A comprehensive map of transient protein-protein interactions and their structural features in the *E. coli* 1-carbon metabolism pathway

Sanchari Bhattacharyya[*], Srivastav Ranganathan, Sourav Chowdhury, Bharat V Adkar, Mark Khrapko and Eugene I Shakhnovich[*]

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford St, Cambridge, MA 02138

[*]Correspondence should be addressed to:
S.B. (sbhattacharyya@fas.harvard.edu) and E.I.S. (shakhnovich@chemistry.harvard.edu)

Abstract

Enzymes in a pathway often form metabolons through weak protein-protein interactions (PPI) that localize and protect labile metabolites. Due to their transient nature, the structural architecture of these enzyme assemblies has largely remained elusive. Here we delineate a complete PPI map of the *E. coli* 1-carbon metabolism pathway using bimolecular fluorescence complementation that can detect transient interactions *in vivo* and show strong intra- and inter-pathway clusters within the folate and purine biosynthesis pathways. Using deep scanning mutagenesis we identify the PPI interface on DHFR. Further, a combination of AlphaFold predictions with meta-dynamics simulations reveal interaction interfaces of all PPI in the pathway, which not only agrees with experimental findings but also reveal that most proteins use conserved “dedicated” interfaces distant from their active sites to interact with multiple partners. Overall, our data reveals the structural architecture of binary PPI complexes in a metabolon in a crucial prokaryotic metabolic pathway.

Introduction

Unlike purified enzymes in a test tube, enzymes inside cells often do not function in isolation. In many cases, they carry out their reaction through formation of multi-enzyme assemblies. Metabolons - a term first coined by Paul Srere - are one such protein complex, which are transiently interacting supramolecular entities formed through weak interactions among sequential enzymes in a metabolic pathway which help to channel metabolites (Robinson and Srere, 1985; Srere, 1987). In higher eukaryotes, there has been evidence for such complexes that dynamically associate and dissociate in response to cellular stimuli, the most notable of them being purinosomes (An et al., 2008; Deng et al., 2012; French et al., 2016). It is generally postulated that formation of a metabolon has several advantages like metabolite channeling, increasing local concentration of metabolites, preventing competing pathways and protection of unstable and cyto-toxic intermediates (Sweetlove and Fernie, 2018; Zhang and Fernie, 2021). Despite this, mechanistic understanding of how proteins in a metabolon interact has been elusive. Moreover, due to the transient nature of the interactions, traditional methods of structural biology have failed to address these questions. Based on information from cross-linking studies and mass spectrometry, Wu and Minteer (Wu and Minteer, 2015) built a low-resolution structure of the tri-carboxylic acid cycle (TCA) metabolon using constraint based protein-protein docking. This study gave the first structural insight into a metabolon and showed evidence of electrostatic channeling upon association of proteins (Wu and Minteer, 2015).

More recently, it has been shown that even prokaryotic cells, which were once thought to be a bag of freely floating enzymes and metabolites, show evidence of metabolon formation and are therefore able to achieve substantial degree of metabolic compartmentalization. In our previous work (Bhattacharyya et al., 2016), using immunoprecipitation experiments, we have shown that Dihydrofolate Reductase (DHFR), a central enzyme of the *E. coli* 1-carbon metabolism pathway, associates with multiple proteins in its functional vicinity via weak interactions, which in turn was also responsible for overexpression toxicity of DHFR. Though precise evidence of substrate channeling was not established, a more recent work (Bhattacharyya et al., 2021) presented compelling circumstantial evidence that much of these interactions involve channeling of substrates, causing diffusion limitation of the substrates inside the cell. To obtain a deeper understanding of the breadth and nature of interactions in prokaryotic enzyme assemblies, in this work we established a complete PPI map of the *E. coli* 1-carbon metabolism pathway using

biomolecular fluorescence complementation assay. Further, using high throughput library generation, screening, along with computational simulations using AlphaFold and metadynamics, we answer several key questions about the structural architecture of such transient enzyme assemblies. We uncover that most proteins utilize conserved structural interfaces dedicated to interacting with multiple partners, thereby behaving as date hubs as opposed to more permanent multi-enzyme assemblies or party hubs. This study not only represents the first detailed investigation of PPIs forming metabolons in prokaryotic enzyme pathways, but also represents the first high-resolution structural insight into the precise molecular architecture of such fleeting miniature protein assemblies.

