co-operative behavior. Their repeated emergence implies that *Pf*DHFR relies on these locations as buffering points that allow costly resistance mutations to be accommodated without collapsing catalytic function.

Table 1. Recurrent stabilizing residues supporting adaptive pathways in *Pf*DHFR

Rank	Position	Evidence in Adaptive Walks
1	I215	Very frequent
2	F214	Frequent
3	I216	Frequent
4	P189	Frequent
5	R177	Regular
6	K123	Regular
7	T211	Occasional
8	V94	Occasional
9	L31	Occasional
10	R29	Occasional
11	V36	Rare but recurrent
12	P38	Rare but recurrent
13	G30	Rare but recurrent
14	C184	Rare but recurrent
15	F197	Rare but recurrent

Rank	Position	Evidence in Adaptive Walks
16	L212	Rare but recurrent
17	N209	Rare
18	T54	Rare

We also examined how these resistance-associated positions are selected and maintained during long-term evolution under fluctuating drug pressure. Simulations using a Wright–Fisher evolutionary model (32) were performed with a population size *N* of 10,000 per-site mutation rate ($\mu = 10^{-5}$) and a simple drug-cycling regime consisting of three generations with pyrimethamine "on" followed by one generation "off" repeated throughout the run. For each generation we tracked the fraction of the population carrying a non-wild-type amino acid at six key positions, that is, 7, 41, 42, 50, 99 and 155 which correspond reported resistance–associated sites (Fig. 7F). The trajectories revealed a clear temporal hierarchy of mutations at known resistance sites. The earliest and strongest sweep occurred at position 42, which climbed from near zero to >0.7 frequency by generation ~8,000 and eventually approached fixation. Position 41 rose later and more modestly, forming a broad transient peak centered around generations 10,000–11,000. Positions 50 and 155 showed only weak, delayed increases at low frequency near the end of the simulation. In contrast, positions 7 and 99 remained essentially neutral with mutation frequencies fluctuating around zero across generations.

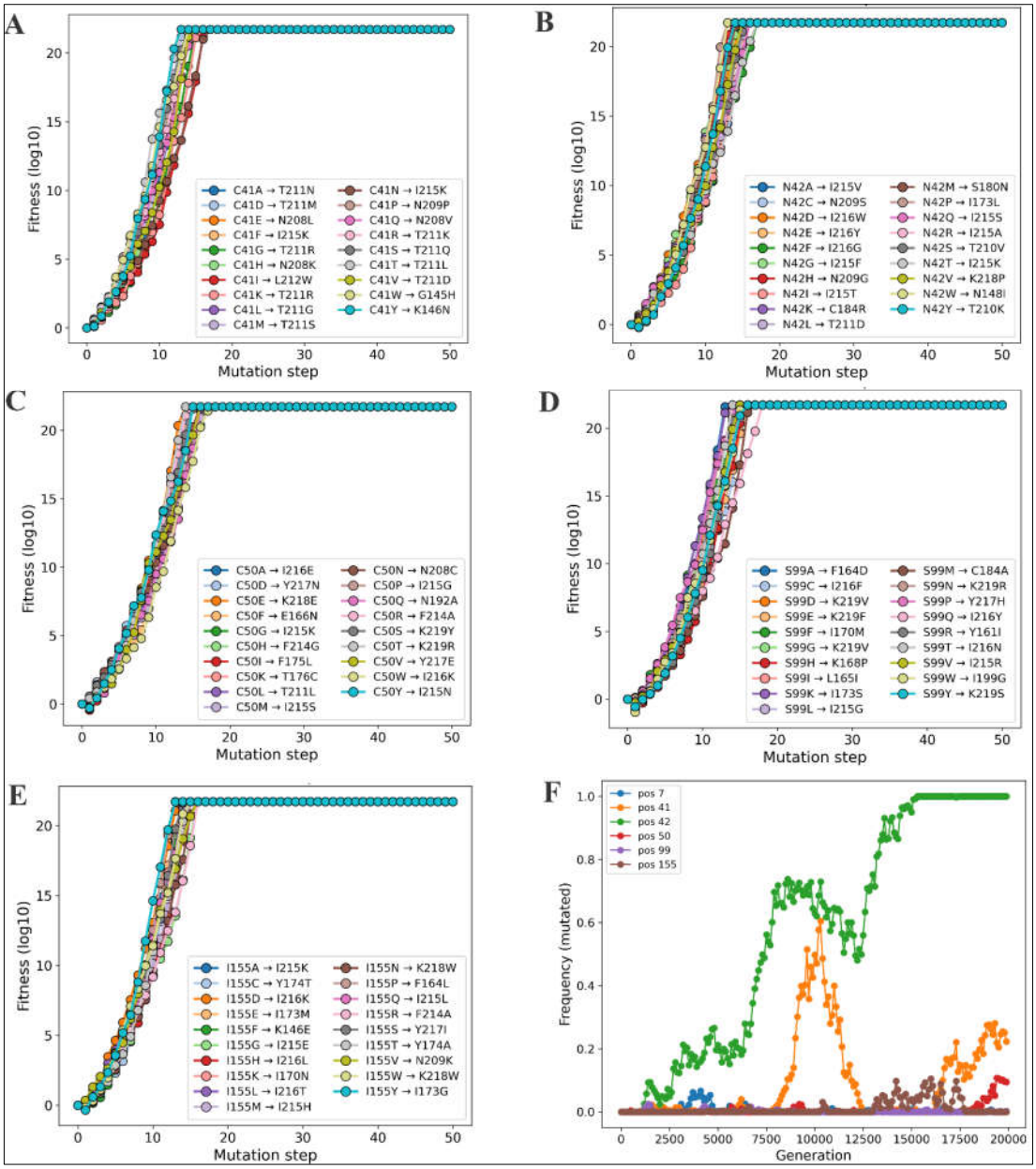

Fig. 7. Adaptive-walk trajectories and mutation-frequency dynamics for mutation positions. Panels **A–E** shows stochastic adaptive walks initiated at residues 41, 42, 50, 99 and 155. For each starting position, all 19 amino-acid substitutions rapidly climb toward a common high-fitness plateau but take distinct early paths that converge onto shared compensatory intermediates. (**F**) This shows the population-level mutation frequencies under Wright–Fisher evolution with drug cycling. Position 42 shows the earliest and

strongest sweep followed by position 41. Positions 50 and 155 increase only late and weakly. Position 99 remains largely absent and position 7 shows no meaningful selection.

