Molecular Biology and Evolution, 2025, **42**, 1–15
https://doi.org/10.1093/molbev/msaf215
Advance access publication 16 September 2025
Discoveries

MBE OXFORD

From Binding to Catalysis: Emergence of a Rudimentary Enzyme Conferring Intrinsic Antibiotic Resistance

Claudèle Lemay-St-Denis [1,2,3,4] Stella Cellier-Goetghebeur,[1,2,3] Maxime St-Aubin,[1,2,3] Keigo Ide,[5] Janine N. Copp,[6] Soichiro Tsuda,[5] Nir Ben-Tal,[7] Rachel Kolodny [*4] Joelle N. Pelletier*[1,2,3,8]

[1]PROTEO, The Québec Network for Research on Protein, Function, Engineering and Applications, Québec, Canada
[2]CGCC, Center in Green Chemistry and Catalysis, Montréal, Canada
[3]Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, Canada
[4]Department of Computer Science, University of Haifa, Haifa, Israel
[5]bitBiome, Tokyo, Japan
[6]Michael Smith Laboratories, University of British Columbia, Vancouver, Canada
[7]School of Neurobiology, Biochemistry and Biophysics, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel
[8]Chemistry Department, Université de Montréal, Montréal, Canada

***Corresponding authors**: E-mails: joelle.pelletier@umontreal.ca; trachel@cs.haifa.ac.il.
Associate editor: Klara Hlouchova

Abstract

How does enzymatic activity emerge? To shed light on this fundamental question, we study type B dihydrofolate reductases (DfrB), which were discovered for their role in antibiotic resistance. These rudimentary enzymes are evolutionarily distinct from the ubiquitous, monomeric FolA dihydrofolate reductases targeted by the antibiotic trimethoprim. DfrB is unique: it homotetramerizes to form a highly symmetrical central tunnel that accommodates its substrates in close proximity and the right orientation, thus promoting the metabolically essential production of tetrahydrofolate. It is the only known enzyme built from the ancient Src Homology 3 fold, typically a binding module. Strikingly, by studying the evolution of this enzyme family, we observe that no active-site residues are conserved across catalytically active homologs. Integrating experimental and computational analyses, we identify an intricate relationship between homotetramerization and catalytic activity, where formation of a tunnel featuring positive electrostatic potential proves to be a powerful predictor of activity. We demonstrate that the DfrB enzymes have not evolved in response to the synthetic antibiotic to which they confer strong resistance, and propose that DfrB domains evolved the capacity for rudimentary catalysis from a binding capacity. That (rudimentary) catalysis can emerge from the homotetramerization of a binding domain, and that it has been recently recruited by pathogenic bacteria, manifests the opportunistic nature of evolution.

Keywords: enzyme catalysis, enzyme evolution, antibiotic resistance, folate metabolism, DfrB, SH3 fold

Introduction

Enzymes are central to life as we know it. Their evolved capacity to catalyze essential chemical reactions has been key to the adaptation and diversification of life over the past 4 billion years. However, fundamental gaps remain in our understanding of the mechanisms that led to the emergence of catalytic capacity. These gaps are partly attributable to the fact that in many systems, the acquisition of catalysis dates back billions of years, challenging the ability to trace the molecular evolution of its emergence. Nevertheless, a few studies have documented examples of evolutionary processes in which enzymes emerged from noncatalytic proteins, upcycling a binding property into a catalytic property. In these cases, the ligand becomes the substrate, and the structural elements mediating the binding function eventually foster the catalytic site (Kaur and Subramanian 2014; Ortmayer et al. 2016; Clifton et al. 2018; Kaltenbach et al. 2018). The proposed mechanism of emergence begins with a binding site that already contains key residues involved in the catalysis (Harms 2018). Substitutions that cause subtle changes in the binding site favor the transition state, yielding an enzyme with weak catalytic activity. Subsequent substitutions lead to a more productive orientation of the substrates and the tuning of productive and unproductive protein motions to optimize catalysis.

Here we propose a new model for the evolution and emergence of catalysis, in an enzyme that is likely to have evolved from a noncatalytic domain through multimerization: type B dihydrofolate reductase (DfrB). DfrB enzymes were first discovered in the 1970s, 10 years after the clinical introduction of the fully synthetic antimicrobial trimethoprim (Noall et al. 1962; Fleming et al. 1972) which selectively inhibits the ubiquitous monomeric dihydrofolate (DHF) reductase FolA. DfrB enzymes are structurally unrelated to FolA (supplementary figure S1, Supplementary Material online). Indeed, DfrB enzymes are unaffected by trimethoprim and were shown to provide strong resistance via an alternative enzymatic pathway for the metabolically essential reduction of DHF (Howell 2005).

DfrB enzymes possess unusual biophysical and enzymological properties that have inspired numerous characterization

Received: January 27, 2025. **Revised:** June 3, 2025. **Accepted:** July 21, 2025

studies, primarily of DfrB1 (Bradrick et al. 1996; Li et al. 2001; Chopra et al. 2006). The DfrB enzyme is a Src Homology 3 (SH3) fold, a fold that accounts for 1.5%, or ~9,000, of the entries in the (99% nonredundant) ECOD protein domain database (Cheng et al. 2014). This very ancient fold is a small ß-barrel formed by 5 antiparallel ß-strands connected by 4 loops, one of which contains a 3_{10} helix (Fig. 1a, left) (Alvarez-Carreño et al. 2021). The DfrB domain tetramerizes to form a central, symmetrical active-site tunnel through the interface of 4 identical copies of the SH3 B4 strands (Fig. 1a). Tetramerization occurs in 2 steps: 2 unfolded monomers homomerize to form a dimer via ß-sheet extension, followed by the assembly of 2 dimers into a tetramer via loop interactions (Bodenreider et al. 2002). The tunnel formed upon tetramerization accommodates 2 ligands: the DHF substrate and the reducing cofactor NADPH. Placing DHF and NADPH in proximity to each other in the correct orientation enables hydride transfer from the nicotinamide group of NADPH to reduce the pterin group of DHF (supplementary figure S2, Supplementary Material online). This produces tetrahydrofolate, an essential cofactor for cell proliferation. Unlike microbial FolA DHF reductases, DfrB activity is not inhibited by trimethoprim.

Among the many protein domains that share the SH3 fold, the DfrB family (which includes 2 nonredundant ECOD database domains) is unique in its capacity to catalyze chemical reactions. Indeed, in most cases, the SH3 fold serves as a binding module. In eukaryotes, the SH3 fold is found in multidomain signaling proteins, where it recognizes proline-rich motifs within the same and other proteins (Dionne et al. 2022). Its role in protein-protein interaction is key for the assembly of complexes, where the L1 and L2 loops determine the binding specificity. In prokaryotes, SH3-fold domains are functionally more diverse, mediating binding to peptides, proteins, DNA, RNA, and metals (Kishan and Agrawal 2005). DfrB is also unique in its capacity to form specific homotetramers: the vast majority of SH3 domains do not form homocomplexes. The few that have been reported to form homomers do so by a diversity of surfaces and mechanisms, a further indication that such homo-oligomerization in DfrB is atypical for the SH3 fold.

The DfrB family is atypical also compared to other enzymes. First, its catalysis is proximity-based, assisted by the inherent disorder of the bound substrates (Duff et al. 2016; Lemay-St-Denis and Pelletier 2023). Moreover, it does not involve any essential chemistry of the enzyme; it mediates substrate binding mainly with its backbone rather than its side chains (Krahn et al. 2007). Notably, it has limited evolutionary potential for improvement due to its 4-fold symmetry (Lemay-St-Denis and Pelletier 2023). Thus, compared with the more efficient FolA enzymes, DfrB is considered a "rudimentary" DHF reductase (see Supplementary Material online for further discussion).

How did this rudimentary but unusual catalyst evolve and persist until this day, in parallel with the ubiquitous FolA? To address this question, we searched for distantly related homologs of DfrB. We studied these homologs computationally, using the state-of-the-art AlphaFold-multimer to predict the structure of their complexes and to identify homologs sharing the conformation of DfrB1's homotetramer. We further characterized these homologs and their catalytic activity experimentally, using in vivo and in vitro assays. Notably, not all DfrB homologs have catalytic activity; we collected clues as to their possible noncatalytic functions. Overall, our investigation, summarized in Fig. 2, reveals the properties of this unique antibiotic resistance-conferring enzyme and allows us to propose a new model for the evolution of catalytic activity, based on oligomerization of an SH3 fold.

Results

Curating a Set of DfrB Homologs

Searching with Hidden Markov Model (HMM) profiles in the genomic and metagenomic databases for sequences similar to previously characterized DfrB family members, we collected a representative set of 386 DfrB homologs. Figure 1b shows these homologs as a phylogenetic tree. Only 203 of these have a documented source (e.g. soil, water), and only 186 have an annotated host organism (supplementary figure S3, Supplementary Material online). Among the DfrB homologs with an assigned host organism, 63% (117) are found in Proteobacteria, of which 80% are in Alpha-proteobacteria. Actinobacteria (23%) and viruses (11%) are the other 2 most common taxonomic groups. Of the 203 homologs with annotated sample sources, 76% (152) were reported to be from soil, root, or plant-related samples, and 15% (30) from water-related samples. These data are consistent with recent reports that DfrB genes are abundant in rhizosphere and aquatic environments (Cellier-Goetghebeur et al. 2022; Kneis et al. 2022, 2023; Pham et al. 2023).

The Majority of DfrB Homologs are Unrelated to Antibiotic Resistance Contexts

As noted above, DfrB enzymes offer an alternative, evolutionarily advantageous, way to catalyze DHF reduction for bacteria exposed to trimethoprim, which selectively targets FolA. Genetic mobility mechanisms may transfer antibiotic resistance genes (ARGs) that provide such an evolutionary advantage to bacteria exposed to antibiotics. However, we argue that most DfrB homologs in our set are endogenous to their genomes, namely, their presence is not due to recent (post-trimethoprim) genomic mobilization. In particular, the vast majority of DfrB homologs (94.2%) do not lie near mobile genetic elements (MGEs). 139 of the 386 DfrB homologs have an analyzable genomic context (*Methods*); among these, only 1 (0.7%) has an integrase gene nearby and only 5 (3.6%) are near a transposase gene. Three homologs (2.2%) are near an *attC* site—an integron element that leads to gene excision and allows gene mobility when recognized by an integrase.