Results

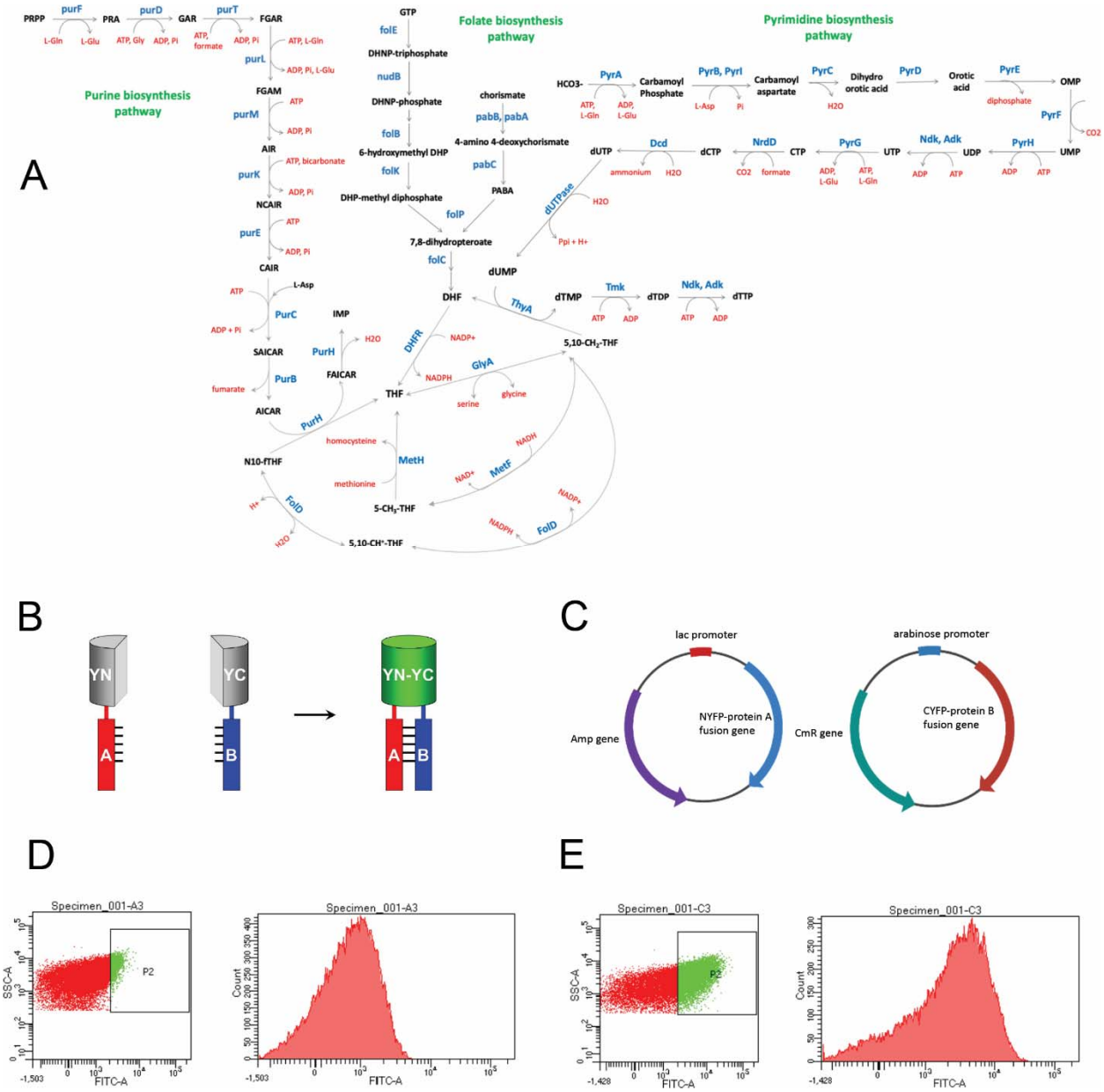

Figure 1: Overview of the *E. coli* 1-carbon metabolism pathway and YFP complementation method used to detect protein-protein interaction (PPI). (A) Purine, pyrimidine and folate biosynthesis pathways converge on to the 1-carbon metabolic pathway. (B) PPI is detected by bimolecular fluorescence complementation assay (BiFc) using the split YFP system wherein N-terminus of YFP molecule is fused to protein A, while the C-terminus is fused to protein B. A mature fluorescent YFP molecule is formed by complementation only if protein A and protein B interact. (C) A mutually compatible two-plasmid based system used for expression of YFP-fusion proteins. NYFP-protein A is expressed from an IPTG-inducible lac promoter, while CYFP-protein B is expressed from an arabinose inducible pBAD promoter. Representative flow-cytometry data for (D) a weakly interacting PPI pair and (E)

strongly interacting pair. The mean fluorescence intensity (MFI) of the histogram is used to assess the extent of interaction.

Split YFP system to measure interactions in the 1-carbon metabolism pathway.

The 1-carbon metabolism pathway is an essential pathway consisting of 3 inter-dependent biosynthetic pathways that catalyze the transfer of 1-carbon units from methyl donors (5,10-methelene THF, 5-methyl THF, N-formyl THF) to generate purine and pyrimidine nucleotides as well as methionine and pantothenic acid. The 1-carbon units themselves are produced through the folate biosynthesis pathway that starts with GTP and chorismate and involves condensation of para-amino benzoate (PABA) with 6-hydroxymethyl-dihydropterin diphosphate to form 7,8-hihydropteroate, which is eventually converted to dihydrofolate (DHF) and tetrahydrofolate (THF) (Fig 1). The pyrimidine biosynthesis pathway intersects the folate pathway at its second last step catalyzed by thymidylate synthase (ThyA), where uridine monophosphate (dUMP) accepts a methyl group from 5,10-methylene THF to convert to thymidine monophosphate (dTMP). The purine biosynthesis pathway utilizes folate derivatives twice: in the third step where the enzyme PurT catalyzes conversion of Glycineamide ribonucleotide (GAR) to formyl-GAR with the help of formate; in the penultimate step where PurH catalyzes the conversion of AICAR to formyl-AICAR using N-formyl THF as a cofactor.

Our previous work (Bhattacharyya et al., 2016) has shown that enzymes in the folate biosynthesis pathway not only interact among themselves but also with those from functionally related purine and pyrimidine biosynthesis pathways. To systematically map interactions among all proteins in these three related pathways, we used a high-throughput way to measure these interactions. We used bimolecular fluorescence complementation assay (BiFc) using the split YFP system to measure interaction between pairs of proteins. In this system, one protein is fused to the N-terminus of YFP, while the other protein is fused to the C-terminus. A mature YFP is formed only if proteins A and B interact (Fig 1B). The two proteins are expressed from two mutually compatible plasmids and using two different inducers (IPTG based and Arabinose based) (Fig 1C). Stronger interactions give higher fluorescence, while weak or non-specific interactions gives low fluorescence (Fig 1D).