Discussion

Dihydrofolate reductase has long served as a model for studying adaptive evolution under antifolate pressure (33). Previous studies treated *Pf*DHFR resistance as discrete outcomes with resistance inferred indirectly using surrogates (34, 35). In this work, we combined evolutionary features, network level constraints, molecular dynamics, thermodynamic stability and experimental kinetic data to construct a multidimensional fitness landscape that describes *Pf*DHFR evolution maintaining functionality. Collectively, our findings suggest that *Pf*DHFR is under strong functional constraints. That is, resistance mutations in the binding pocket will only be distinctly tolerated in the presence of compensatory mutations in other regions of the protein (36). The evolutionary and network analyses show that resistance in this protein arises from a collaborative mutational ensemble and not independent substitution.

Co-evolutionary patterns revealed that several mutation sites are statistically coupled forming coordinated clusters rather than acting in isolation. Many of these linked residues fall in or near the binding pocket. This implies that, when one region shifts, other regions somewhere within the protein adjust to maintain overall function. Such coordinated adjustments allow the enzyme to redistribute the structural and dynamic strain, which makes sense of the high-order mutation patterns commonly seen in resistant parasites (35). Structure network and correlation analyses also indicate that resistance mutations rearrange the short-range and long-range communication in *Pf*DHFR. Since these communication pathways link distal sites to the catalytic core, even single substitution can have an indirect effect on the behavior of enzymes by affecting the internal network. Some of these mutations cluster within flexible loop regions that control access to the binding pocket where changes in dynamics influence ligand binding and catalytic activity. Such distal, dynamics-driven effects are consistent with previous studies that have demonstrated long-range allosteric interactions in DHFR (37). Structural correlation maps aided in differentiating between the destabilizing and preservation or restoration of internal connectivity substitutions. The disruptive mutation, N42I, breaks important couplings and

discontinue communication channels. This can offer a mechanistic explanation as to why catalytic activity is less as previous studies reported (34, 35). Thus, if a single communication pathway is impaired, alternative pathways emerge to aid in signal transduction, which is an enzymatic property that contributes to resilience. Conversely, other mutations including C41R, C50R, S99N, and I155L do not only destroy the existing contacts, but also form new ones that distant parts of the protein. This type of rearrangement maintains communication over the long range and helps to sustain enzymatic activity. This portrays *Pf*DHFR as a modular system with a stable catalytic core that is supported by adaptable loop regions. Such an architecture could explain why resistance mutations emerge and at the same time preserve sufficient catalytic function. Mutational fitness profiling further emphasizes this modularity, revealing unequal tolerance to substitution across the protein fold. The most conserved residues like 99 and 155 show very little natural variation, yet resistance mutations are reported at these very locations. This apparent contradiction could be explained by epistasis where the effect of a mutation depends on the presence of other mutations (38).

The MD simulations of apo wild-type enzyme and resistant mutants indicate that the resistance mutations do not affect the global fold as was seen by similar values of the RMSD (10, 39). Rather, mutations distribute flexibility to surface-exposed loop regions at active sites and main contact interfaces. This enables the resistant variants to search over a wider number of local conformations and retain the overall structural integrity. Principal components analysis and free-energy analysis indicate that the wild-type enzyme is found mostly in one low-energy basin, and its resistant mutants can freely access multiple low-energy substates separated by relatively small barriers. These flatter energy landscapes provide active flexibility in a direction that does not collapse the conformational ensemble. Prior works have also indicated such flexibility in resistance mutations under selective pressures. Meanwhile, high-order mutations have been associated with reduced conformational flexibility and increased structural rigidity (10). The adaptive simulations via *Pf*PATH showed that antifolate resistance emerges through a restricted set of mutational routes shaped by stability, co-evolution and dynamic couplings. This composite landscape mirrors trends observed in field isolates and laboratory selection experiments (35, 40) in which certain mutations tend to emerge and persist early, followed by the subsequent accumulation of other mutations. Adaptive walks on this landscape reveal that

the DHFR protein rarely acquires mutations on their own. Instead, they usually move through a set of residues that act of stabilizing support points before major resistance mutations emerge. These compensatory steps are not confined to the canonical resistance sites but also many other intermediates across the protein fold. These residues form the "latent compensatory ridge" which is a group of structurally permissive positions that buffer destabilizing mutations and at the same time preserve function. Similar observations have been made in DHFR which showed that epistasis and structural constraints restrict viable adaptive routes yielding reproducible mutational trajectories across populations (41, 42). All in all, *Pf*DHFR adaptation relies not only on active-site chemistry but also on the broader architectural framework that distributes mechanical strain and tunes conformational ensembles.