Moreover, a gene endogenous to its host organism will have a Guanine–Cytosine content (%GC) similar to that of the genome in which it is embedded. In our dataset, the %GC of the DfrB homologs is generally similar to that of their genomes; the ratio $\%GC_{gene}/\%GC_{genome}$ varies from 0.83 to 1.14 (Fig. 1c), supporting that these homologs are endogenous to their host organisms, or that their integration is sufficiently old for their GC content to have adapted. Note that, in contrast, the above conditions do not hold for *dfrB1* to *dfrB9*, previously identified in clinical settings. These were overwhelmingly found near MGE, and their %GC differs from that of their host organism (supplementary figure S4, Supplementary Material online), suggesting they were recently acquired and have not yet adapted to the GC content of their pathogenic host (Lemay-St-Denis et al. 2021). Notably, a recent report proposes that ß-Proteobacteria of the order *Burkholderiales* would have been the environmental hosts of the *dfrB* genes identified in

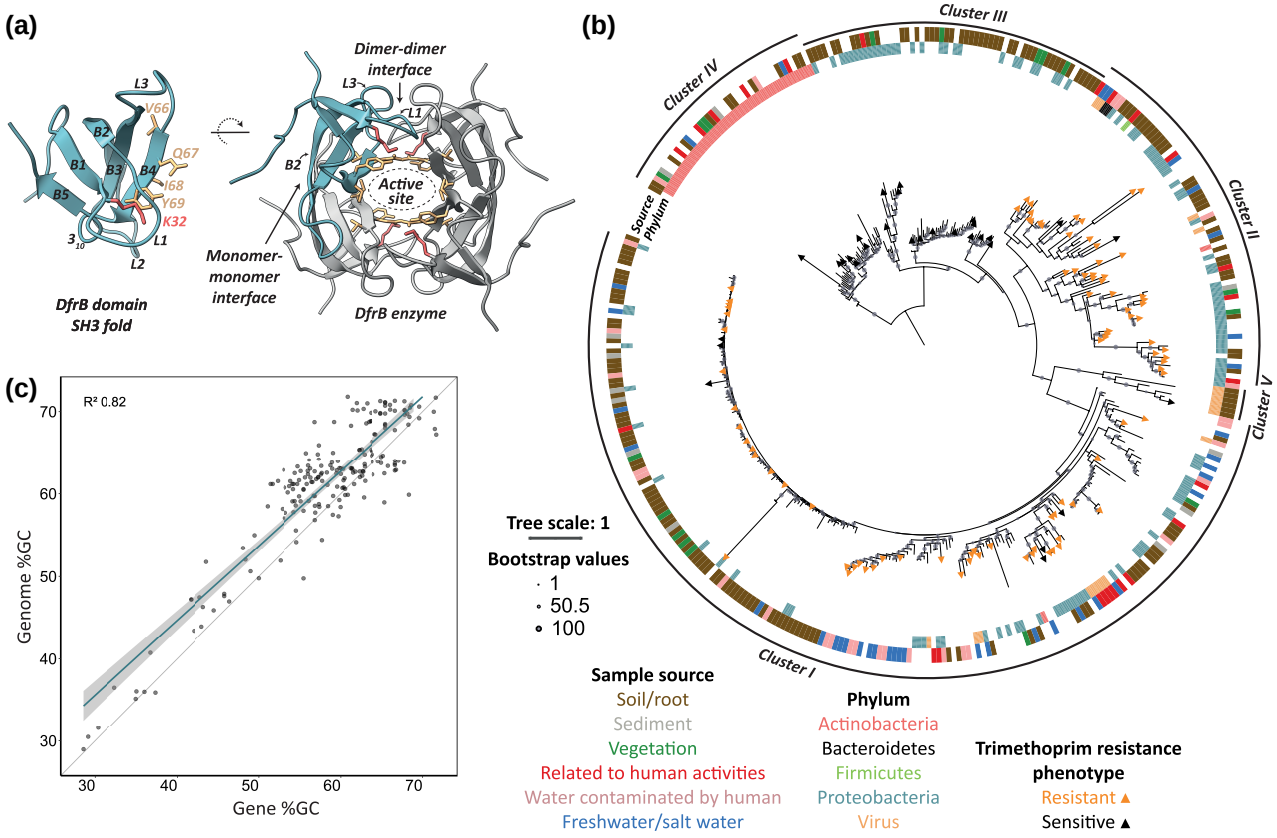

Fig. 1. The DfrB domain and its prevalence in environmental organisms. a) (left) The annotated topology of the DfrB1 domain (pdb 2rk1). The side chains of the active site motif are represented as sticks. (right) The homotetrameric DfrB1 enzyme structure, with one protomer colored in blue. Important elements for the homotetramer formation of this enzyme are indicated. b) Phylogenetic tree of the DfrB domain, rooted at the intersection of sequences from Actinobacteria and from Proteobacteria, along with the source of the sample and the taxonomic phylum of the organism, and the trimethoprim resistance phenotype of proteins when expressed in bacteria. Clusters are annotated according to the representation in Fig. 3. c) The GC content of *dfrB* homologs is plotted against the GC content of their cognate genome.

clinical samples or sharing the highest sequence identity with these (Kneis et al. 2024).

Multimerization Shapes the DfrB Sequence Space

Figure 2 shows the sequence similarity network (SSN) of the DfrB homologs. Each node represents a homolog, and edges connect nodes with similar sequences (MMseqs2 E-value lower than 10^{-6}). The homologs form 5 distinct clusters, with the clinically relevant DfrB members in cluster *I*, including DfrB1, whose structure is known (PDB ID 2rk1). Supplementary figure S5b and c, Supplementary Material online shows the homologs that are predicted to form DfrB1-like homodimers and homotetramers. The assembly of 2 protomers into homodimers, followed by the dimerization of 2 such homodimers into a homotetramer, is a hallmark of the thoroughly studied DfrB1 (Lemay-St-Denis and Pelletier 2023). Here, we used AlphaFold-multimer (Evans et al. 2021) to predict homodimer and homotetramer complex formation of the SH3 fold of each homolog (the region homologous to residues 20 to 78 in DfrB1). We consider the predicted complex to be a DfrB1-like tetramer when the monomer–monomer interface is formed by contacts between 2 B2 strands of 2 protomers, and the 2 dimer–dimer interfaces are formed by contacts between the L1 and L3 loops (supplementary figure S6, Supplementary Material online). The vast majority of homologs in clusters *I* and *II* are predicted to form a DfrB1-like dimer, for a total of 274 homologs; in cluster *III*, only some of the homologs are predicted to form such dimers (supplementary figure S5c, Supplementary Material online). As for the formation of DfrB1-like tetramers, we observe a clear demarcation between clusters *I* and *II*, where the DfrB1-like tetramer is predicted with high confidence, and the remaining clusters, where the DfrB1-like tetramer is not predicted (Fig. 3a; supplementary figure S7, Supplementary Material online). To our knowledge, no other SH3 domain utilizes the same contacts to form a homocomplex.

Catalytic Activity Correlates with Homotetramer Formation

Next, we investigated the relationship between catalysis and the formation of a DfrB1-like homotetrameric complex by experimentally characterizing 148 homologs across the entire sequence space sampled (representing 38% of our dataset). We found that the formation of a DfrB1-like tetramer is both a necessary and (almost universally) sufficient condition for catalytic activity. Among 123 homologs predicted by AlphaFold-multimer to form a DfrB1-like tetramer (all in clusters *I* and *II*), 114 provide trimethoprim resistance when expressed in *E. coli*, a convenient read-out for in vivo DHF reductase activity (Fig. 3b). The 9 remaining homologs (123−114 = 9) appear to be insoluble or not expressed, as evidenced by the absence of overexpression bands following resolution on a tris-glycine gel (supplementary figure S8, Supplementary Material online). None of the 25 homologs that were not predicted to form a DfrB1-like tetramer

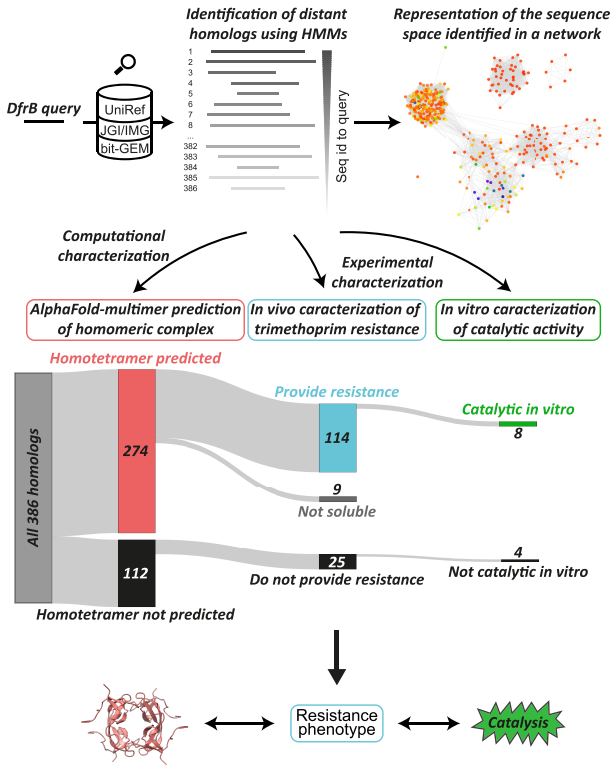

Fig. 2. Summary of the approach and key findings. Using the characterized members of the DfrB family, we searched genomic and metagenomic databases for homologs using HMM profiles, and studied the similarities amongst them. Then, for a set of 386 nonredundant DfrB homologs, we computationally predicted homomeric complexes using AlphaFold-multimer (salmon). We performed in vivo examinations with ~40% of the homologs (blue), investigating their capacity to confer antibiotic resistance. Finally, we characterized the DHF reduction capacity of a dozen of these homologs (green). Overall, our results show that the formation of a DfrB-like tetramer was a sufficient requirement for catalysis to emerge, and that this capacity is probably promiscuous to a principal (yet to be discovered) binding function.

(from clusters *III* to *V*) exhibited DHF reductase activity in vivo. Overall, AlphaFold-multimer predictions of having DfrB1-like homotetramers correlated in 139 of the 148 experimentally tested cases (94%) with the presence of DfrB catalytic function (Fig. 3).

Notably, the sequence segments forming both monomer–monomer and dimer–dimer interfaces of catalytically active homologs are not conserved in either residue identity or length (see Supplementary Material online for further discussion; supplementary figure S9, Supplementary Material online). For example, although A0A5E7Z8W7 has a 7-residue insertion on a loop at the dimer–dimer interface when compared to DfrB1, this homolog is predicted to form a DfrB1-like tetramer (supplementary figure S10b and c, Supplementary Material online). Indeed, it exhibits the same overall kinetic parameters as all active DfrB domains and assembles into a tetramer in solution (Table 1). More generally, not a single residue is strictly invariant throughout the SH3 fold in all catalytically active DfrB homologs (supplementary figure S9b, Supplementary Material online).

Characteristics of the "Active Site"

Most enzymes include specific amino acids, called catalytic residues, that directly facilitate catalysis, exchanging atoms and/or electrons with the substrates. Very specific functional groups, and thereby very particular amino acids, are required to this end, and they have to be placed in specific locations in 3D space. Thus, the catalytic residues are, in essence, unanimously conserved in evolution, and their immediate environment, i.e. the active site, is also highly conserved.

DfrB, being a "rudimentary" enzyme, is very different from most enzymes in that it merely promotes catalysis based on proximity and specific motions between the substrate and cofactor (Kamath et al. 2010). Thus, its "active site" residues do not require any essential chemistry and therefore there are no strict restrictions on their nature; the most conserved residue in DfrB is outside the "active-site", at the mouth of the tunnel, where a positively charged residue at position 32—generally a lysine—is required to anchor the negatively charged tails of both ligands. Nevertheless, the DfrB tunnel is the site of catalytic product formation. As such, it is commonly referred to as an "active site" (Narayana et al. 1995; Park et al. 1997: 67; Smiley et al. 2002), and we will continue to use this term for consistency.

The 114 homologs that experimentally catalyze DHF reduction in vivo do not share the same active-site residues; we identified 18 active-site motifs, coded in the B4 strand, that are conducive to catalysis (Fig. 4a). VQIY is the most common active-site motif and it is exclusive to homologs in cluster *I*, which includes the clinically observed DfrB homologs (DfrB1 to DfrB10). Homologs in cluster *II* display a wider range of active-site motifs. This finding is consistent with our previous work which demonstrated, through sequence alterations, the permissiveness of the active site to patterns of substitution (Schmitzer et al. 2004). Altogether, while there are no single specific active-site residues on the B4 strand that are essential for catalysis, in most cases the active site is characterized by the 4-residue motif: nonpolar/polar/nonpolar/aromatic (Fig. 4a).