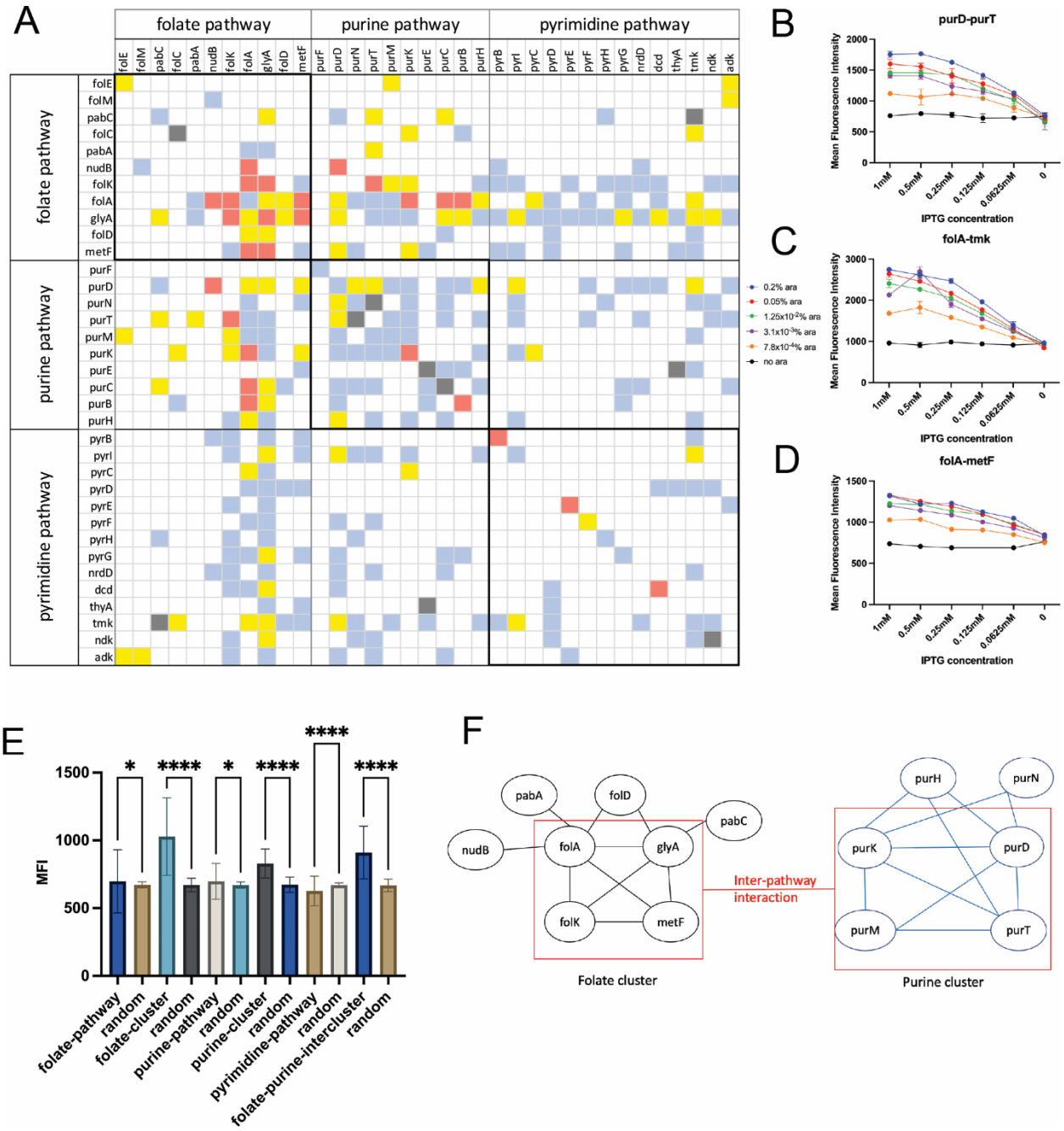

Figure 2: Exploration of the complete PPI network shows strong intra- and inter- pathway cluster of interactions. (A) A total of 35 proteins from the three different pathways were cloned as both N-terminal and C-terminal YFP fusions and interaction of every protein with every other protein (including itself which indicates homodimer probability) was probed using the split-YFP system. The matrix was color coded based on the mean YFP fluorescence of the pair. The interaction was evaluated in both formats (NYFP-A/CYFP-B and NYFP-B/CYFP-A) and the maximum fluorescence data for each pair was used to generate this matrix. Red boxes indicate MFI > 1100, yellow for 900 < MFI < 1100, blue for 750 < MFI < 900 while white boxes indicates MFI < 750. Titration of inducer levels (arabinose and IPTG) to test dependence of fluorescence

values on intracellular protein expression for (B) purD-purT (C) folA-tmk and (D) folA-metF pairs. As expected, higher inducer concentration leads to more complex formation and stronger YFP fluorescence. (E) Certain protein clusters show highly statistically significant interaction (**** indicates p-value < 0.0001) when compared to an equivalent number of random interaction pairs from the table. While there is no significant clustering for all 55 pairs in the folate pathway, a set of 4 proteins (10 PPI pairs for FolK, FolA, GlyA, FolD and MetF) show strong clustering. Similarly, a set of 4 proteins in the purine biosynthesis pathway (PurK, PurD, PurM, PurD) show cluster formation. While no statistically significant clustering is observed among pyrimidine biosynthesis pathway proteins, there is strong evidence of inter-pathway interaction among selected folate and purine biosynthesis enzymes. For each of the clusters, statistical significance is calculated with respect to the mean and standard deviation of a set of *equal* number of interactions that are picked randomly 500 times from the whole data set (represented in the plots as random). (F) A pictorial graph representation of the strongest interactions. Corners represent protein identity while edges represent strong interaction.

Matrix of 1225 interactions shows large variation in PPI strengths.