The simulations at population level also indicate a hierarchical order of replacements which agrees with patterns observed in natural *P. falciparum* populations. The allele-frequency patterns indicated a clear hierarchy of substitutions. Position 42 (51) rose first and most rapidly, followed by a delayed rise in position 41 (50). Positions 50 (59), 99 (108) and 155 (164) were almost flat indicating their great catalytic cost in the composite landscape. Through this ordering, the overall trend of natural *P. falciparum* populations is observed with the DHFR mutations generally accumulating in the sequence of N51I -C59R -S108N -I164L, N51I (IRNL) (10, 40). Although S108N is often described as the first resistance mutation in natural populations (10), this haplotype in our experiment stayed at low frequency across generations. Under the fitness landscape and drug-cycling regime used, stronger early gains were consistently realized at positions 41 and 42. Hence, it is obvious that the simulations favored substitutions that balance resistance with catalytic viability and penalize destabilizing changes unless supported by the compensatory ridge. This could explain why some resistance sites are highly accessible and others remain conditional or rare. It could also explain why natural parasites repeatedly follow similar evolutionary routes and why only a small subset of all possible DHFR mutational combinations is observed in field isolates.

Our study elucidates how DHFR in *Plasmodium falciparum* adapts under antifolate pressure by evolving along a narrow and constrained evolutionary path. This trajectory allows the enzyme to balance the fitness cost of resistance while retaining sufficient catalytic activity for survival. Nonetheless, this study has limitations. Our analyses focused on the apo form of *Pf*DHFR and

did not explicitly consider ligand- or drug-bound states. Despite these limitations, the strong agreement between evolutionary features, structural dynamics and experimentally established resistance patterns supports a mechanistic view of *Pf*DHFR adaptation to antifolate pressure. Crucially, the presence of a narrow compensatory ridge suggests that resistance evolution in this protein is restricted to a limited set of structurally permissive pathways. Experimentally targeting this latent ridge may offer a rational strategy to both validate these predictions and guide the design of antifolates that constrain the parasite's ability to access viable escape routes.

Materials and Methods

Sequence preprocessing

For this study, we used the crystal structure of the wild-type *P. falciparum* DHFR-TS protein (PDB ID: 3QGT) as the reference for our analyses. The protein structure consists of two domains: the DHFR domain (first 231 residues) and the TS domain (last 288 residues), separated by an 89-residue interjunction region (Fig. S2). Given the focus on DHFR, we truncated the DHFR domain, which stretches from amino acid position 10 to 228 in the full-length protein as annotated in UNIPROT (43). This truncation was performed using the PYMOL Molecular Graphics System Version 3.0.3 **(**44**).** For consistency, the truncated sequence was re-indexed from 1 to 219, a scheme used throughout this study.

Homology Modeling and Structure Validation

Upon examination, the DHFR protein contained ten missing residues in its sequence from position 86 to 95. To fill in these gaps, homology modeling was performed using Modeller (Version 10.5) (45, 46). This software generated several protein models and ranked them using the Discrete Optimized Protein Energy (DOPE) scoring function. The model with the lowest DOPE score was selected as the most accurate representation of the full DHFR structure (Fig. S2). To ensure the accuracy of the modeled structure, we calculated the Root Mean Square Deviation (RMSD) between the modeled residues and the crystal structure (Fig. S3). An RMSD value of 0.306Å indicated high structural similarity. After validation, we extracted the amino

acid sequence of length 219 (as shown below) from the modeled DHFR structure for further analyses. The modeled structure also served as the seed structure for MD simulations.

>A7UD81_PLAFA

DIYAICACCKVESKNEGKKNEVFNNYTFRGLGNKGVLPWKCNSLDMKYFCAV
TTYVNESKYEKLKYKRCKYLNKETVDNVNDMPNSKKLQNVVVMGRTSWESIP
KKFKPLSNRINVILSRTLKKEDFDEDVYIINKVEDLIVLLGKLNYYKCFIIGGSVV
YQEFLEKKLIKKIYFTRINSTYECDVFFPEINENEYQIISVSDVYTSNNTTLDFIIYK
K

Sequence Space Analysis

To explore the mutational landscape of DHFR, we performed a sequence space analysis using the multiple sequence alignment (MSA) of homologous DHFR proteins. First, we retrieved 5000 non-redundant protein sequences homologous to DHFR using the NCBI BlastP tool (47). The sequences were cleaned to ensure high quality, non-redundancy and biological relevance. We then aligned the cleaned sequences using the ClustalW algorithm (48) and the resulting MSA served as the basis for the subsequent analyses.

Occupancy Calculation

Occupancy calculations were performed to assess how frequently specific amino acids occupy each residue position in homologous DHFR sequences (49). Positions with high occupancy imply conserved residues which are likely essential for the stability of protein. We identified residues where mutations may have significant impacts on the protein's structure and function by comparing occupancy data with mutation sites.

Shannon Entropy Calculation

Shannon entropy (S (i)) (49, 50) was calculated to quantify the variability of amino acids at each sequence position. Positions with low entropy values represent conserved residues, critical for structural and functional integrity, while high entropy values indicate positions prone to mutations. Shannon entropy was calculated following this formulation,

$$S(i) = -\sum_{a=1}^{K} p_i\,(a) \log_b p_i(a) \qquad (1)$$

Where $\boldsymbol{i}$ represents the sequence position, a indexes the amino-acid alphabet of size K (with K=20 for standard protein sequences) and $p_i(a)$ is the probability of observing amino acid a at position i in the multiple sequence alignment. Lower values of $S(i)$ correspond to highly conserved residues, while higher values indicate increased sequence variability.

Mutual Information Calculation

To investigate co-evolving residues, we calculated mutual information (MI) (49, 51, 52) for each pair of positions in the MSA. Mutual information quantifies the degree of correlation between two sequence positions, revealing residues that might be functionally or structurally interdependent. These co-varying residues could indicate regions where mutational compensation occurs, maintaining overall protein function despite mutations in specific positions.

$$MI(i,j) = \sum_{a=1}^{K} \sum_{b=1}^{L} p_{ij}\,(a,b) \log_b \left(\frac{p_{ij}(a,b)}{p_i(a)\, p_j(b)} \right) \qquad (2)$$

In this formulation, a and b index the amino-acid alphabets at positions i and j, of sizes K and L, respectively (with K=L=20 for protein sequences). Here, $p_{ij}(a,b)$ denotes the joint probability of observing amino acids a and b at positions i and j, while $p_i(a)$ and $p_j(b)$ represent the corresponding marginal probabilities at each position. The MI *(i,j)* values range from 0 (uncorrelated residues) to MI_{max} (the most interdependent residue pairs). In this study, a cutoff MI value of 0.5 was chosen to focus on highly co-varying positions. The MI values were visualized as a heatmap and a co-evolutionary matrix, generated using python programming.