We found that the tunnel formed by the catalytic DfrB homologs, though lined by different active-site motifs, features positive electrostatic potential (Fig. 4b; supplementary figure S11, Supplementary Material online). This is consistent with the tunnel's capacity to bind the negatively charged substrates: the co-crystal structure of DfrB1 shows the phosphate groups of NADPH and the carboxylate groups of DHF held within this space (Krahn et al. 2007). Contrary to the FolA DHF reductase, which is closely tailored to those ligands by side-chain interactions, the DfrB1 tunnel is wide and lined with backbone atoms, and it holds dozens of water molecules (Schnell et al. 2004; Howell 2005). Its modest catalytic activity appears to tolerate the wide evolutionary variation seen in clusters *I* and *II* (Table 1; Fig. 4). For example, the variant with the lowest catalytic efficiency (A0A6J5PTB3) has a highly positive electrostatic potential along the whole tunnel. Among DfrB homologs predicted to form dimers but not tetramers, not all B4 strand motifs create surfaces with positive electrostatic potential. This observed permissiveness in the identity of residues within the central and symmetric tunnel of DfrB homologs provides further support for the idea that catalysis relies on creating an environment conducive to catalysis rather than requiring specific chemistry within it. The most conserved residue among all catalytically active homologs is the lysine at position 32, and its only substitution observed is arginine, underscoring the importance of this positive charge in binding NADPH and DHF (supplementary figure S12, Supplementary Material online) (Hicks et al. 2004).

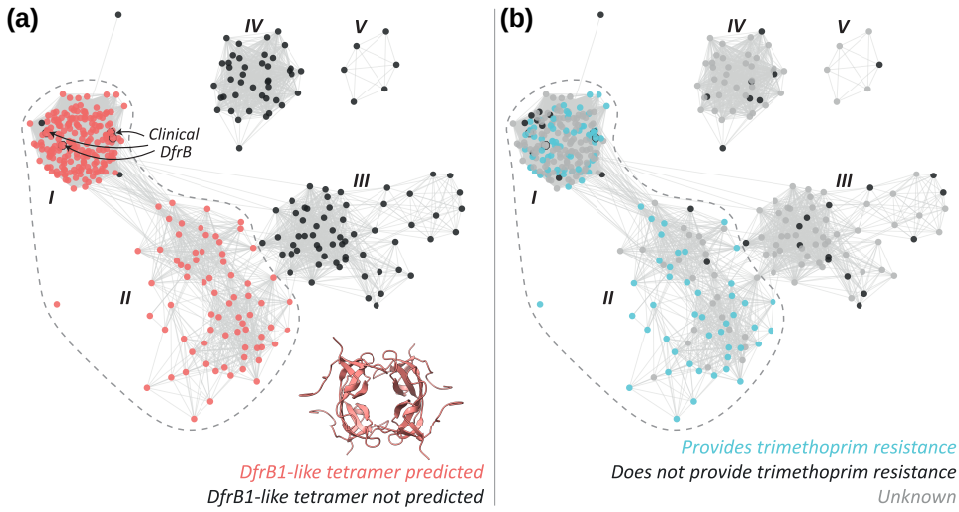

Fig. 3. Tetramerization is a defining property for the emergence of a catalytic function in the DfrB enzymes. a) A network visualization of the DfrB sequence space is presented, with each node representing a protein homolog. Edges connect nodes corresponding to proteins with similar sequences (E-value alignment score lower than 10^{-6}). The clinically identified DfrB members (DfrB1 to DfrB10) or their representative homologs are indicated with arrows, all in cluster *I*. Homologs are color-coded based on the AlphaFold-multimer predictions to form/not to form tetramers similar to DfrB1. b) Homologs are color-coded based on their ability to confer trimethoprim resistance in *E. coli* at a minimum concentration of 75 μg/mL trimethoprim.

In-depth Characterization of Representative Homologs

To further demonstrate that catalysis in DfrB depends on the enzyme's distinct homotetrameric structure rather than on the active-site chemistry, we conducted in vitro characterization of twelve homologs spanning the enzyme sequence space (supplementary figure S10, Supplementary Material online). Eight of these were predicted to form a DfrB-like tetramer (supplementary figure S13, Supplementary Material online): we experimentally validated their tetramer formation by size exclusion chromatography (Table 1). Despite the diversity of these homologs' active-site motifs, all but one exhibited virtually indistinguishable catalytic efficiencies for DHF reduction (Table 1). The only exception was A0A6J5PTB3, which binds 2 NADPH in its active site more favorably than 1 NADPH and 1 DHF, making it more of a binder than a catalyst (supplementary figure S14, Supplementary Material online).

The four in vitro-characterized homologs for which DfrB1-like tetramer formation was not predicted did not reduce DHF (supplementary figure S13, Supplementary Material online). Interestingly, size exclusion chromatography indicates that these catalytically inactive homologs form homomers, suggesting that the DfrB domain (of the SH3 fold) has a propensity for multimerization that may go beyond formation of a DfrB1-like complex (Table 1). It is worth noting that the 2 noncatalytic homologs for which tetrameric assembly was observed by SEC (E2SFM2 and A0A1R3X8F5), both in cluster *IV*, and sharing 74% sequence similarity, also have the highest pLDDT according to the homotetramer prediction (supplementary figure S13, Supplementary Material online). Their predicted conformations are similar to each other (RMSD 0.37 Å), but bear no relation to the DfrB-like tetramer. With the information currently available, we cannot propose a function for cluster *III*, *IV*, and *V* homologs.

Experimental Characterization of Additional Properties for Catalytic Homologs

We measured the thermostability of the set of catalytically active DfrB homologs and found that it varies with no discernible trend (Fig. 5b; supplementary figure S15, Supplementary Material online). Thermostability was found to be an unusual and distinguishing feature of the previously characterized clinically relevant members of the DfrB family (in cluster *I*), which can fully tolerate incubation at 95 °C (Lemay-St-Denis et al. 2023). We evaluated thermostability by comparing the DHF reductase activity in *E. coli* lysate following induced expression, before and after heat treatment at 50 and 75 °C. Thermotolerance was observed in 59% of the characterized DfrB homologs following incubation at 50 °C (70% of cluster *I* and 38% of cluster *II*), and in one-third following incubation at 75 °C (35% of cluster *I* and 38% of cluster *II*), demonstrating that thermotolerance is not a conserved property in the evolution of the DfrB domain. Therefore, we did not pursue further the investigation of this property in the context of the evolution of the DfrB domain. Note that this in vitro lysate assay was less sensitive than the in vivo trimethoprim resistance assay, with only 32 of the 114 trimethoprim-resistant homologs showing spectrophotometrically detectable DHF reductase activity.

DfrB1 first binds the NADPH cofactor, and then binds the DHF substrate to form the catalytic complex (Lemay-St-Denis and Pelletier 2023). We characterized the productive affinity of DfrB homologs for NADPH in lysate, revealing that 79% (19 of the 24 that were characterized in this assay) had a half maximum concentration constant ($K_{0.5}$) value of less than 30 μM (Fig. 5a), similar to that of DfrB1. That so many of the homologs share metabolically relevant affinity for NADPH implies that the interaction with this dinucleotide is key to their function. The affinity for DHF, on the other hand, varies widely between the homologs (Table 1), suggesting that it might be secondary in importance.

Investigating the Principal Function of the DfrB Domain

The highly specific FolA enzymes are roughly 100-fold more efficient at DHF reduction than DfrB enzymes (Howell 2005). Thus, we hypothesize that DfrB and its homologs evolved a different function—perhaps the promotion of

Table 1. Sequence-diverse DfrB homologs can catalyze DHF reductase with the same efficiency

Cluster	Protein ID	Local sequence identity with DfrB1 (%)	B4 strand motif	K_M^{DHF} (μM)	K_M^{NADPH} (μM)	k_{cat}^{DHF} (s^{-1})	k_{cat}^{DHF}/K_M^{DHF} ($s^{-1}\mu M^{-1}$)	Trimethoprim resistant	Predicted to form a DfrB1-like tetramer	Multimeric state determined by SEC
I	DfrB1	—	VQIY	8.2 ± 0.11[a]	1.6 ± 0.02[a]	0.83 ± 0.01[a]	0.10[a]	Yes	Yes	Tetramer
I	A0A1I4V6W3	47	VQIY	21 ± 7[b]	12 ± 1[b]	1.1 ± 0.2[b]	0.05[b]	Yes[b]	Yes	Tetramer and higher order states[b]
I	UPI001402A2DF	38	VQIY	130 ± 20	11 ± 1	2.5 ± 0.3	0.02	Yes	Yes	Tetramer
I	A0A839U7A6	52	VQLY	44 ± 5	10 ± 5	1.9 ± 0.1	0.04	Yes	Yes	Tetramer
I	A0A0N7CEH1	55	VQVY	6 ± 2	5.7 ± 0.6	0.47 ± 0.02	0.07	Yes	Yes	Tetramer
Singleton	A0A5E7Z8W7	30	IQNF	7 ± 4	18 ± 2	1.3 ± 0.03	0.17	Yes	Yes	Tetramer
II	A0A6J5PTB3	24	LHIF	160 ± 10	ND	1.07 ± 0.02	0.007	Yes	Yes	Tetramer
II	DfrB-TS (BBD027md-00173_03685)	31	LHIY	28 ± 3	4.8 ± 0.4	2.3 ± 0.1	0.08	Yes	Yes	Tetramer and octamer
II	BBD029md-00326_13692	35	QMIA	36 ± 4	20 ± 10	4.0 ± 0.3	0.11	Yes	Yes	Tetramer
III	A0A1M7UVF4	29	LQFL			NS		No	No	Dimer
IV	E2SFM2	28	VTYF			NS		No	No	Tetramer
IV	A0A1R3X8F5	24	VGYF			NS		No	No	Tetramer
V	A0A345GTK3	22	TLTL			NS		No	No	Dimer

The clusters refer to the nomenclature in Fig. 3. The protein IDs refer to either the UniRef ID or the bit-GEM ID. The B4 strand motif corresponds to the active site residues in catalytic DfrB enzymes. Trimethoprim resistance was determined by MIC. Homotetrameric predictions were made by AlphaFold-multimer using the SH3 fold region of each DfrB homolog. SEC was performed at 3 mg/mL. For homologs of clusters *I* and *II*, concentration ranged between 70 and 296 μM; homologs of clusters *III* to *V* were characterized at concentrations between 344 and 365 μM, indicating that protein concentration was not a limiting factor in multimerization.
[a]From Toulouse et al. (2020).
[b]From Lemay-St-Denis et al. (2023).
NS: No significant activity detected using at least 16-fold the protein concentration (molar) that was used to characterize active homologs.
ND: Not determined. Substrate inhibition was observed at >12.5 μM NADPH, see supplementary figure S14, Supplementary Material online.