Out of total 41 proteins in the three pathways, genes corresponding to 35 proteins were cloned successfully as both N- and C-YFP fusions, which were used to generate a 35×35 matrix of 1225 interactions (Fig 2A; this is a symmetric matrix generated by selecting the highest fluorescence value for any given pair). For a few pairs, the bacterial cultures co-transformed with both plasmids grew very poorly, hence significant number of cells could not be analyzed (grey boxes). For the remaining pairs, mean fluorescence intensity (MFI) showed a wide range of variation (Fig S1A). As seen from Fig 2A, vast majority of the interactions are weak (MFI < 750). Several diagonal elements in the table show strong interactions, which correspond to formation of homo-oligomers. Interestingly, several off-diagonal protein pairs showed highly discernible interaction (MFI > 750). We classified them into weak (blue for 750 < MFI < 900), moderate (yellow for 900 < MFI < 1100) and strong (red boxes indicate MFI > 1100) interactions.

Since the N- and C-YFP fusion proteins are induced using IPTG and Arabinose, we next attempted to find out how the fluorescence intensity changes as a function of protein expression levels (or inducer concentration). As expected, we found that increase in inducer concentration causes a sharp increase in YFP fluorescence (Fig 2B, C and D), indicating that the observed fluorescence is not non-specific, rather it arises from specific protein-protein interactions that are strongly concentration dependent.

Folate and purine biosynthesis pathways show multiple strong intra- and inter-pathway interactions.

We observed that certain proteins (FolA and GlyA) have a statistically higher propensity to be involved in PPI than others (Fig S1C). Next, we sought to find out if folate, purine and pyrimidine biosynthesis pathways have significant intra-pathway interactions. To that end, we calculated the average strength of interaction among all proteins in those pathways and compared the mean and standard deviation of an equal number of PPI pairs chosen randomly from the matrix. We found that while the entire pathways do not have significantly stronger interactions than random pairs (Fig 2E), smaller clusters within the pathways do exhibit strong interactions that are statistically significant. For example, five proteins of the folate pathway, namely FolK, FolA, GlyA, FolD and MetF interact strongly with each other, presumably forming an intra-pathway cluster (cluster referred to folate-cluster in Fig 2E, which refers to 10 PPI pairs among the five proteins listed above). Interestingly, these are also sequential proteins in the folate pathway. Similarly, several proteins of the purine biosynthesis pathway (namely PurD, PurN, PurT, PurM and PurK) also form an intra-pathway cluster (purine-cluster in Fig 2E), though the interactions are overall weaker than those of the folate pathway. No such cluster was formed among enzymes from the pyrimidine biosynthesis pathway. The only exception in the pyrimidine pathway is Thymidylate Kinase (Tmk) which does interact with several proteins within its own pathway along with proteins from purine and folate pathways. Interestingly it was shown that Tmk overexpression is toxic, probably due to mis-interactions upon overexpression with other proteins of the folate pathway (Bhattacharyya et al., 2021; Bhattacharyya et al., 2016). Next, we checked for significant inter-pathway interactions among the three pathways. While overall there are no statistically significant inter-pathway interactions between proteins (Fig. S1B), certain clusters of proteins in the folate and purine biosynthesis pathway do show statistically significant inter-pathway interactions (16 PPI pairs referred to as folate-purine inter-cluster in Fig 2E compared to a set of 16 randomly chosen PPI pairs from the whole matrix). Based on the strong interactions observed, we constructed a graphical representation of possible clusters within the pathways and between folate and purine pathways (Fig. 2F). According to this, FolA, FolK, GlyA and MetF form a core cluster in the folate pathway, where each protein interacts with every

other protein. FolD, PabA, PabC and NudB proteins interact with one or two of the core proteins, but do not interact among themselves, hence they form peripheral members of the cluster. A similar analysis shows a core cluster in the purine biosynthesis pathway, comprising PurK, PurD, PurM and PurT, while PurH and PurN are peripheral members.

Since YFP fluorescence is strongly dependent on protein expression levels, one possibility is that the observed interactions in our study are largely due to variations in expression of the different proteins. To that end, we carried out a western blot analysis to quantify expression levels of the 35 N- and C-YFP fusion proteins (Fig S2). The overall expressions of the CYFP-fusion proteins were less than NYFP fusions. Though there were some variations in the expression levels within one set, they were not substantial enough to explain the variations in fluorescence intensities. For example, NYFP-FolA which shows strong interactions with multiple proteins (Fig 2A) had comparable expression to most proteins in the dataset.

Purine biosynthesis metabolon interactome is made up of structurally similar proteins.

We next asked, are there physico-chemical characteristics of the enzymes that dictate their interactions? Theoretical studies have shown that structurally similar proteins tend to interact (Lukatsky et al., 2007). To that end, we computed the structural similarity index of all proteins against all others using a structural similarity measure TM-score (Zhang and Skolnick, 2005) (Fig S3). We first asked if the protein clusters identified from fluorescence data are structurally similar or not. We found that the 5 proteins in the folate cluster (FolK, FolA, GlyA, FolD and MetF, total 10 pairs) have a mildly significant TM score distribution higher than any random 10 pairs from the dataset (p-value = 0.0437) (Fig S4), while the entire set of 55 proteins have no significant structural similarity above the background. On the contrary, the set of 5 proteins from the purine cluster (PurD, PurN, PurT, PurM and PurK, total 9 pairs) have a highly statistically significant greater TM-score distribution than a random set (p-value < 0.0001) (Fig S4). The folate-purine inter-pathway cluster (16 interactions) however has no significant structural similarity. It is worth mentioning at this point that previous studies demonstrated that certain enzymes of the purine biosynthesis pathway evolved from a single common ancestor by gene duplication and divergence (Kappock et al., 2000; Zhang et al., 2008).