Intrinsic Disorder Prediction

Intrinsic disorder prediction analysis was conducted to explore the presence of disordered regions within the DHFR protein. Disordered regions, known for their inherent flexibility, can tolerate higher levels of mutation and often play significant roles in protein-protein interactions. For a comprehensive evaluation, we employed six widely used disorder predictors: PONDR® FIT (21), PONDR® VSL2 (22), PONDR® VL3 (22), PONDR® VLXT (23), IUPred Short (24), and IUPred Long (24). These predictors use distinct predictive approaches to identify disordered regions within the protein. PONDR® VSL2 is particularly optimized for proteins containing both structured and disordered regions (53). We combined the results of all the six predictors to

obtain an average disorder profile (MDP), which yielded a unanimous disorder score to each residue. The scores above 0.5 were considered as disordered and the ones between 0.15 and 0.5 as flexible but considered relatively ordered. This disorder profile was further analyzed to look at possible correlations between disordered regions, known mutation sites, and active site residues to help understand the role of protein flexibility in resistance mutations and the maintenance of protein function.

Sequence-based Network Analysis

Following the MI calculations, we performed a sequence-based network analysis to identify co-evolving residues. Nodes in the network represent individual residues, while edges denote mutual information-based connections between them. MI thresholds were set to define the strength of these interactions, with different ranges indicating very strong (MI $\geq$ 0.8), strong (0.6 < MI < 0.8), moderate (0.4 < MI < 0.6) and weak (0.0 < MI < 0.4) interactions. This helped us visualize a network of interactions for selected residues of interest, particularly mutation sites and active site residues. The resulting networks provided insight into how mutations at specific positions might influence stability and drug resistance through interdependence with nearby residues. To investigate densely connected sub-networks within the protein, we employed Maximal Clique Analysis (MCA) (25, 26) and Community Detection (27) methods. In MCA, we aimed to identify groups of residues that form maximal cliques, where every residue within a clique is directly connected to every other residue. This method focuses on identifying highly interconnected sets of residues that are structurally and functionally interdependent. Each residue in the protein is represented as a node and the edges between these nodes represent significant interactions, which are based on proximity in the protein structure or known co-evolutionary relationships. Community network analysis was also performed to divide the protein's residue interaction network into distinct subgroups or communities. Unlike MCA, which focuses on fully connected sub-networks, community detection identifies groups of residues that are more likely to interact within their community than with residues outside it.

Structure Network and Correlation Analysis

Dynamic correlation and network analysis of *Pf*DHFR were done with the Bio3D package (version 2.4-1.9000) of R (30, 54). Bio3D offers a platform to comparative structure analysis, normal mode analysis and analysis of MD-based correlation network. Minimization of the

protein structures of the wild-type and mutant PfDHFR variants was initially performed followed by alignment of the backbone atoms to achieve the same frames of reference. In our case, we used the dccm() function of Bio3D to calculate residue-residue cross-correlation matrices (C_{ij}) to an extent that quantifies concerted atomic motions of individual structures. These matrices were then offered as inputs to correlation network analysis through the cna() function which represents each residue as a node and then pairs with weighted edges between the nodes based on how they move with each other. We used a cut (of $|C_{ij}| \geq 0.3$) to remove weak correlations and concentrate on couplings that have a functional meaning. The resulting networks were partitioned into communities with the Girvan-Newman edge betweenness algorithm (55) which represents clusters of highly correlated residues that are associated with dynamically consistent modules. The visualizations of networks were created with the plot.cna() and pymol() functions and examined in PyMOL 3D representations of the community of residues and communication pathways. In-house Python/R scripts were used to do other visualization and customized correlation mapping to create dynamic cross-correlation maps and network overlays.

Molecular Dynamic Simulations and Trajectory Analysis

A 1 microsecond MD simulation of apo *Pf*DHFR was performed to examine the structural dynamics using GROMACS 2025.1 (56) for both wild type and mutant structures. The system was prepared with the Amber99SB-ILDN force field (57) that provides superior side-chain torsion parameters to proteins. The *Pf*DHFR structure was put in a cubic periodic box (side length of around 8.5 nm) and solvated with 25, 476 explicit SPC/E water molecules (58). Neutralizing counter-ions (8 Cl^-) were added to mimic a 0.15 M ionic strength environment and long-range electrostatic interactions were treated with the particle-mesh Ewald (PME) method (59, 60). The system was first energy minimized until convergence (< 1000 kJ/mol/nm maximum force) was achieved before production. The equilibration was then carried out in two stages, that is, 100ps NVT at 300 K and 100ps NPT at 1 bar. A 2fs integration timestep was used, with all covalent bonds involving hydrogen constrained via the LINCS algorithm (61) to stabilize the system. Energies and pressures were stored as well as trajectory coordinates every 10ps and every 1ps respectively, to be analyzed later. During the production run, we undertook the standard structural descriptors to describe the dynamic activity of the protein. The backbone

root-mean-square deviation was computed to determine the general conformational drift and equilibration. Per-residue root-mean-square fluctuations were determined to determine the flexible regions of the protein. The DSSP algorithm was also used to track secondary structure content and thus monitor any alterations in helices, 2- sheet or loops (62). The other descriptors that were considered were the radius of gyration, solvent-accessible surface area, intra-molecular hydrogen bonds and principal component analysis to obtain the major motions and the major conformational states. The first two major components projected the free energy landscape to show the relative stabilities of the various conformational substates. These analyses presented in totality a global view of the conformational dynamics of PfDHFR throughout the simulation.