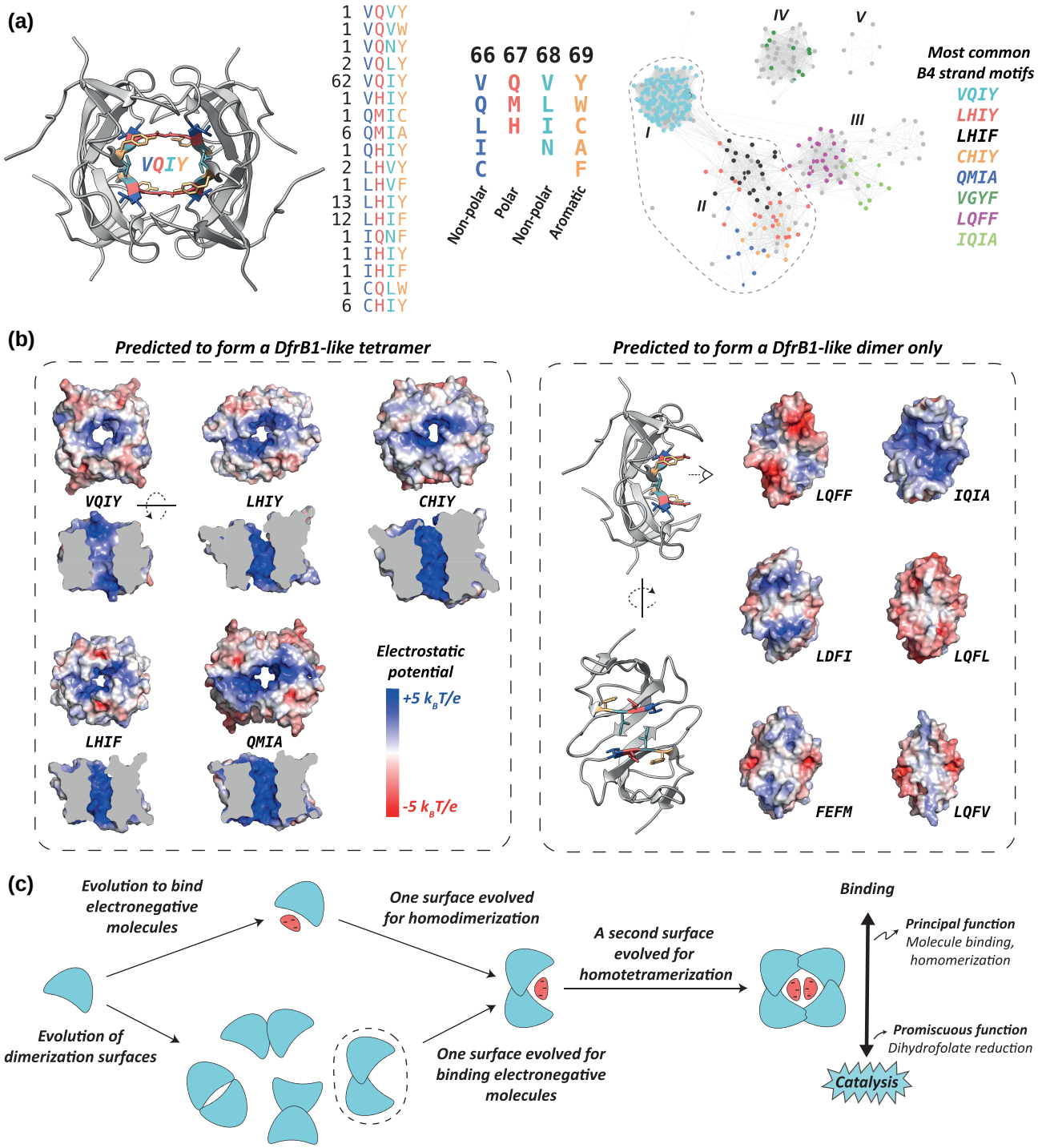

Fig. 4. DfrB's tunnel manifests sequence variation but a conserved positive electrostatic potential. a) (left) Marked on each of the DfrB protomers are the B4 strands that form the active-site tunnel. (center) All 18 active-site motifs that yield an active DHF reductase are presented; they correspond to residues 66 to 69 in DfrB1. The number to the left of each motif corresponds to the number of characterized catalytic homologs having this motif. (right) Homologs displaying one of the 8 most common B4 strand motifs are colored on the SSN. b) (left) The surface of predicted DfrB1-like tetramers of homologs representing 5 of the 18 active-site motifs is colored according to their electrostatic potential. The 13 remaining active-site motifs are similarly represented in supplementary figure S11, Supplementary Material online. (right) The surface of predicted DfrB1-like dimers of 6 representative homologs not predicted to form a DfrB1-like tetramer is colored according to their electrostatic potential. c) A possible evolutionary pathway for the emergence of the DfrB domain capable of catalyzing the reduction of DHF. The promiscuous catalytic activity of the complex would have led to the genomic recruitment of DfrB by pathogenic bacteria exposed to the antibiotic trimethoprim. This homotetrameric complex would also have later led to fusion with other domains, as described in "Domain fusions." The formation of such a ligand-binding homotetramer would result in the promiscuous function of catalysis.

protein–protein interactions, which is typical for the SH3 fold topology—while promiscuously allowing DHF reductase activity. We searched for hints as to the main native functions of the DfrB homologs.

Genomic Context

To gain insight into a putative native function of DfrB homologs, we examined their genomic contexts. In bacteria, genes with related functions are often located near each other and

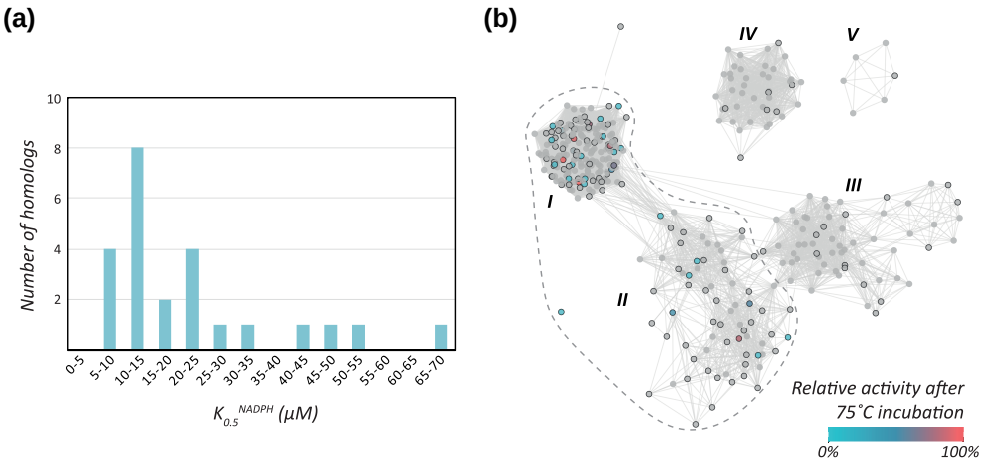

Fig. 5. Tight dinucleotide binding is conserved in the evolution of the DfrB domain, but thermostability is not. a) The distribution of half-maximal concentration constant for NADPH binding ($K_{0.5}$) for homologs having a detected DHF reductase activity in lysate. The $K_{0.5}^{NADPH}$ value for DfrB1 is 10.8 μM. b) Homologs with detectable DHF reductase activity in lysate are color-coded according to their thermostability. Following induction of DfrB homolog expression, *E. coli* lysates were subjected to a 10-min incubation at 75 °C to determine thermotolerance. All homologs experimentally tested for DHF reductase activity are indicated with a black circumference.

organized in operons (Dandekar 1998). The homologs predicted to form a DfrB1-like tetramer show a significantly stronger association with genes involved in replication, recombination, and DNA repair compared to homologs for which DfrB1-like tetrameric assembly is not predicted (supplementary figure S16, Supplementary Material online). Among the 139 genomic contexts analyzed, 17 cases neighbor the *dnaN* gene (median distance from the *dfrB* homolog of 2.1 kb), known for its homodimeric ring-like structure and DNA clamping action (Burnouf et al. 2004). Interestingly, *dnaN* was exclusively detected in the proximity of *dfrB* homologs predicted to form a DfrB1-like tetramer, such that 27% of those genomic contexts harbor *dnaN* within 10 kb of the *dfrB* homologs. This proximity is consistent with a functional relationship. Similarly, the *xerC* gene, a site-specific tyrosine recombinase that is implicated in the resolution of dimeric chromosomes (Blakely et al. 1993), was found in 23 occurrences (median distance of 2.0 kb). Of these, 20 were in the context of homologs predicted to form a DfrB1-like tetramer, and the remaining 3 were in the context of homologs that are not predicted to form DfrB1-like complexes. While some genomic contexts are conserved between homologs identified in bacteria of the same genus (supplementary figure S17, Supplementary Material online), there is no clear pattern across all contexts.

Many DNA viruses encode genes involved in the folate pathway, presumably to achieve self-sufficiency in nucleotide biosynthesis. Some viruses possess a *ThyX* gene, which codes for a thymidylate synthase (TS) that catalyzes the production of dTMP and tetrahydrofolate, an essential cofactor in the folate cycle. Others have a *ThyA* gene, which codes for a TS that produces dTMP and DHF, which is subsequently reduced by a DHF reductase to generate tetrahydrofolate (Levin et al. 2013). In total, we identified 20 DfrB homologs in viruses, with 19 belonging to the *Caudoviricetes* class of DNA viruses (supplementary table S1, Supplementary Material online). We analyzed the genes adjacent to the viral DfrB homologs, focusing on TS genes, which would suggest a functional relationship with DfrB (supplementary table S1, Supplementary Material online). Notably, all 4 DfrB homologs identified in viruses in cluster *II* are located near a *ThyA* gene (<20 kb), with 3 within 3 kb—one separated by 5 base pairs only. Among the 10 DfrB-containing viral genomes from cluster *I*, 3 harbor a *ThyA* gene, 2 of which are located within 20 kb of a DfrB homolog. Additionally, 6 noncatalytic DfrB homologs from cluster *V* are found in *Streptomyces* phages, which are also DNA viruses. All these genomes encode a *ThyX* gene that does not require a DHF reductase for thymine production, suggesting that the DfrB domain in these contexts has a role distinct from DHF reduction. This is consistent with the observation that DfrB homologs from cluster *V* are not predicted to form the DfrB1-like tetrameric structure and do not exhibit trimethoprim resistance. Overall, the genomic proximity of the viral *DfrB* homologs in cluster *II*, and to a lesser degree in cluster *I*, to *ThyA* genes, combined with the absence of a *FolA* DHF reductase in their respective DNA genomes, suggests that DfrB plays an essential metabolic role in dTMP production in these viruses.

Domain Fusions

We identified a relationship between the predicted formation of a DfrB1-like tetramer and the sequence length of the protein homologs (supplementary figure S18, Supplementary Material online). This relationship suggests that homotetramerization may have conferred an evolutionary advantage, favoring later evolutionary steps such as gene fusion with functional domains that benefit from the multimerization of the DfrB domain (supplementary figure S5d, Supplementary Material online). In this scenario, the native function of the DfrB domain would be to promote protein-protein interactions, as befits an SH3 fold.

In cases where DfrB domains are fused to other domains within a single protein chain, we searched using Foldseek (Van Kempen et al. 2024) for the functions of these fused domains to gather hints as to the function of the DfrB domain. Of the 45 proteins in our dataset that comprised more than 160 residues, and thus were likely to contain at least one more domain in addition to the 60-residue SH3 fold, only 7 were fused to a domain with a predicted function, while many others were predicted to be alpha-helix bundles of unknown function. Importantly, these 7 fusions were experimentally confirmed

to be functional as DHF reductases by their trimethoprim-resistance phenotype. In these cases, the domains identified in fusion with the DfrB domain were NTP pyrophosphohydrolase, ChaP oxygenase, *N*-acetyl transferase, and DnaA domains (see Supplementary Material online for further discussion; supplementary figures S19 and S20, Supplementary Material online).

Most interestingly, the DfrB domain was found in fusion with a bacterial TS domain of the ThyA family. We named this fusion protein DfrB-TS. As mentioned above, the TS and DHF reductase enzymes play sequential roles in the folate pathway, strongly supporting the proposition that the function of the DfrB domain in this specific fusion is to reduce the DHF produced by the TS domain (Fig. 6). Fusions of TS and DHF reductase domains are known, but to date have exclusively involved the fusion of TS with a FolA-like domain (DHFR-TS). First reported in *Crithidia fasciculata* (Ferone and Roland 1980) and later in several protozoans, plants, and viruses (Ivanetich and Santi 1990; Balestrazzi et al. 1995; Nomburg et al. 2024), such a fusion was never found in bacteria until here (supplementary figure S20d, Supplementary Material online). Notably, the TS domain functions as a symmetrical dimer (Carreras and Santi 1995). We experimentally verified that both domains within the DfrB-TS are functional, confirming its bifunctional nature (Fig. 6; supplementary figure S21, Supplementary Material online). The kinetic parameters of the TS domain are slightly lower than those of previously characterized TS domains, making its catalytic efficiency 6-fold lower than that reported for the TS domain in the DHFR-TS from *Toxoplasma gondii* (Trujillo et al. 1996).