To understand this further, we analyzed the fluorescence intensity of all those PPI pairs which have significant structural similarity (TM score > 0.5), both in the extended pathway as well as

within individual pathways. Additionally, we also analyzed TM score distributions of all PPI pairs which have MFI values > 750 (Fig S4). The analysis shows that while a higher TM score overall does not guarantee interaction, it does dictate interaction propensity in the purine biosynthesis pathway.

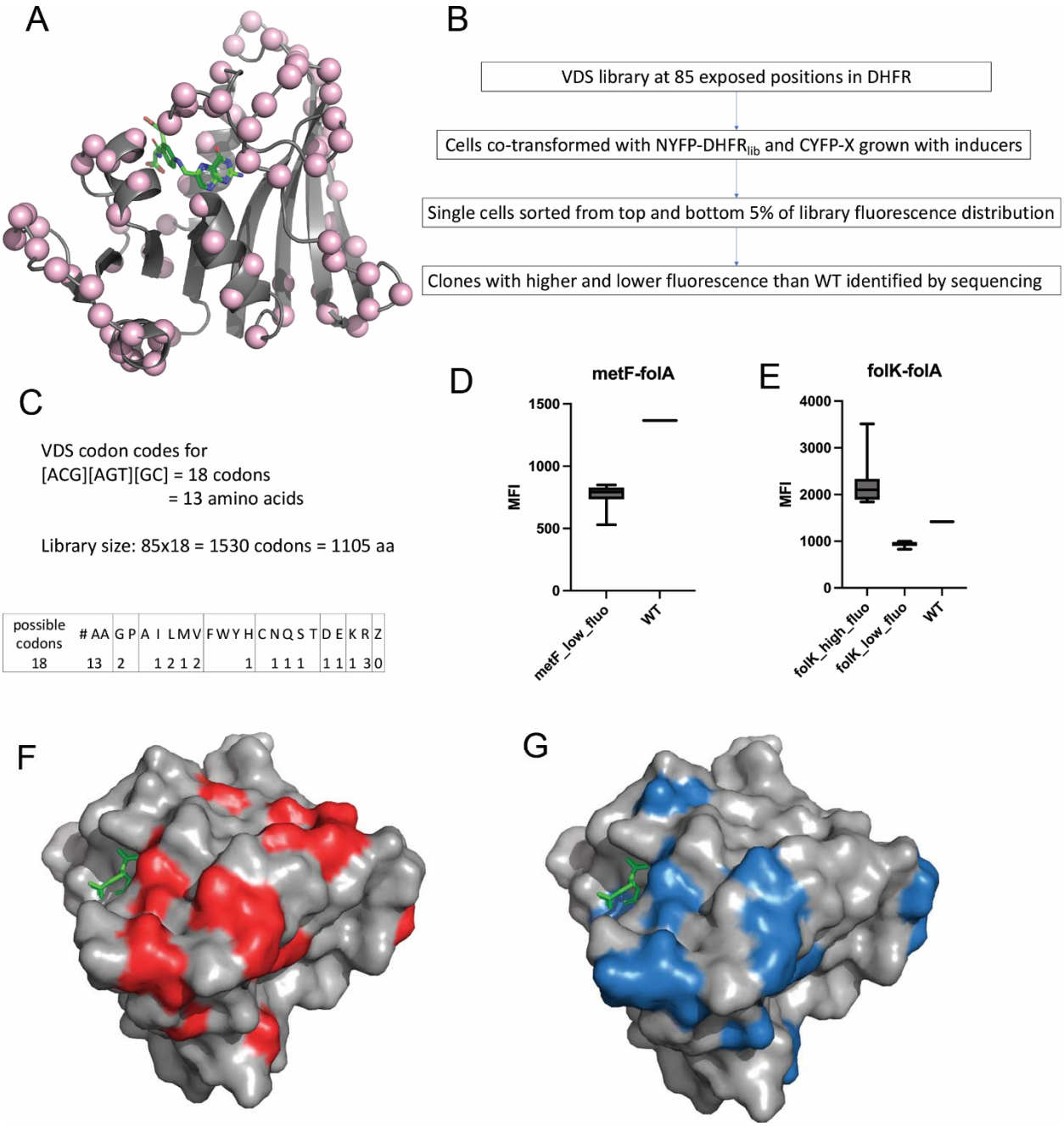

Figure 3: ***Elucidation of PPI binding interface using saturation mutagenesis-based library screening.*** *At each of the 85 exposed locations (% accessibility > 20%) in DHFR protein (Cα residue shown by light pink sphere in (A)), a VDS codon was introduced (one position at a time) by PCR. (C) VDS codes for 18 codons (13*

different amino acids, no stop codons) which results in a total library size of 1105 amino acids. Panel (B) shows a flow-chart of the method used to map the binding interface. The NYFPlib plasmid was transformed with either CYFP-MetF or CYFP-folK expressing plasmids. Single cells were sorted out from the top and bottom 5% of the fluorescence histogram of the library variants, and clones with MFI significantly lower or higher than WT DHFR were sequenced to reveal their identities. Panels (D) and (E) represent box-plots of MFI for the sequenced clones. The mutations were mapped back on the DHFR structure and surprisingly, the two interaction surfaces show very significant non-random overlap (p-value=2e-7, see Methods and text) that is away from the folate binding site (DHF shown in green sticks). Panels F and G are DHFR structures where the MetF and FolK binding residues identified from the library screening are shown in red and blue color respectively.

Saturation mutagenesis reveals dedicated PPI interfaces on DHFR that are conserved between its PPI partners.