Simulating Evolutionary Trajectories in Heterogeneous Landscapes

We developed Plasmodium Fitness Pathway Analysis of Trajectories in Heterogeneous Landscapes (*Pf*PATH) to investigate how *Plasmodium falciparum* DHFR adapts under antifolate pressure. *Pf*PATH is a population-genetic and adaptive-walk simulator for *Pf*DHFR designed to computationally analyze how the protein adapts under drug pressure, structural constraints and mutational noise in high-dimensional heterogeneous fitness landscapes. This computational model integrated different layers of sequence information, structure features and experimentally derived kinetic and drug-response constraints into a unified fitness landscape and propagates parasite genotypes through a Wright-Fisher evolutionary process (32). The conceptual model is a generalization of the biophysical trajectory models already used to study DHFR evolution in other systems (34), except that it captures aspects of evolution, structural constraints and experimentally measured drug-response behavior. All genotypes were represented as full DHFR amino-acid sequences with mutations defined relative to the wild-type (WT). For a genotype g, the set of substitutions indexes

$$\mathcal{M}(g) = \{(p, a_p) \mid a_p \neq a_p^{WT}\} \qquad (3)$$

the positions and amino acids mutated from WT. Each mutation was assigned a set of single-site scores reflecting functional impact, stability changes and structure-based energetic effects. Since these sources differed in scale, each score was converted to a dimensionless quantity by z-normalization and hyperbolic-tangent transformation,

$$\tilde{s} = \tanh\left(\frac{s-\mu}{2\sigma}\right), \qquad (4)$$

which scaled down the data to the range of −1 to 1 so that no individual dataset dominated the terrain. Stability-related quantities were inverted in sign such that stabilizing mutations made a positive contribution. The composite score of genotype g was calculated by summing the contributions of all mutated sites. The single-site effects represented tendencies in functionality as inferred from mutational scanning and the anticipated alterations in thermodynamic stability. Pairwise terms were then added to the model to approximate epistasis (41). These consisted of co-evolutionary couplings and structural–dynamics couplings. A small positive contribution was assigned when both members of a strongly interacting pair were mutated, whereas a small penalty was imposed when only one member of the pair was mutated. Multi-mutant ΔG predictions were also incorporated into the model. This allowed the landscape to capture non-additive mutation–mutation interactions. These relations rendered the terrain rugged and directional. To restrict exploration to biologically feasible mutational loads, *Pf*PATH applied a mild penalty to genotypes containing more than six substitutions. This cutoff was used to reflect the rarity of *Pf*DHFR haplotypes with more than six mutations in nature. Genotypes containing more than six mutations were assigned a linearly proportional mutation-load penalty per additional substitution. An explicit stability filter based on ΔΔG predictions was also applied. For each genotype, the most destabilizing predicted ΔΔG was identified as,

$$\Delta\Delta G_{\max}(g) = \max_{(p,a)\in\mathcal{M}(g)} \Delta\Delta G_{p,a} \qquad (5)$$

A smooth penalty was applied whenever this value exceeded a soft threshold, and any genotype whose maximum ΔΔG exceeded a hard cutoff was excluded. This stability gate prevented evolutionary paths from passing through highly unstable or catalytically incompetent intermediates. The final composite score was

$$S(g) = S_{\text{single}}(g) + S_{\text{pair}}(g) - \text{penalty}_{\text{count}}(g), \qquad (6)$$

and intrinsic fitness is mapped via a Boltzmann-like transformation,

$$W_{\text{intrinsic}}(g) = \exp[\beta\, S(g)]\, \phi_{\text{stab}}(g), \qquad (7)$$

with selection strength $\beta = 2$. This formulation ensured that moderate differences in composite score translated into meaningful but not extreme differences in reproductive success. Drug pressure was applied by adjusting intrinsic fitness based on standard DHFR resistance haplotypes. The mutations N51I, C59R, S108N, and I164L were used to classify genotypes as wild-type-like, double-, triple-, or quadruple-resistant. These resistant haplotypes were assigned a multiplicative fitness benefit in the presence of drug, proportional to experimentally measured increases in inhibitor dissociation constants. In the absence of drug, these haplotypes incurred a fitness cost corresponding to reduced catalytic efficiency, as previously reported for these mutants (40). During drug-on periods,

$$W_{\text{on}}(g) = W_{\text{intrinsic}}(g)\,(1 + a_{h(g)}), \qquad (8)$$

and in drug-off periods,

$$W_{\text{off}}(g) = W_{\text{intrinsic}}(g)\,(1 - c_{h(g)}), \qquad (9)$$

where $a_{h(g)}$ and $c_{h(g)}$ depend on the haplotype class. Simulations assumed either constant drug pressure or cyclical exposure, allowing exploration of how intermittent treatment reshaped evolutionary trajectories. Evolution in *Pf*PATH was modeled using a Wright–Fisher framework, which describes a population with discrete, non-overlapping generations in which each generation is formed by sampling individuals from the previous generation according to relative fitness (63). The population size N was fixed at 10,000 individuals per generation, and offspring were sampled using multinomial sampling with replacement. Mutations occurred independently at each amino-acid site with probability μ per generation. Although natural *P. falciparum* mutation rates are low, an elevated effective mutation rate ($\mu = 10^{-5}$ per site) was used to capture both de novo mutation and standing variation, enabling multi-step adaptive paths to be observed on computationally feasible timescales. After mutation, each unique genotype was assigned a fitness value using the composite fitness scorer, and the next generation of N individuals was sampled proportionally to these fitness values. Allele and haplotype frequencies at selected positions of interest were tracked across generations to capture time-resolved evolutionary dynamics in the presence and absence of drug.