Discussion

Using computational and experimental tools, we sought to shed light on the evolutionary emergence of the DfrB domain—a unique instance in which the ancient SH3 fold, which typically mediates protein-protein interactions, functions as an enzyme. We identified diverse homologs, sharing as little as 30% sequence identity with the first discovered DfrB1 enzyme, that can function as DHF reductases with the same catalytic efficiency. Based on these homologs, we concluded that in DfrB, DHF reductase activity is, in essence, dependent on the domain's tetramerization, which creates a central tunnel that can accommodate the DHF substrate and the NADPH reducing cofactor. Similar to DfrB1, the surface of the tunnel formed by these catalytic homologs features a positive electrostatic potential, complementary to the negative charges of DHF and NADPH.

Our findings suggest that the evolution of the DfrB domain may not have been driven by DHF reductase function. First, not all DfrB homologs form a tetramer, implying that they do not have enzymatic activity. Furthermore, those that form the DfrB1-like tetramer and thus produce tetrahydrofolate are 100-fold less catalytically efficient than the FolA enzymes that are ubiquitous in bacteria.

More specifically, we propose the following model for the evolution of DfrB's catalytically competent active site. The pattern of predicted DfrB1-like dimer and tetramer formation seen in the SSN (supplementary figure S5b and c, Supplementary Material online) is consistent with a monomeric SH3-fold domain evolving a dimerization interface and a positively charged ligand-binding surface. The ligand-binding dimer would then have evolved a second dimerization interface, forming the highly symmetrical homotetramer where the 2 equivalent positively charged binding surfaces line a central tunnel (Fig. 4c). The formation of a DfrB1-like tetramer, with a central tunnel that can accommodate one molecule of NADPH and one molecule of DHF, fortuitously formed a catalytic complex. This proposed model is consistent with the dimer, formed via the monomer–monomer interface ($K_D^{dimer \rightarrow monomer}$ 50 pM (Reece et al. 1991)), being more stable than the tetramer, formed via the dimer–dimer interface ($K_D^{tetramer \rightarrow dimer}$ 10 to 50 μM [Nichols et al. 1993; Dam and Blondel 2004]). Similar to what has been described in other systems (Levy et al. 2008), we suggest that the hierarchy in affinities between the DfrB units determines the order of assembly. Furthermore, the evolution of DfrB-like dimers in cluster *III* to DfrB-like tetramers, in cluster *II* homologs (Fig. 4b), may have promoted the acquisition of additional domains that could benefit from the DfrB complexes.

We propose 2 putative binding functions for which (at least some) DfrB domains may have been selected. The first proposed function is nucleotide binding. Evidence supporting this proposition includes the fact that the best characterized DfrB member—DfrB1—weakly binds ADP-ribose and ATP-ribose (Jackson et al. 2005). The conserved low micromolar affinity of the homologs for NADPH (Table 1, Fig. 4a) suggests that the function of these homologs is more likely related to the NADPH dinucleotide cofactor rather than to DHF. Notably, 2 NADPH molecules can simultaneously bind within DfrB1's symmetrical active-site tunnel, with negative cooperativity such that the loosely bound second NADPH could be readily donated (Jackson et al. 2005; Lemay-St-Denis and Pelletier 2023). As shown in Table 1 and supplementary figure S14, Supplementary Material online, the A0A6J5PTB3 homolog represents an exception, as binding of 2 NADPH molecules (forming a noncatalytic ternary complex) is favored over binding of one copy of both NADPH and DHF (required for the catalytic ternary complex). This is another indication that in DfrB, the DHF reduction function might be secondary to nucleotide binding. In addition, our genomic neighborhood analysis showed that homologs predicted to form a DfrB1-like tetramer are more frequently associated with genes whose function is related to DNA replication, recombination, and repair than their predicted monomeric/dimeric counterparts. Finally, the domains fused to the DfrB domain typically bind nucleotide-based molecules such as nucleotides, DNA, and acetyl-CoA (see Supplementary Material online for further discussion). Possibly, the DfrB domain homologs act as reservoirs for those ligands, binding them with moderate specificity to allow their release for use by their fused partner domain.

The second native function we propose for DfrB is promoting oligomerization. In particular, it is possible that the domains fused to DfrB1-like tetramer-forming homologs benefit from the capacity of the DfrB domain to oligomerize. Indeed, among these fused domains, those that have predicted functions typically assemble as symmetric dimers or as higher-order complexes (see Supplementary Material online for more details). Since both fused domains oligomerize, it is possible that the function of one of the 2 domains is the oligomerization itself, such that it promotes oligomerization of the other domain. Whether either of these hypotheses prove to be correct will require future experimental investigation.

Our data suggest that DfrB is unique in that its catalytic activity is promiscuous to its principal noncatalytic function. We refer to DfrB's catalytic activity as promiscuous, since it is

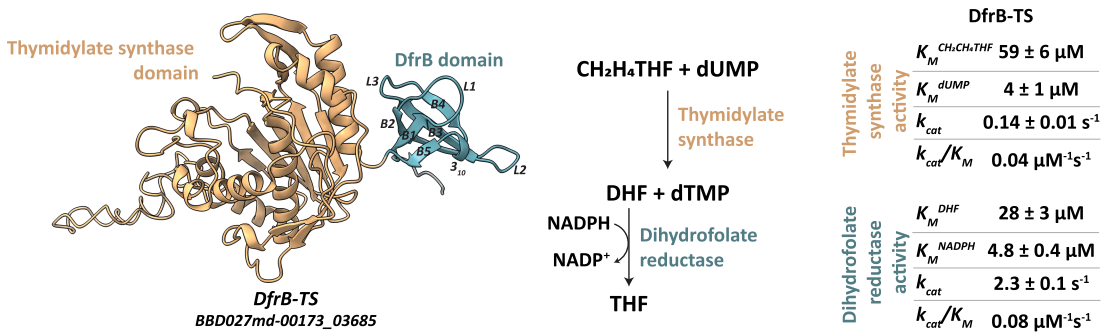

Fig. 6. A fusion between the DfrB and TS domains (DfrB-TS), catalyzing coupled reactions. (Left) A structural model of the DfrB domain fused to a TS domain. (Right) The kinetic parameters for the TS and DHF reductase activities of DfrB-TS.

irrelevant under endogenous physiological conditions, as FolA efficiently catalyzes DHF reduction. As reviewed by Tawfik and others (O'Brien and Herschlag 1999; Khersonsky et al. 2006; Copley 2015), functional promiscuity is typically observed in enzymes that catalyze multiple reactions, or, alternatively, in binding proteins capable of interacting with multiple partners (one at a time). That is, the promiscuity remains within the same "discipline"—catalysis or binding. We argue here that DfrB deviates from this general dogma in that it mixes disciplines. We propose that its principal function is binding, and that in addition it has promiscuous enzymatic activity. It is therefore through this binding property that the DfrB domain may have exerted its catalytic capacity.

Trimethoprim, which selectively targets the bacterial DHF reductase FolA, pressures bacteria to recover this essential metabolic catalytic activity. Evidently, some bacteria use the DfrB enzyme to compensate for the inhibition of FolA, thus providing antibiotic resistance. While it is reasonable to suggest that DfrB has mutated to provide this activity under trimethoprim selection pressure, we have shown here that DfrB domains from a variety of environments not exposed to trimethoprim have this enzymatic capability intrinsically. Thus, we conclude that the DfrB domain can carry out DHF reductase activity as is: the clinical introduction of trimethoprim merely resulted in the genomic recruitment of the existing DfrB domain to carry out DHF reductase activity in trimethoprim-exposed bacteria as a replacement for the chemically inhibited FolA. This perfectly illustrates that the environment provides a reservoir of functional antibiotic-resistance protein for novel antibiotics from which bacteria can sample (Canteón 2009).

Fusing domains that catalyze consecutive steps in a single pathway renders a more efficient system. Thus, for example, a DHFR-TS fusion (where a FolA-like domain is fused to a ThyA domain) has been identified in protozoan, plant organisms, and eukaryotic viruses (Lazar et al. 1993; Yuvaniyama et al. 2003; Nomburg et al. 2024). In this respect, the fusion protein DfrB-TS, analogous to the DHFR-TS, suggests that the DHF reductase activity of the DfrB domain has already been used by evolution at least once before in the context of the folate pathway. Note that this fusion also debunks a previous assumption that such fusions are only found in eukaryotic species (Stechmann and Cavalier-Smith 2002). It is also noteworthy that the DfrB-TS and DHFR-TS fusions constitute an elegant example of convergent metabolic evolution, where evolutionarily unrelated domains performing the same catalytic function (DfrB and FolA-like) are fused to the same TS domain.

The identification of DfrB in viruses provides another compelling case for the DfrB domain's role as a DHF reductase. Notably, all 4 homologs in cluster *II* identified within DNA viruses are genomically linked to a TS gene, with distances between the *DfrB* and *ThyA* genes ranging from 5 to 15.7 kb. These cluster *II* DfrB homologs are predicted to form a tetramer, suggesting catalytic activity; we experimentally tested and validated this activity for 3 of these 4 homologs. We hypothesize that the catalytic DfrB domains represent an integrative aspect of DNA metabolism in DNA viruses that harbor a *ThyA* gene but lack a *FolA* DHF reductase. For these, the DHF reductase activity of the DfrB homolog is essential for producing the tetrahydrofolate cofactor required for dTMP synthesis. Interestingly, noncatalytic DfrB homologs from cluster *V* are identified in viruses containing the *ThyX* gene, which enables thymine production without the need for DHF reduction. The presence of DfrB homologs in these viruses further supports the notion that the DfrB domain may have function(s) beyond DHF reductase activity.

Overall, our computational and experimental analyses show that homomerization can be a critical step in creating an environment conducive to (slow) catalysis. In particular, for the DfrB family, homotetramerization leads to the formation of a symmetrical tunnel with positive electrostatic potential, suitable for accommodating the NADPH cofactor and the DHF substrate, and facilitating their reaction. Thus, a minimalistic approach to a reductase reaction, essential for bacterial survival, can take place with relatively few restrictions. In particular, there are no strict restrictions on the identity of the active-site residues, which show remarkable sequence variation, since their side chains do not contribute directly to catalysis. These characteristics make the DfrB domain uniquely different from the catalytic regions of typical enzymes, including FolA, which are highly conserved. In a broader context, this study provides a glimpse of Nature's creativity in generating rudimentary catalysis in protein domains, unexpectedly providing strong antibiotic resistance to a recently introduced synthetic molecule.

Methods

Identification of Homologs and Synthesis of Representatives

We used the state-of-the-art homology detection software HMMER and HH-suite (Eddy 2011; Steinegger et al. 2019). A multiple sequence analysis consisting of the region corresponding to the SH3 fold in the trimethoprim-resistant DfrB1 to DfrB21 (positions 24 to 78) was used to search UniRef30 (2022_02) with HH-blits, using 5 iterations. The 367 identified sequences were filtered using HH-filter to retain

hits with 85% coverage and 20% sequence identity (-cov 85 -qid 20). The 182 filtered sequences were used as input for HMM-search, using an E-value of 10^{-3} to search against UniRef90 (2022_05). Using HH-suite and HMMER, a total of 195 proteins were identified. The protein IDs of the hits identified by HH-blits and hmmsearch were used to retrieve the corresponding sequences from UniRef. Entries recently removed from the database were not retrieved, resulting in a list of a total of 212 nonredundant homologs.