Since enzymes in a pathway interact via weak and transient interactions, no structural information is available for such PPI pairs by traditional methods. We reasoned that using our split YFP based methodology to screen for binding, we could highlight mutational hotspots in the PPI pairs that render these complexes stronger or weaker. To that end, we focused on two PPI pairs, FolA-MetF and FolA-FolK, which form one of the strong complexes in our dataset. We selected all surface exposed residues of FolA (accessibility > 80%, shown by pink spheres in Fig 3A), except its active site and contacting residues, and generated a near saturation mutagenesis library using VDS codon. VDS codes for all different amino acid types (hydrophobic, polar, charged), however does not incorporate large hydrophobics (F, W or Y) or proline, and does not incorporate any stop codon (Fig 3C). This is to ensure that the change in fluorescence (if any) would not arise due to protein destabilization or due to a truncated protein. When *E. coli* cells were co-transformed with NYFP-FolAlib and CYFP-MetF/CYFP-FolK (Fig 3B), the overall histogram of fluorescence intensity of the library variants did not appear to be different from WT (Fig S5). On the other hand, single cells sorted from the top and bottom 5% of the distribution were found to have a large variation in fluorescence intensities. We selected 33 clones for the DHFR-MetF interaction and 28 clones for the DHFR-FolK interaction that had considerably higher or lower fluorescence intensities than the WT pair (Fig 3D, 3E), and sequencing was carried out to reveal the identities of the mutants. Surprisingly, we found that a vast majority of the residues overlap between FolA surfaces interacting with FolK and MetF (26, 29, 114, 124,

131, 132, 140-145, 149, 157, 159). When these mutational hotspots were mapped back on to the structure of DHFR, we found the binding site is mostly located in the C-terminal beta hairpin, as well as in the helix close to the beta hairpin. The binding site does not overlap with the active site (folate or NADPH binding site).

To test whether there is a non-random, statistically significant overlap between the FolA interaction surfaces with Folk and MetF, we applied statistical test described in Methods and found that DHFR uses the same surfaces to interact with MetF and FolK: the overlap between two interaction surfaces is highly significant with a p-value of 10^{-7} (see Methods).

A possible alternative explanation for the drop in fluorescence in mutants is that mutations could lead to loss of DHFR protein abundance to unfolding/ misfolding/aggregation caused by DHFR destabilization. To address this possibility, we performed a western blot analysis of the mutants with anti-DHFR antibodies. None of the mutations showed any substantial change in intensity/abundance (Fig S5), indicating that the mutations genuinely perturb binding without affecting abundance.

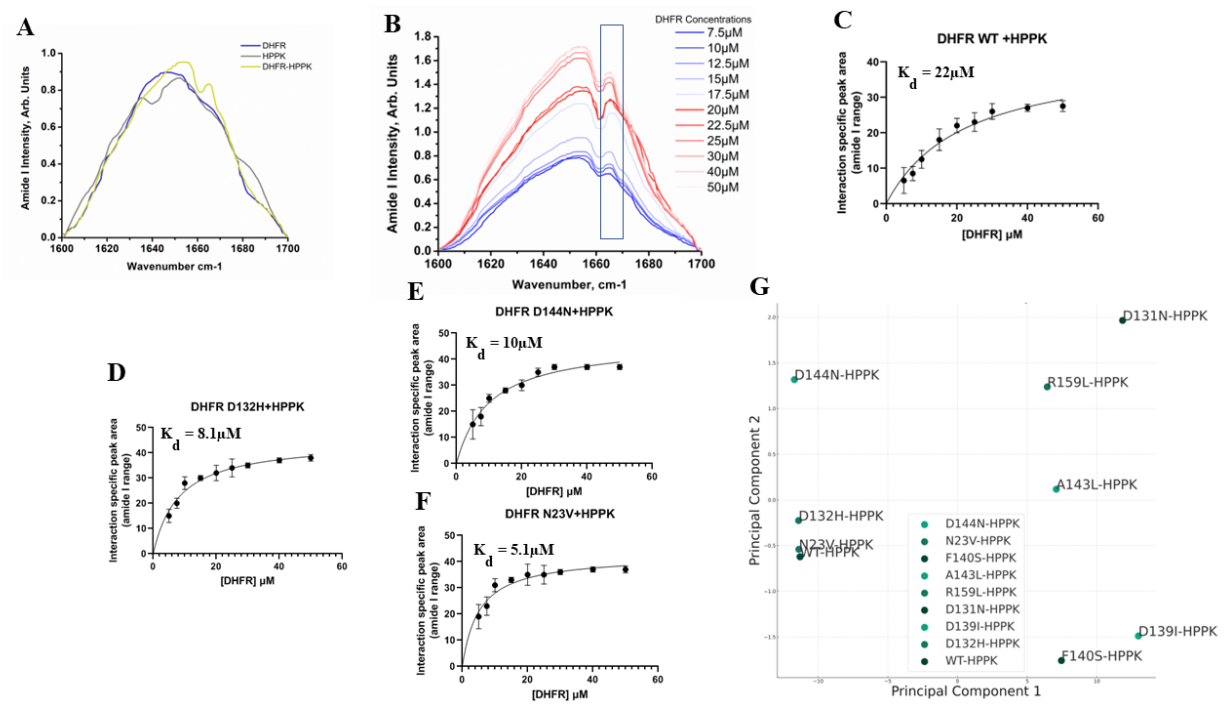

Figure 4: ***Raman Spectroscopic Analysis shows HPPK interacts with WT and select mutants of DHFR. (A)*** *Amide I Raman Spectroscopic profiles of WT-DHFR (blue trace), HPPK (grey trace), and DHFR-HPPK complex (yellow trace) are shown. Intensities (arbitrary unit, Arb. Units) for WT DHFR and HPPK in the amide I range (1600-1700 cm^{-1}) were recorded individually and then for the WT DHFR-HPPK complex. The spectrum for the WT DHFR-*