References

1. Vitiello, A., Zovi, H., 2025. The status of malaria, challenges and opportunities to end malaria: an editorial. Eur J Clin Microbiol Infect Dis. 44(8), 1761–1762. https://doi.org/doi:%252010.1007/s10096-025-05160-2

2. Kreutzfeld, O., Tumwebaze, P.K., Byaruhanga, O., Katairo, T., Okitwi, M., Orena, S., Rasmussen, S.A., Legac, J., Conrad, M.D., Nsobya, S.L., Aydemir, O., Bailey, J.A., Duffey, M., Cooper, R.A., Rosenthal, P.J., 2022. Decreased Susceptibility to Dihydrofolate Reductase Inhibitors Associated with Genetic Polymorphisms in Ugandan Plasmodium falciparum Isolates. Journal of Infectious Diseases 225, 696–704. https://doi.org/10.1093/infdis/jiab435

3. Fernández-Villa, D., Aguilar, M.R., Rojo, L., 2019. Folic Acid Antagonists: Antimicrobial and Immunomodulating Mechanisms and Applications. Int J Mol Sci 20, 4996. https://doi.org/10.3390/ijms20204996

4. Wróbel, A., Arciszewska, K., Maliszewski, D., Drozdowska, D., 2020. Trimethoprim and other nonclassical antifolates an excellent template for searching modifications of dihydrofolate reductase enzyme inhibitors. J Antibiot 73, 5–27. https://doi.org/10.1038/s41429-019-0240-6

5. Bilsland, E., Van Vliet, L., Williams, K., Feltham, J., Carrasco, M.P., Fotoran, W.L., Cubillos, E.F.G., Wunderlich, G., Grøtli, M., Hollfelder, F., Jackson, V., King, R.D., Oliver, S.G., 2018. Plasmodium dihydrofolate reductase is a second enzyme target for the antimalarial action of triclosan. Sci Rep 8, 1038. https://doi.org/10.1038/s41598-018-19549-x

6. Hyde, J.E., 2005. Exploring the folate pathway in Plasmodium falciparum. Acta Trop 94, 191–206. https://doi.org/10.1016/j.actatropica.2005.04.002

7. Nzila, A., Ward, S.A., Marsh, K., Sims, P.F., Hyde, J.E., 2005. Comparative folate metabolism in humans and malaria parasites (part I): pointers for malaria treatment from cancer chemotherapy. Trends in parasitology 21, 292. https://doi.org/10.1016/j.pt.2005.04.002

8. Yuthavong, Y., Tarnchompoo, B., Vilaivan, T., Chitnumsub, P., Kamchonwongpaisan, S., Charman, S.A., McLennan, D.N., White, K.L., Vivas, L., Bongard, E., Thongphanchang, C., Taweechai, S., Vanichtanankul, J., Rattanajak, R., Arwon, U., Fantauzzi, P., Yuvaniyama, J., Charman, W.N., Matthews, D., 2012. Malarial dihydrofolate reductase as a paradigm for drug development against a resistance-compromised target. Proc Natl Acad Sci U S A 109, 16823–16828. https://doi.org/10.1073/pnas.1204556109

9. Sirawaraporn, W., Prapunwattana, P., Sirawaraporn, R., Yuthavong, Y., Santi, D.V., 1993. The dihydrofolate reductase domain of Plasmodium falciparum thymidylate

synthase-dihydrofolate reductase. Gene synthesis, expression, and anti-folate-resistant mutants. Journal of Biological Chemistry 268, 21637–21644. https://doi.org/10.1016/S0021-9258(20)80589-4

10. Amusengeri, A., Tata, R.B., Tastan Bishop, Ö., 2020. Understanding the Pyrimethamine Drug Resistance Mechanism via Combined Molecular Dynamics and Dynamic Residue Network Analysis. Molecules 25, 904. https://doi.org/10.3390/molecules25040904

11. Boukoumba, F.M., Lekana-Douki, J.B., Matsiegui, P.B., Moukodoum, D.N., Adegnika, A.A., Oyegue-Liabagui, S.L., 2021. High prevalence of genotypes associated with sulfadoxine/pyrimethamine resistance in the rural area of Fougamou, Gabon. Journal of Global Antimicrobial Resistance 25, 181–186. https://doi.org/10.1016/j.jgar.2021.03.003

12. Cetin, E., Atilgan, A.R., Atilgan, C., 2022. DHFR Mutants Modulate Their Synchronized Dynamics with the Substrate by Shifting Hydrogen Bond Occupancies. J Chem Inf Model 62, 6715–6726. https://doi.org/10.1021/acs.jcim.2c00507

13. Kayode, A.T., Ajogbasile, F.V., Akano, K., Uwanibe, J.N., Oluniyi, P.E., Eromon, P.J., Folarin, O.A., Sowunmi, A., Wirth, D.F., Happi, C.T., 2021. Polymorphisms in Plasmodium falciparum dihydropteroate synthetase and dihydrofolate reductase genes in Nigerian children with uncomplicated malaria using high-resolution melting technique. Sci Rep 11, 1–8. https://doi.org/10.1038/s41598-020-80017-6

14. Nana, R.R.D., Hawadak, J., Foko, L.P.K., Kumar, A., Chaudhry, S., Arya, A., Singh, V., 2022. Intermittent preventive treatment with Sulfadoxine pyrimethamine for malaria: a global overview and challenges affecting optimal drug uptake in pregnant women. Pathog Glob Health 117, 462–475. https://doi.org/10.1080/20477724.2022.2128563