To explore a larger sequence space, we queried metagenomic data. A total of 2702 metagenomes from the JGI/IMG database (https://img.jgi.doe.gov/) were retrieved in May 2020 using a Pfam search for "DHFR_2" (Mistry et al. 2021). Filtering of their putative genes using the Pfam keyword "pfam06442" yielded 1524 complete genes starting with a methionine, of which 688 encoded nonredundant proteins. We further increased our exploration by searching the microbial single-cell genome database bit-GEM (Nishikawa et al. 2022) with the 182 hh-filtered sequences using HMMsearch with an E-value of 10^{-3}.

We combined the homologs obtained from UniRef, JGI, and bit-GEM and filtered the dataset with CD-HIT (85% sequence similarity and 85% length difference cutoff) (Fu et al. 2012), for a total of 386 sequences. A total of 148 representative sequences, covering the entire dataset and the diversity of active site motifs identified, were synthesized as N-terminally His-tagged in pET29b by Twist Biosciences, downstream from the T7 expression system. All homologs were codon-optimized by the vendor for expression in *E. coli*, and were expressed and characterized in full length, without sequence alteration.

Generating the Sequence Similarity Network

Briefly, pairwise MMseqs (Steinegger and Söding 2017) E-values were calculated between all possible pairs of the 386 representative complete sequences. Pairwise similarities were used to generate a network with Cytoscape (Shannon et al. 2003) in which a node represents a protein sequence, and an edge represents a pairwise MMseqs E-value. Edges with E-values scores of 10^{-6} and lower (closer homologs) are shown (Fig. 3).

Generating a DfrB-representative Phylogenetic Tree

The phylogenetic tree was generated from the MAFFT (Katoh and Standley 2013) alignment of the region of each sequence corresponding to the SH3 fold (positions 24 to 78 in DfrB1) using IQ-TREE (Nguyen et al. 2015) (ultrafast bootstrap analysis, 1000 alignments). The tree was visualized with iTOL (Letunic and Bork 2016) (Fig. 1b).

Genomic Annotation and Analysis

The sample source of isolation from which each protein was detected was retrieved. The genomic sequences of representative proteins were retrieved using their NCBI sequence ID along with their taxonomic assignment. Metagenomic contigs from JGI and bitBiome were too short for annotation purposes. GC% ratios were calculated for genes whose genomic sequence was at least 10 times longer (Fig. 1c).

From the genomic sequences, 10 kb upstream and downstream of the DfrB homolog gene, for a total of 20 kb, were annotated with Prokka (Seemann 2014) using the Galaxy server(The Galaxy Community et al. 2022). Genomic sequences with shorter genomic contexts surrounding the DfrB homolog genes were not used. The predicted genes were used as queries against the COG database, and a COG ID was annotated if the blastp E-value was less than 10^{-5} (Altschul 1997; Galperin et al. 2021).

COG functional categories were assigned using the COG reference list (https://www.ncbi.nlm.nih.gov/research/cog/cogcategory/J/) (Galperin et al. 2021). In addition to annotating the genomic sequences with Prokka, we sought to confirm the presence of MGEs and ARGs, as some *dfrB* genes have been previously associated with such elements. For this purpose, we used IntegronFinder (Néron et al. 2022) from the Galaxy server (The Galaxy Community et al. 2022) and the Resistance Gene Identifier from the Comprehensive Antibiotic Resistance Database (Alcock et al. 2019), respectively. Integrons were identified in contigs using the local detection (–local-max) and search for promoter and attI sites (–promoter-attI) options. Genomic annotations were compiled using R (version 4.1.3) and visualized using the ggplot2 (Wickham 2016) package (supplementary figures S17 and S19d, Supplementary Material online).

The analysis of the genomic context of DfrB homolog genes identified in viruses was carried out using their annotated genome in the "Nucleotide" section of NCBI.

Protein Structure Prediction

The DfrB1-like dimer and DfrB1-like tetramer for each protein were predicted with AlphaFold-multimer (Evans et al. 2021), using the region of each sequence corresponding to the SH3 fold. The Mgnify (Richardson et al. 2023), BFD (Jumper et al. 2021), UniRef90 (Suzek et al. 2015), Uniclust30 (Mirdita et al. 2017), UniProt (The UniProt Consortium et al. 2023), and PDB databases were used for these predictions. Out of 25 relaxed models, the model with the highest confidence score (0.8 ipTM [predicted interface TM score] + 0.2 pTM [predicted TM score]) was used for analysis. High confidence scores (Jumper et al. 2021; Wallner 2023) were assigned to the dimeric and tetrameric predictions in clusters *I* and *II* (supplementary figure S7, Supplementary Material online). The ability of the homologs to form a DfrB1-like dimer and a DfrB-like tetramer was determined by generating a contact map from each prediction, as described in supplementary figure S6, Supplementary Material online. The full-length sequences of the protein fusions were used to predict homodimeric complexes. The electrostatic potential of each predicted complex was calculated by the Adaptive Poisson–Boltzmann Solver software on PyMOL (Fig. 4b; supplementary figure S11, Supplementary Material online) (Jurrus et al. 2018). The monomeric structure of DfrB-TS shown in Fig. 5 was generated with ColabFold (Mirdita et al. 2022).

Trimethoprim Resistance Assay in *E. coli*

The in vivo trimethoprim resistance assays were performed in triplicate according to (Wiegand et al. 2008) using the agar method. Briefly, *E. coli* BL21(DE3) cells transformed with the homologs were propagated overnight in Luria–Bertani (LB) media with 50 μg mL^{-1} kanamycin. An inoculum of 10^4 colony-forming units was spotted on 5% methanol LB agar plates containing 0.25 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (ThermoFisher) and trimethoprim (Millipore Sigma) in 2-fold concentration steps up to 600 μg mL^{-1}; the latter is the highest concentration of trimethoprim soluble in a final concentration

of 5% methanol. The trimethoprim concentration inhibiting bacterial growth following 42 h incubation at 37 °C was considered to be the minimal inhibitory concentration (MIC). Homologs with a MIC of 75 µg/mL and higher are considered to be trimethoprim-resistant (Fig. 3b). The cTEM-19m ß-lactamase was used as a negative control (Clouthier et al. 2012), and DfrB1 was used as a positive control.

Thermotolerance

The substrates DHF (synthesized as previously reported [Blakley 1960]) and NADPH (Chem Impex) were quantified spectrophotometrically in 50 mM pH 7 potassium phosphate buffer ($\varepsilon^{DHF}_{282nm}$ 28,400 $M^{-1}cm^{-1}$ and $\varepsilon^{NADPH}_{340nm}$ 6,230 $M^{-1}cm^{-1}$). Of the 148 homologs, 9 were previously characterized (Toulouse et al. 2020; Cellier-Goetghebeur et al. 2022). We thus carried out the analysis of the remaining 139 homologs. The lysis of *E. coli* expressing the 139 homologs and the thermotolerance assay in clarified lysate were performed as previously reported (Lemay-St-Denis et al. 2023). Briefly, for the standard kinetic assay, 10 µL of clarified lysate was added to 90 µM NADPH and 90 µM DHF in 50 mM potassium phosphate buffer, pH 7, in a 96-well UV-transparent plate (Corning) to a final volume of 100 µL. Enzyme activity was determined by monitoring the depletion of DHF and NADPH at 340 nm using a plate reader (Beckman Coulter DTX 880). The initial rate of the reaction was determined during the first 20% of the reaction (substrate to product conversion) with the depletion of NADPH and DHF at 340 nm in 96-well plates ($\Delta\varepsilon_{340nm}$ 12,300 $M^{-1}cm^{-1}$ to determine product formation), which corresponds to the activity before heat treatment. Depletion of NADPH and DHF in negative controls, lysates from E. coli overexpressing cTEM-19m ß-lactamase (Clouthier et al. 2012), was subtracted from the initial reaction rate of each homolog to account for nonspecific signal. To determine thermostability, lysates were incubated at 50 or 75 ° for 10 min in 96-well plates. The plates were centrifuged at 2,000 × *g* for 30 min, and the supernatant was collected and characterized as above for residual DHF reductase activity (Lemay-St-Denis et al. 2023). Assays were performed in triplicate. DfrB1 was used as a positive control.

Productive Affinity of DfrB Homologs for NADPH in Lysate

Inspired by Tamer et al. (2019), productive affinity for NADPH was determined to calculate the half maximum concentration constant ($K_{0.5}$) of DfrB homologs for binding to NADPH, as presented in supplementary figure S22, Supplementary Material online. This was performed using "single curve" kinetic analysis to monitor complete depletion of NADPH in lysate incubated with DHF; this approach allows parallel analysis of kinetic data extracted from a single reaction curve instead of requiring multiple reaction curves, facilitating analysis of multiple homologs (Tamer et al. 2019). First, a calibration curve was generated. Seven reactions using NADPH concentrations ranging from 3.1 to 50 µM, always with 250 µM DHF and 0.51 µM purified DfrB1, were followed over 2.5 h. To correct for the nonenzymatic decrease in signal, a control sample was prepared for each substrate combination in which the enzyme was omitted; we note that nonspecific background was negligible. After correction, the variation in absorbance over the entire reaction (initial absorbance–final absorbance) was calculated for each curve and plotted in relation to [NADPH]. The corresponding calibration curve, with R^2 of 0.98 between the difference in absorbance at 340 nm and [NADPH], was generated.

Homologs for which significant DHF reductase activity was detected in lysate, for a total of 24, were expressed in *E. coli* BL21(DE3) in 96 deep-well plates containing 1 mL of ZYP-5053 medium, 20 µL inoculum, and were incubated at 37 °C for 3 h, and for 16 h at 22 °C, always under agitation. The plates were centrifuged for 30 min at 4 °C, at 2,000 × *g*. The cell pellets were resuspended in 300 µL of lysis buffer (100 mM pH 8 potassium phosphate buffer, 10 mM $MgSO_4$, 1 mM DTT, 0.5 mg/mL lysozyme, 1.5 mM benzamidine, 0.25 mM PMSF, and 0.4 U DNAse), agitated for 2 h at 22 °C, and then centrifuged. A Beckman Coulter Biomek NXp robot was used to transfer the clarified lysate into 96-well plates. CalA lipase (Alejaldre et al. 2023) was used as a negative control for activity, while DfrBH-5 (Lemay-St-Denis et al. 2023) (here corresponding to A0A1I4V6W3) and DfrB1 were used as positive controls for activity. Lysate concentrations were 40% (v/v) in 50 µL wells containing 250 µM DHF and 30 µM NADPH in 50 mM pH 7 potassium phosphate buffer. Absorbance depletion at 340 nm was monitored during 85 min to capture the totality of the reaction. We controlled for DHF and NADPH degradation by monitoring the depletion of absorbance for each substrate, incubating it individually with DfrB1 lysate. Assays were carried out in triplicate. The OD measurements were translated in [NADPH] with the calibration curve, and the data were then fitted in the model below in Excel using the Solver tool to extract the parameters V_{max}, $K_{0.5}$, and h, the Hill coefficient, a measure of cooperativity (Fig. 5a).

$$[S]_i = [S]_{i-1} - (t_i - t_{i-1})\frac{V_{max}[S]^h_{i-1}}{K^h_{0.5} + [S]^h_{i-1}}$$

Protein Purification

Expression of His-tagged proteins transformed in *E. coli* BL21(DE3) was carried out as follows. Overnight precultures of 5 mL were inoculated into 500 mL of Terrific Broth media containing 50 µg mL^{-1} kanamycin (Sigma). After initial growth at 37 °C to an OD_{600nm} of 0.6, cells were induced by 1 mM IPTG (ThermoFisher) and expression was carried out at 30 °C overnight. The cells were harvested, resuspended in IMAC A buffer (600 mM NaCl, 50 mM Tris, 1 mM $CaCl_2$, 20 mM imidazole, and pH 8), and incubated for 15 min at 4 °C with 0.5 mg mL^{-1} lysozyme (Sigma). The cells were lysed by sonication and centrifuged at 16,000 × *g* (Sorvall SLA-3000) at 4 °C for 20 min. The supernatant was filtered with a 0.2 µm filter, injected onto a HisTrap FF column (Cytiva), and eluted with IMAC B buffer (600 mM NaCl, 50 mM Tris, 1 mM $CaCl_2$, 500 mM imidazole, and pH 8). Buffer exchange and concentration of protein fractions were carried out with Amicon Ultra Centrifugal Filter Units of either 3 or 30 K molecular weight cutoffs (Fisher), in 50 mM pH 8 potassium phosphate buffer. Pure fractions were pooled together and concentrated. The exact mass was confirmed by the Regional Mass Spectrometry Centre at Université de Montréal.