HPPK complex is not a mathematical sum of individual spectra recorded for WT DHFR and HPPK and shows new spectral features and a characteristic peak at ~1665 cm^{-1} indicating a signal originating from the WT DHFR-HPPK bound complex. ***(B)*** *Amide I spectral profiles of WT DHFR-HPPK complex along an increasing concentration of DHFR (7.5 to 50 µM). The marked section shows the peak at ~1665 cm^{-1} characteristic of WT DHFR-HPPK interaction.* ***(C)*** *Changes in the peak area at ~1665cm^{-1}, characteristic of WT DHFR-HPPK interaction, were plotted against increasing concentrations (7.5µM to 50µM) of DHFR. The peak areas were obtained upon deconvolution of the spectral read-outs (n=3) and the plotted data was subjected to Hyperbolic fitting to derive WT DHFR-HPPK dissociation constant, K_d. K_d for WT DHFR-HPPK interaction was found to be 22µM.* ***(D)*** *Changes in the peak area at ~1665cm^{-1}, characteristic of D132H mutant of DHFR-HPPK interaction, were plotted against increasing concentrations (7.5µM to 50µM) of D132H DHFR. K_d for D132H DHFR-HPPK interaction was found to be 8.1µM.* ***(E)*** *Changes in the peak area at ~1665cm^{-1}, characteristic of D144N mutant of DHFR-HPPK interaction, were plotted against increasing concentrations (7.5µM to 50µM) of D144N DHFR. K_d for D144N DHFR-HPPK interaction was found to be 10µM.* ***(F)*** *Changes in the peak area at ~1665cm^{-1}, characteristic of N23V mutant of DHFR-HPPK interaction, were plotted against increasing concentrations (7.5µM to 50µM) of N23V DHFR. K_d for N23V DHFR-HPPK interaction was found to be 5.1µM.* ***(G)*** *Principal component analysis (PCA) based on the entire Raman spectroscopic profiles (800 to 2000 cm^{-1}) for WT and 8 mutants of DHFR. The positions of the mutants in the plot indicate their relatedness based on their Raman profiles. Mutants that are closely positioned on the plot share similarities in their Raman spectra. PCA shows HPPK-interacting mutants, viz. D132H, N23V, and D144N along with the WT are clustered closely. On the other hand, non-HPPK mutants of DHFR are separately spaced. n=3 biologically independent samples were used for the experiments. All the data are presented as mean values +/- SEM.*

Raman spectroscopy of purified proteins confirms binding site residues.

Next, we sought to validate the interaction between DHFR and HPPK (encoded by the gene folK) and its affinity changing mutants revealed in high throughput mutagenesis. To that end we performed Raman spectroscopy and specifically looked at the spectral features of the amide I range (1600 to 1700 cm^{-1}) for DHFR and HPPK individually and then upon mixing. Amide I spectra originate from the C=0 stretch in the peptide backbone and are extremely informative about protein conformations. Analysis of Amide I spectra helps in the mechanistic understanding of secondary structures and their organization in a protein. Deconvolution of amide I spectra with standard peaks for discrete secondary structural elements viz. helices, sheets, and loops help in the quantification of secondary structure content in a protein (Gallagher, 2009). We performed deconvolution of the amide I spectra of the DHFR and HPPK using Lorentzian peak fitting with pre-assigned peaks to better understand the change in the spectral features upon mixing both the protein solutions to detect non-additivity which would report on interaction between two

proteins. The quality of fitting was checked with reduced chi-squared values, which was typically in between around 1.2.

Secondary structure contents obtained from deconvoluted DHFR (Fig.S6A) and HPPK spectra (Fig.S6B) were compared with secondary structure contents of the crystal structures of DHFR (PDB: 1DRA) and HPPK (PDB: 4M5I) to check the spectral quality and physical condition of the samples. Secondary structure contents for both DHFR (Fig.6SA) and HPPK (Fig.S6B) were found to be comparable to the secondary structure contents of the reference structures.

Upon mixing the DHFR and HPPK protein solutions any change in amide I spectral feature is a potential indication of spectra originating from the conformation of bound protein-protein complex which in turn indicates protein-protein interaction. In the absence of interaction, spectra appear as a mathematical sum of two individual protein spectra. In our study, we therefore investigated spectral features under conditions of DHFR HPPK mixing. Control experiments were performed with ADK mixed individually with DHFR and HPPK. We performed a similar quality check experiment with ADK as was performed with DHFR and HPPK using the crystal structure of ADK as the reference and observed comparable secondary structure contents.

Upon mixing DHFR and HPPK resultant amide I spectra showed new features around 1663cm^{-1} not observed in the individual spectrum of DHFR and HPPK before mixing (Fig.4A). Further the resultant spectrum was not the mathematical sum of individual spectra obtained with DHFR and HPPK before interaction. Interestingly, no new feature in the amide I spectra was observed for control acquisitions where ADK was mixed individually with DHFR (Fig.S6A) and HPPK (Fig.S6B). The spectra obtained in these control acquisitions were equal to the mathematical sum of DHFR and ADK spectra before mixing in the DHFR-ADK mix and similarly HPPK and ADK spectra in the HPPK-ADK mix. The presence of new spectral feature with a peak around 1663cm^{-1} in the Amide I spectral profile of the DHFR-HPPK mix (Fig.4A) is a potential indication of an interaction between HPPK and DHFR and resultant Amide I spectra potentially reflecting the secondary structural organization of the HPPK-DHFR complex in solution.