15. Sanyal, D., Banerjee, S., Bej, A., Chowdhury, V.R., Uversky, V.N., Chowdhury, S., Chattopadhyay, K., 2022. An integrated understanding of the evolutionary and structural features of the SARS-CoV-2 spike receptor binding domain (RBD). International Journal of Biological Macromolecules 217, 492–505. https://doi.org/10.1016/j.ijbiomac.2022.07.022

16. Kim, J.K., Lee, C., Lim, S.W., Andring, J.T., Adhikari, A., McKenna, R., Kim, C.U., 2020. Structural insights into the effect of active-site mutation on the catalytic mechanism of carbonic anhydrase. IUCrJ 7, 985–994. https://doi.org/10.1107/S2052252520011008

17. Malabanan, M.M., Amyes, T.L., Richard, J.P., 2010. A Role for Flexible Loops in Enzyme Catalysis. Curr Opin Struct Biol 20, 702–710. https://doi.org/10.1016/j.sbi.2010.09.005

18. Corbella, M., Pinto, G.P., Kamerlin, S.C.L., 2023. Loop dynamics and the evolution of enzyme activity. Nat Rev Chem 7, 536–547. https://doi.org/10.1038/s41570-023-00495-w

19. Zinovjev, K., Guénon, P., Ramos-Guzmán, C.A., Ruiz-Pernía, J.J., Laage, D., Tuñón, I., 2024. Activation and friction in enzymatic loop opening and closing dynamics. Nature Communications 15, 2490. https://doi.org/10.1038/s41467-024-46723-9

20. Dayhoff II, G.W., Uversky, V.N., 2022. Rapid prediction and analysis of protein intrinsic disorder. Protein Science 31, e4496. https://doi.org/10.1002/pro.4496

21. Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K., Uversky, V.N., 2010. PONDR-FIT: A Meta-Predictor of Intrinsically Disordered Amino Acids. Biochimica et biophysica acta 1804, 996. https://doi.org/10.1016/j.bbapap.2010.01.011

22. Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., Obradovic, Z., 2006. Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics 7, 208. https://doi.org/10.1186/1471-2105-7-208

23. Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., Dunker, A.K., 2001. Sequence complexity of disordered protein. Proteins 42, 38–48. https://doi.org/10.1002/1097-0134(20010101)42:1%253C38::aid-prot50%253E3.0.co;2-3

24. Dosztányi, Z., Csizmók, V., Tompa, P., Simon, I., 2005. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol 347, 827–839. https://doi.org/10.1016/j.jmb.2005.01.071

25. Konc, J., Janezic, D., 2007. An improved branch and bound algorithm for the maximum clique problem. MATCH Commun. Math. Comput. Chem 58, 569-590.

26. Pardalos, P.M., Rodgers, G.P., 1992. A branch and bound algorithm for the maximum clique problem. Computers & Operations Research 19, 363–375. https://doi.org/10.1016/0305-0548(92)90067-F

27. Grant, W.P., Ahnert, S.E., 2019. Modular decomposition of protein structure using community detection. Journal of Complex Networks 7, 101–113. https://doi.org/10.1093/comnet/cny014

28. Munro, D., Singh, M., 2021. DeMaSk: a deep mutational scanning substitution matrix and its use for variant impact prediction. Bioinformatics 36, 5322–5329. https://doi.org/10.1093/bioinformatics/btaa1030

29. Campitelli, P., Modi, T., Kumar, S., Ozkan, S.B., 2020. The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. Annu. Rev. Biophys. 49, 267–288. https://doi.org/10.1146/annurev-biophys-052118-115517

30. Grant, B.J., Skjærven, L., Yao, X.-Q., 2021. The Bio3D packages for structural bioinformatics. Protein Science 30, 20–30. https://doi.org/10.1002/pro.3923

31. Tarnchompoo, B., Sirichaiwat, C., Phupong, W., Intaraudom, C., Sirawaraporn, W., Kamchonwongpaisan, S., Vanichtanankul, J., Thebtaranonth, Y., Yuthavong, Y., 2002. Development of 2,4-Diaminopyrimidines as Antimalarials Based on Inhibition of the S108N and C59R+S108N Mutants of Dihydrofolate Reductase from Pyrimethamine-Resistant Plasmodium falciparum. J. Med. Chem. 45, 1244–1252. https://doi.org/10.1021/jm010131q

32. Tataru P, Simonsen M, Bataillon T, Hobolth A. Statistical Inference in the Wright-Fisher Model Using Allele Frequency Data. Syst Biol. 2017 Jan 1;66(1):e30-e46. doi: 10.1093/sysbio/syw056. PMID: 28173553; PMCID: PMC5837693.

33. Schneider, A., 2018. Dihydrofolate reductase and membrane translocation: evolution of a classic experiment. The EMBO Reports 19, EMBR201745692. https://doi.org/10.15252/embr.201745692

34. Eccleston RC, Manko E, Campino S, Clark TG, Furnham N. A computational method for predicting the most likely evolutionary trajectories in the stepwise accumulation of resistance mutations. Elife. 2023 Dec 22;12:e84756. doi: 10.7554/eLife.84756. PMID: 38132182; PMCID: PMC10807863.

35. Lozovsky ER, Chookajorn T, Brown KM, Imwong M, Shaw PJ, Kamchonwongpaisan S, Neafsey DE, Weinreich DM, Hartl DL. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. Proc Natl Acad Sci U S A. 2009 Jul 21;106(29):12025-30. doi: 10.1073/pnas.0905922106. Epub 2009 Jul 8. PMID: 19587242; PMCID: PMC2715478.