Kinetic Characterization Using Purified DfrB

Kinetic assays were performed in a 1 cm pathlength quartz cuvette at 27 °C in a Cary 100 Bio UV-Visible (Agilent) spectrophotometer.

For the characterization of DHF reductase activity, DHF and NADPH were quantified as described in the section titled "Thermotolerance". During the assays, the initial rate of linear depletion of NADPH and DHF was monitored in triplicate at 340 nm ($\Delta\varepsilon_{340nm}$ 12,300 M^{-1} cm^{-1} (Baccanari et al. 1975)) in 50 mM pH 7 potassium phosphate buffer. Any nonspecific decrease in absorbance was accounted for by monitoring the absorbance for the same period in reactions where the purified enzyme was replaced by buffer. For the determination of K_M^{DHF} and K_M^{NADPH}, the concentration range of the variable substrate varied while the second substrate was kept at a concentration of 50 µM.

For characterization of the TS activity of DfrB-TS, measurements were performed in triplicate in a buffer with 100 mM Tris, 50 mM ß-mercaptoethanol, and 1 mM EDTA at pH 7.3. The substrate (6R)-5,10-CH_2-H_4folate (CH_2H_4THF) was prepared by incubating for 10 min at room temperature a solution of 2 mM of tetrahydrofolate (THF, Sigma) in 0.038% formaldehyde (Sigma), and dUMP (Sigma) was prepared in the buffer. The assay was monitored at 340 nm ($\Delta\varepsilon_{340nm}$ 6,400 M^{-1} cm^{-1} (Spencer et al. 1997)), where the production of DHF is detected. Any nonspecific decrease in absorbance was accounted for by monitoring the absorbance for the same period in reactions where the purified enzyme was replaced by buffer. For the determination of K_M^{DHF} and K_M^{NADPH}, the concentration range of the variable substrate varied while the second substrate was kept at a concentration of either 40 or 50 µM.

Data were fitted to the Michaelis-Menten equation using nonlinear regression analysis GraphPad Prism version 7 for Mac (GraphPad Software, San Diego, CA, USA, Table 1). Standard deviation is shown.

For the assay following the bifunctional activity of DfrB-TS (supplementary figure S21, Supplementary Material online), the substrates were prepared in a buffer with 100 mM Tris, 50 mM ß-mercaptoethanol, and 1 mM EDTA at pH 7.3. Absorbance was monitored at 340 nm for 10 min for different combinations of 80 µM CH_2H_4THF, 50 µM NADPH, and 50 µM dUMP mixed with the enzyme.

Size Exclusion Chromatography

The oligomerization states of the DfrB homologs were analyzed using analytical SEC with an ÄKTA fast protein liquid chromatography system (Table 1). The 2.4 mL size exclusion column (Superdex 200 Increase 3.2/300, Cytiva) was calibrated with the Cytiva Gel Filtration Calibration Kit and with lysozyme (Fisher). Injections of 10 µL of protein at 3 mg/mL (corresponding to concentrations ranging from 70 to 365 µM) were applied to the column equilibrated with 50 mM potassium phosphate, pH 8, at a flow rate of 0.075 mL min^{-1}. Each protein was injected in triplicate.

Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

This research is dedicated to the memory of Maxime St-Aubin, our beloved colleague. We thank Samy Cecioni and Andreea R. Schmitzer for providing access to their instruments, Jeffrey W. Keillor for fruitful insights into single curve kinetic analysis, and Uri Gophna for proofreading.

Author Contributions

C.L.-S.-D.: conceptualization, methodology, investigation, formal analysis, visualization, writing—original draft, and writing—review and editing; S.C.-G.: investigation, formal analysis, visualization, and writing—review and editing; M.S.-A.: investigation and formal analysis; K.I.: investigation; J.N.C.: investigation; S.T.: resources; N.B.T.: conceptualization and writing—review and editing; R.K.: conceptualization, methodology, software, and writing—review and editing; and J.N.P.: conceptualization, supervision, funding acquisition, and writing—review and editing.

Funding

This work was supported by the Natural Science and Engineering Research Council of Canada (NSERC) discovery grant RGPIN-N-2018-04686 and the Canada Research Chair in Engineering of Applied Proteins (J.N.P.). R.K. and N.B.-T. were supported by the Israel Science Foundation (Grant/Award Number: 1764/21). N.B.-T.'s research is supported in part by the Abraham E. Kazan Chair in Structural Biology, Tel Aviv University, and the Minerva Foundation. C.L.-S.-D was supported by scholarships from NSERC, Hydro-Québec, APRENTICE, and MITACS, as well as by the Zuckerman STEM Leadership Program. S.C.-G. was supported by a scholarship from NSERC. M.S.-A. was supported by a scholarship from APRENTICE.

Conflicts of Interest

K.I. and S.T. were employees of bitBiome, Inc. at the time of the study.

Data Availability Statement

The data underlying this article are available in the article and in its online supplementary material.

References

Alcock BP *et al.* CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 2019:48:D517–D525. https://doi.org/10.1093/nar/gkz935.

Alejaldre L, Lemay-St-Denis C, Pelletier JN, Quaglia D. Tuning selectivity in CalA lipase: beyond tunnel engineering. *Biochemistry.* 2023:62:396–409. https://doi.org/10.1021/acs.biochem.2c00513.

Altschul S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997:25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.

Alvarez-Carreño C, Penev PI, Petrov AS, Williams LD. Fold evolution before LUCA: common ancestry of SH3 domains and OB domains. *Mol Biol Evol.* 2021:38:5134–5143. https://doi.org/10.1093/molbev/msab240.

Baccanari D, Phillips A, Smith S, Sinski D, Burchall J. Purification and properties of Escherichia coli dihydrofolate reductase. *Biochemistry.* 1975:14:5267–5273. https://doi.org/10.1021/bi00695a006.

Balestrazzi A, Branzoni M, Carbonera D, Parisi B, Cella R. Biochemical evidence for the presence of a bifunctional dihydrofolate reductase-thymidylate synthase in plant species. *J Plant Physiol.* 1995:147:263–266. https://doi.org/10.1016/S0176-1617(11)81515-4.

Blakely G *et al.* Two related recombinases are required for site-specific recombination at dif and cer in *E. coli* K12. *Cell.* 1993:75:351–361. https://doi.org/10.1016/0092-8674(93)80076-Q.

Blakley RL. Crystalline dihydropteroylglutamic acid. *Nature.* 1960:188:231–232. https://doi.org/10.1038/188231a0.

Bodenreider C, Kellershohn N, Goldberg ME, Méjean A. Kinetic analysis of R67 dihydrofolate reductase folding: from the unfolded monomer to the native tetramer. *Biochemistry*. 2002:41:14988–14999. https://doi.org/10.1021/bi020453b.

Bradrick TD, Beechem JM, Howell EE. Unusual binding stoichiometries and cooperativity are observed during binary and ternary complex formation in the single active pore of R67 dihydrofolate reductase, a D_2 symmetric protein. *Biochemistry*. 1996:35:11414–11424. https://doi.org/10.1021/bi960205d.

Burnouf DY *et al.* Structural and biochemical analysis of sliding clamp/ligand interactions suggest a competition between replicative and translesion DNA polymerases. *J Mol Biol*. 2004:335:1187–1197. https://doi.org/10.1016/j.jmb.2003.11.049.

Canteón R. Antibiotic resistance genes from the environment: a perspective through newly identified antibiotic resistance mechanisms in the clinical setting. *Clin Microbiol Infect*. 2009:15:20–25. https://doi.org/10.1111/j.1469-0691.2008.02679.x.

Carreras CW, Santi DV. The catalytic mechanism and structure of thymidylate synthase. *Annu Rev Biochem*. 1995:64:721–762. https://doi.org/10.1146/annurev.bi.64.070195.003445.

Cellier-Goetghebeur S *et al.* Discovery of highly trimethoprim-resistant DfrB dihydrofolate reductases in diverse environmental settings suggests an evolutionary advantage unrelated to antibiotic resistance. *Antibiotics*. 2022:11:1768. https://doi.org/10.3390/antibiotics11121768.

Cheng H *et al.* ECOD: an evolutionary classification of protein domains. *PLoS Comput Biol*. 2014:10:e1003926. https://doi.org/10.1371/journal.pcbi.1003926.

Chopra S, Lynch R, Kim S-H, Jackson M, Howell EE. Effects of temperature and viscosity on R67 dihydrofolate reductase catalysis. *Biochemistry*. 2006:45:6596–6605. https://doi.org/10.1021/bi052504l.

Clifton BE *et al.* Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. *Nat Chem Biol*. 2018:14:542–547. https://doi.org/10.1038/s41589-018-0043-2.

Clouthier CM *et al.* Chimeric β-lactamases: global conservation of parental function and fast time-scale dynamics with increased slow motions. *PLoS One*. 2012:7:e52283. https://doi.org/10.1371/journal.pone.0052283.

Copley SD. An evolutionary biochemist's perspective on promiscuity. *Trends Biochem Sci*. 2015:40:72–78. https://doi.org/10.1016/j.tibs.2014.12.004.

Dam J, Blondel A. Effect of multiple symmetries on the association of R67 DHFR subunits bearing interfacial complementing mutations. *Protein Sci*. 2004:13:1–14. https://doi.org/10.1110/ps.03309504.

Dandekar T. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci*. 1998:23:324–328. https://doi.org/10.1016/S0968-0004(98)01274-2.

Dionne U, Percival LJ, Chartier FJM, Landry CR, Bisson N. SRC homology 3 domains: multifaceted binding modules. *Trends Biochem Sci*. 2022:47:772–784. https://doi.org/10.1016/j.tibs.2022.04.005.

Duff MR, Chopra S, Strader MB, Agarwal PK, Howell EE. Tales of dihydrofolate binding to R67 dihydrofolate reductase. *Biochemistry*. 2016:55:133–145. https://doi.org/10.1021/acs.biochem.5b00981.

Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol*. 2011:7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

Evans R, *et al.* 2021 Mar 10. Protein complex prediction with AlphaFold-Multimer. Bioinformatics [preprint]. bioRxiv. https://doi.org/10.1101/2021.10.04.463034

Ferone R, Roland S. Dihydrofolate reductase: thymidylate synthase, a bifunctional polypeptide from *Crithidia fasciculata*. *Proc Natl Acad Sci U S A*. 1980:77:5802–5806. https://doi.org/10.1073/pnas.77.10.5802.

Fleming MP, Datta N, Gruneberg RN. Trimethoprim resistance determined by R factors. *BMJ*. 1972:1:726–728. https://doi.org/10.1136/bmj.1.5802.726.

Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012:28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

Galperin MY *et al.* COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res*. 2021:49:D274–D281. https://doi.org/10.1093/nar/gkaa1018.

Harms MJ. Enzymes emerge by upcycling. *Nat Chem Biol*. 2018:14:526–527. https://doi.org/10.1038/s41589-018-0064-x.

Hicks SN, Smiley RD, Stinnett LG, Minor KH, Howell EE. Role of lys-32 residues in R67 dihydrofolate reductase probed by asymmetric mutations. *J Biol Chem*. 2004:279:46995–47002. https://doi.org/10.1074/jbc.M404484200.

Howell EE. Searching sequence space: two different approaches to dihydrofolate reductase catalysis. *ChemBioChem*. 2005:6:590–600. https://doi.org/10.1002/cbic.200400237.

Ivanetich KM, Santi DV. Bifunctional thymidylate synthase-dihydrofolate reductase in protozoa. *FASEB J*. 1990:4:1591–1597. https://doi.org/10.1096/fasebj.4.6.2180768.

Jackson M *et al.* Calorimetric studies of ligand binding in R67 dihydrofolate reductase. *Biochemistry*. 2005:44:12420–12433. https://doi.org/10.1021/bi050881s.

Jumper J *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021:596:583–589. https://doi.org/10.1038/s41586-021-03819-2.

Jurrus E *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci*. 2018:27:112–128. https://doi.org/10.1002/pro.3280.

Kaltenbach M *et al.* Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat Chem Biol*. 2018:14:548–555. https://doi.org/10.1038/s41589-018-0042-3.

Kamath G, Howell EE, Agarwal PK. The tail wagging the dog: insights into catalysis in R67 dihydrofolate reductase. *Biochemistry*. 2010:49:9078–9088. https://doi.org/10.1021/bi1007222.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013:30:772–780. https://doi.org/10.1093/molbev/mst010.

Kaur G, Subramanian S. Repurposing TRASH: emergence of the enzyme organomercurial lyase from a non-catalytic zinc finger scaffold. *J Struct Biol*. 2014:188:16–21. https://doi.org/10.1016/j.jsb.2014.09.001.

Khersonsky O, Roodveldt C, Tawfik D. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol*. 2006:10:498–508. https://doi.org/10.1016/j.cbpa.2006.08.011.

Kishan K, Agrawal V. SH3-like fold proteins are structurally conserved and functionally divergent. *Curr Protein Pept Sci*. 2005:6:143–150. https://doi.org/10.2174/1389203053545444.

Kneis D *et al.* Trimethoprim resistance in surface and wastewater is mediated by contrasting variants of the dfrB gene. *ISME J*. 2023:17:1455–1466. https://doi.org/10.1038/s41396-023-01460-7.

Kneis D, Berendonk TU, Forslund SK, Hess S. Antibiotic resistance genes in river biofilms: a metagenomic approach toward the identification of sources and candidate hosts. *Environ Sci Technol*. 2022:56:14913–14922. https://doi.org/10.1021/acs.est.2c00370.

Kneis D, Tskhay F, Barron DLC, Berendonk M, U T. Bacteria of the order Burkholderiales are original environmental hosts of type II trimethoprim resistance genes (dfrB). *ISME J*. 2024:18:wrae243. https://doi.org/10.1093/ismejo/wrae243.

Krahn JM, Jackson MR, DeRose EF, Howell EE, London RE. Crystal structure of a type II dihydrofolate reductase catalytic ternary complex. *Biochemistry*. 2007:46:14878–14888. https://doi.org/10.1021/bi701532r.

Lazar G, Zhang H, Goodman HM. The origin of the bifunctional dihydrofolate reductasethymidylate synthase isogenes of *Arabidopsis thaliana*. *Plant J*. 1993:3:657–668. https://doi.org/10.1111/j.1365-313X.1993.00657.x.

Lemay-St-Denis C *et al.* A conserved SH3-like fold in diverse putative proteins tetramerizes into an oxidoreductase providing an antimicrobial resistance phenotype. *Philos Trans R Soc B Biol Sci*. 2023:378:20220040. https://doi.org/10.1098/rstb.2022.0040.

Lemay-St-Denis C, Diwan S-S, Pelletier JN. The bacterial genomic context of highly trimethoprim-resistant DfrB dihydrofolate reductases highlights an emerging threat to public health. *Antibiotics*. 2021:10:433. https://doi.org/10.3390/antibiotics10040433.

Lemay-St-Denis C, Pelletier JN. From a binding module to essential catalytic activity: how nature stumbled on a good thing. *Chem*

Commun. 2023:59:12560–12572. https://doi.org/10.1039/D3CC04209J.
Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 2016:44:W242–W245. https://doi.org/10.1093/nar/gkw290.
Levin I, Giladi M, Gophna U. Lateral gene transfer and the synthesis of thymidine. In: Gophna U, editor. *Lateral gene transfer in evolution*. New York (NY): Springer; 2013. p. 3–14. https://link.springer.com/10.1007/978-1-4614-7780-8_1
Levy ED, Erba EB, Robinson CV, Teichmann SA. Assembly reflects evolution of protein complexes. *Nature*. 2008:453:1262–1265. https://doi.org/10.1038/nature06942.
Li D *et al*. Interligand overhauser effects in type II dihydrofolate reductase. *Biochemistry*. 2001:40:4242–4252. https://doi.org/10.1021/bi0026425.
Mirdita M *et al*. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res*. 2017:45:D170–D176. https://doi.org/10.1093/nar/gkw1081.
Mirdita M *et al*. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022:19:679–682. https://doi.org/10.1038/s41592-022-01488-1.
Mistry J *et al*. Pfam: the protein families database in 2021. *Nucleic Acids Res*. 2021:49:D412–D419. https://doi.org/10.1093/nar/gkaa913.
Narayana N, Matthews DA, Howell EE, Xuong N. A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel D2-symmetric active site. *Nat Struct Mol Biol*. 1995:2:1018–1025. https://doi.org/10.1038/nsb1195-1018.
Néron B *et al*. IntegronFinder 2.0: identification and analysis of integrons across bacteria, with a focus on antibiotic resistance in Klebsiella. *Microorganisms*. 2022:10:700. https://doi.org/10.3390/microorganisms10040700.
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015:32:268–274. https://doi.org/10.1093/molbev/msu300.
Nichols R *et al*. Titration of histidine 62 in R67 dihydrofolate reductase is linked to a tetramer .tautm. two-dimer equilibrium. *Biochemistry*. 1993:32:1695–1706. https://doi.org/10.1021/bi00058a002.
Nishikawa Y *et al*. Validation of the application of gel beads-based single-cell genome sequencing platform to soil and seawater. *ISME Commun*. 2022:2:92. https://doi.org/10.1038/s43705-022-00179-4.
Noall EWP, Sewards HFG, Waterworth PM. Successful treatment of a case of Proteus septicaemia. *BMJ*. 1962:2:1101–1102. https://doi.org/10.1136/bmj.2.5312.1101.
Nomburg J *et al*. Birth of protein folds and functions in the virome. *Nature*. 2024:633:710–717. https://doi.org/10.1038/s41586-024-07809-y.
O'Brien PJ, Herschlag D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol*. 1999:6:R91–R105. https://doi.org/10.1016/S1074-5521(99)80033-7.
Ortmayer M *et al*. An oxidative N-demethylase reveals PAS transition from ubiquitous sensor to enzyme. *Nature*. 2016:539:593–597. https://doi.org/10.1038/nature20159.
Park H, Bradrick TD, Howell EE. A glutamine 67–> histidine mutation in homotetrameric R67 dihydrofolate reductase results in four mutations per single active site pore and causes substantial substrate and cofactor inhibition. *Protein Eng Des Sel*. 1997:10:1415–1424. https://doi.org/10.1093/protein/10.12.1415.
Pham DN, Wu Q, Li M. Global profiling of antibiotic resistomes in maize rhizospheres. *Arch Microbiol*. 2023:205:89. https://doi.org/10.1007/s00203-023-03424-z.
Reece LJ, Nichols R, Ogden RC, Howell EE. Construction of a synthetic gene for an R-plasmid-encoded dihydrofolate reductase and studies on the role of the N-terminus in the protein. *Biochemistry*. 1991:30:10895–10904. https://doi.org/10.1021/bi00109a013.
Richardson L *et al*. MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Res*. 2023:51:D753–D759. https://doi.org/10.1093/nar/gkac1080.
Schmitzer AR, Lépine F, Pelletier JN. Combinatorial exploration of the catalytic site of a drug-resistant dihydrofolate reductase: creating alternative functional configurations. *Protein Eng Des Sel*. 2004:17:809–819. https://doi.org/10.1093/protein/gzh090.
Schnell JR, Dyson HJ, Wright PE. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu Rev Biophys Biomol Struct*. 2004:33:119–140. https://doi.org/10.1146/annurev.biophys.33.110502.133613.
Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014:30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
Shannon P *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003:13:2498–2504. https://doi.org/10.1101/gr.1239303.
Smiley RD, Stinnett LG, Saxton AM, Howell EE. Breaking symmetry: mutations engineered into R67 dihydrofolate reductase, a D_2 symmetric homotetramer possessing a single active site pore. *Biochemistry*. 2002:41:15664–15675. https://doi.org/10.1021/bi026676j.
Spencer HT, Villafranca JE, Appleman JR. Kinetic scheme for thymidylate synthase from *Escherichia coli*: determination from measurements of ligand binding, primary and secondary isotope effects, and pre-steady-state catalysis. *Biochemistry*. 1997:36:4212–4222. https://doi.org/10.1021/bi961794q.
Stechmann A, Cavalier-Smith T. Rooting the eukaryote tree by using a derived gene fusion. *Science*. 2002:297:89–91. https://doi.org/10.1126/science.1071196.
Steinegger M *et al*. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinform*. 2019:20:473. https://doi.org/10.1186/s12859-019-3019-7.
Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. 2017:35:1026–1028. https://doi.org/10.1038/nbt.3988.
Suzek BE *et al*. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*. 2015:31:926–932. https://doi.org/10.1093/bioinformatics/btu739.
Tamer YT *et al*. High-order epistasis in catalytic power of dihydrofolate reductase gives rise to a rugged fitness landscape in the presence of trimethoprim selection. *Mol Biol Evol*. 2019:36:1533–1550. https://doi.org/10.1093/molbev/msz086.
The Galaxy Community *et al*. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res*. 2022:50:W345–W351. https://doi.org/10.1093/nar/gkac247.
The UniProt Consortium *et al*. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res*. 2023:51:D523–D531. https://doi.org/10.1093/nar/gkac1052.
Toulouse JL *et al*. Dual-target inhibitors of the folate pathway inhibit intrinsically trimethoprim-resistant DfrB dihydrofolate reductases. *ACS Med Chem Lett*. 2020:11:2261–2267. https://doi.org/10.1021/acsmedchemlett.0c00393.
Trujillo M, Donald RGK, Roos DS, Greene PJ, Santi DV. Heterologous expression and characterization of the bifunctional dihydrofolate reductase –thymidylate synthase enzyme of *Toxoplasma gondii*. *Biochemistry*. 1996:35:6366–6374. https://doi.org/10.1021/bi952923q.
Van Kempen M *et al*. Fast and accurate protein structure search with foldseek. *Nat Biotechnol*. 2024:42:243–246. https://doi.org/10.1038/s41587-023-01773-0.
Wallner B. AFsample: improving multimer prediction with AlphaFold using massive sampling. *Bioinformatics*. 2023:39:9. https://doi.org/10.1093/bioinformatics/btad573.
Wickham H. *Ggplot2: elegant graphics for data analysis*. 2nd ed. Springer; 2016.
Wiegand I, Hilpert K, Hancock REW. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc*. 2008:3:163–175. https://doi.org/10.1038/nprot.2007.521.
Yuvaniyama J *et al*. Insights into antifolate resistance from malarial DHFR-TS structures. *Nat Struct Mol Biol*. 2003:10:357–365. https://doi.org/10.1038/nsb921.