36. Wagner, A., 2023. Evolvability-enhancing mutations in the fitness landscapes of an RNA and a protein. Nat Commun 14, 3624. https://doi.org/10.1038/s41467-023-39321-8

37. Kazan, I.C., Mills, J.H., Ozkan, S.B., 2023. Allosteric regulatory control in dihydrofolate reductase is revealed by dynamic asymmetry. Protein Sci 32, e4700. https://doi.org/10.1002/pro.4700

38. Matange, N., Bodkhe, S., Patel, M., Shah, P., 2018. Trade-offs with stability modulate innate and mutationally acquired drug resistance in bacterial dihydrofolate reductase enzymes. Biochemical Journal 475, 2107–2125. https://doi.org/10.1042/BCJ20180249

39. Wingert, B., Krieger, J., Li, H., Bahar, I., 2021. Adaptability and specificity: how do proteins balance opposing needs to achieve function? Current Opinion in Structural Biology, Theory and Simulation/Computational Methods Macromolecular Assemblies 67, 25–32. https://doi.org/10.1016/j.sbi.2020.08.009

40. Sirawaraporn, W., Sathitkul, T., Sirawaraporn, R., Yuthavong, Y., Santi, D.V., 1997. Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase. Proc Natl Acad Sci U S A 94, 1124–1129.

41. Starr, T.N., Thornton, J.W., 2016. Epistasis in protein evolution. Protein Science : A Publication of the Protein Society 25, 1204. https://doi.org/10.1002/pro.2897

42. D'Costa, S., Hinds, E.C., Freschlin, C.R., Song, H., Romero, P.A., 2023. Inferring protein fitness landscapes from laboratory evolution experiments. https://doi.org/10.1371/journal.pcbi.1010956

43. The UniProt Consortium, 2023. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research 51, D523–D531. https://doi.org/10.1093/nar/gkac1052

44. Yuan, S., Chan, H.C.S., Hu, Z., 2017. Using PyMOL as a platform for computational drug design. WIREs Computational Molecular Science 7, e1298. https://doi.org/10.1002/wcms.1298

45. Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U., Sali, A., 2006. Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics Chapter 5, Unit-5.6. https://doi.org/10.1002/0471250953.bi0506s15

46. Fiser, A., Do, R.K., Sali, A., 2000. Modeling of loops in protein structures. Protein Sci 9, 1753–1773. https://doi.org/10.1110/ps.9.9.1753

47. McGinnis, S., Madden, T.L., 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Research 32, W20. https://doi.org/10.1093/nar/gkh435

48. Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22, 4673–4680. https://doi.org/10.1093/nar/22.22.4673

49. Bakan A, Meireles LM, Bahar I ProDy: Protein Dynamics Inferred from Theory and Experiments. Bioinformatics 2011 27(11):1575-1577.

50. Strait, B.J., Dewey, T.G., 1996. The Shannon information entropy of protein sequences. Biophysical Journal 71, 148–155. https://doi.org/10.1016/S0006-3495(96)79210-X

51. Gloor, G.B., Martin, L.C., Wahl, L.M., Dunn, S.D., 2005. Mutual Information in Protein Multiple Sequence Alignments Reveals Two Classes of Coevolving Positions. Biochemistry 44, 7156–7165. https://doi.org/10.1021/bi050293e

52. Simonetti, F.L., Teppa, E., Chernomoretz, A., Nielsen, M., Marino Buslje, C., 2013. MISTIC: mutual information server to infer coevolution. Nucleic Acids Research 41, W8–W14. https://doi.org/10.1093/nar/gkt427

53. Johnson, D.E., Xue, B., Sickmeier, M.D., Meng, J., Cortese, M.S., Oldfield, C.J., Gall, T.L., Dunker, A.K., Uversky, V.N., 2012. High-Throughput Characterization of Intrinsic Disorder in Proteins from the Protein Structure Initiative. Journal of structural biology 180, 201. https://doi.org/10.1016/j.jsb.2012.05.013

54. Skjærven, L., Yao, X.-Q., Scarabelli, G., Grant, B.J., 2014. Integrating protein structural dynamics and evolutionary analysis with Bio3D. BMC Bioinformatics 15, 399. https://doi.org/10.1186/s12859-014-0399-6

55. Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences 99, 7821–7826. https://doi.org/10.1073/pnas.122653799

56. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

57. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., Shaw, D.E., 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78, 1950–1958. https://doi.org/10.1002/prot.22711

58. Fuentes-Azcatl, R., Mendoza, N., Alejandre, J., 2015. Improved SPC force field of water based on the dielectric constant: SPC/ε. Physica A: Statistical Mechanics and its Applications 420, 116–123. https://doi.org/10.1016/j.physa.2014.10.072

59. Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092. https://doi.org/10.1063/1.464397

60. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G., 1995. A smooth particle mesh Ewald method. J. Chem. Phys. 103, 8577–8593. https://doi.org/10.1063/1.470117

61. Hess, B., Becker, H., Berendsen, H.J., Fraaije, J.G., 1997. LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry 18. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

62. Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637. https://doi.org/10.1002/bip.360221211

63. Messer P.W. 2016. Neutral Models of Genetic Drift and Mutation. Encyclopedia of Evolutionary Biology 2016, Pages 119-123. https://doi.org/10.1016/B978-0-12-800049-6.00031-7

Acknowledgements

DM acknowledges One Health Trust, Bangalore, India. DM and SC acknowledge the Director, Birla Institute of Technology and Science-Pilani, Hyderabad, India.

Funding: Authors acknowledge the Director, Birla Institute of Technology and Science-Pilani, and Head, Department of Biological Sciences, BITS-Pilani Hyderabad, India. The work is supported by BITS NFSG (N4/24/1032). DP acknowledges Birla Institute of Technology and Science-Pilani Hyderabad for doctoral fellowship. DM acknowledges One Health Trust for doctoral fellowship.

Author Contributions: Conceptualization: SC and DS. Methodology: DM. Formal analysis: DM. Figure Preparation: DM and VNU. Writing—original draft: DM. Writing—review & editing: SC, DS, VNU, PU, and DP.

Competing Interests: The authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.