Next, we carried out the Raman spectroscopy using a range of DHFR concentrations (from 7.5 to 50 μM) and keeping HPPK concentrations fixed at 15 μM. We observed area under the characteristic peak at 1663 cm^{-1} increases monotonically (Fig.4B). The changes in the interaction-specific peak area at ~1663 cm^{-1} were plotted against the increasing DHFR

concentrations (Fig.4C). As the peak chosen was found to be specific for the interaction and no non-specific/unbounded fraction-specific signal was included in the plot, the hyperbolic model was used to fit the data. K_d was obtained upon hyperbolic fitting of the data and was found to be 22 µM for WT DHFR-HPPK interaction (Fig.4C). Quality of the model deployed for fitting the data was assessed by R^2 value and it was found to be 0.9. Next, we carried out similar interaction analyses with HPPK for select mutants of DHFR that are predicted to weaken or strengthen binding based on FACS data (D144N, N23V and D132H increase fluorescence, while D131N, E139I, F140S, A143L, R159L showed loss in fluorescence). Interestingly, D132H mutant of DHFR did show stronger interaction with HPPK with a K_D of 8.1µM (Fig.4D). Similarly compared to WT DHFR, D144N and N23V mutants of DHFR were also found to have a stronger interaction with HPPK with K_D values of 10 and 5.1 µM (Figs. 4E and F). For mutants that show loss in fluorescence, quite strikingly, the characteristic peak at 1663 cm^{-1} was absent upon interaction, and hence an exact K_D could not be derived. While this directly hints at loss in binding, we attempted to compare more global signatures of the Raman spectra among mutants, hence we resorted to Principal Component Analysis (PCA). PCA compares the Raman spectral profiles of select DHFR variants upon interaction with HPPK (Fig.4G). The entire Raman Spectral profile spanning from 800 to 1900cm^{-1} was used for the PCA analysis allowing a comparison of both secondary and higher-order structural characteristics in each of the DHFR-HPPK solution mixes. The horizontal and vertical axes represent the scores of the first and second principal components (PC1 and PC2), respectively. These components, which are orthogonal linear combinations of the original variables, capture the most significant variance inherent in the dataset. The distribution of the DHFR-HPPK complexes along these axes reveals the inherent structural differences and similarities in their Raman spectral profiles. Clusters of closely positioned data points viz. WT-HPPK, N23V-HPPK, D132H-HPPK, and D144N-HPPK suggest these DHFR variants likely form stable complexes with HPPK in solution (Fig.4G), implying similar interaction dynamics and potentially shared structural characteristics. In contrast, those distantly positioned from such clusters might indicate variants with distinct interaction profiles or those that do not readily form complexes in solution and indeed those were the ones that were predicted to show loss in binding (D131N, E139I, F140S, A143L, R159L). The distribution and clustering observed in this PCA plot offer insights into the propensity of

these DHFR variants to interact with HPPK, highlighting the potential influence of the point mutations on the complex formation.

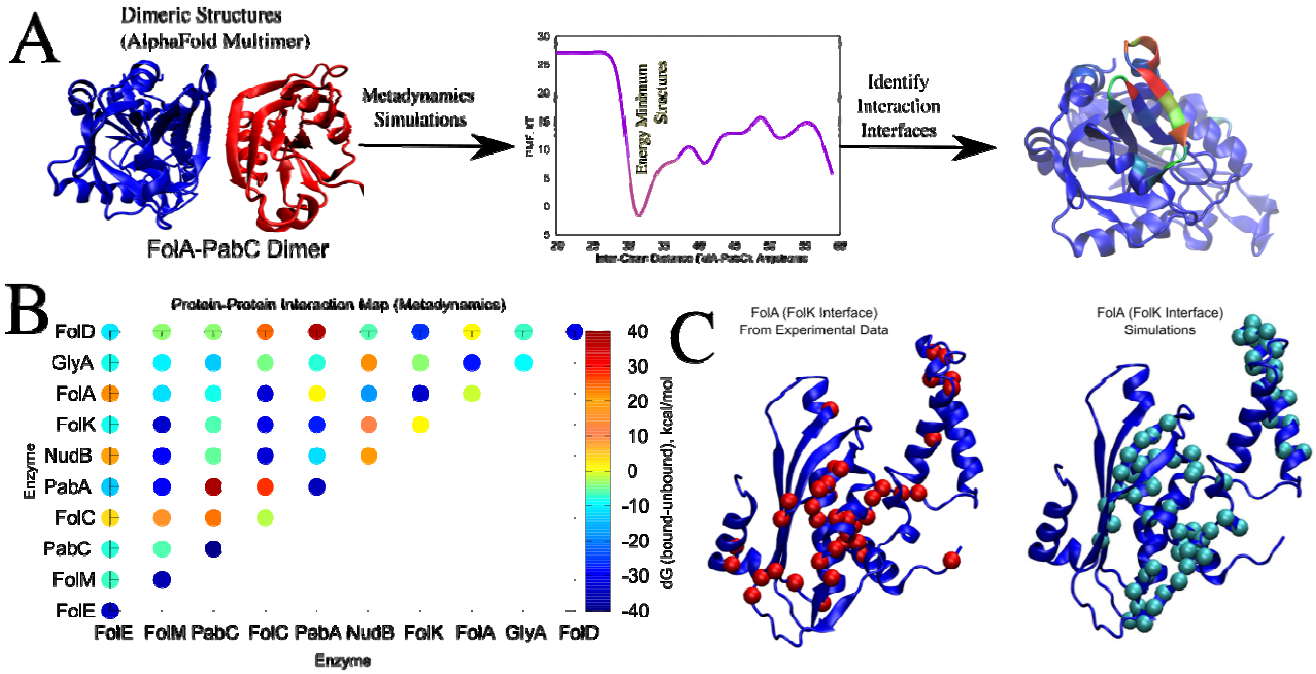

Figure 5. Metadynamics simulations used to compute interaction interfaces and unbinding free energies for folate pathway dimers. *A) AlphaFold multimer is used to make initial dimeric structure predictions for all possible folate pathway enzyme pairs. To further refine these structures and compute their stabilities, metadynamics simulations were performed, with AlphaFold predicted structures as starting points and with inter-chain distance as the collective variable. The structures corresponding to the free energy minima in the PMFs were then used to study the interaction interfaces for folate pathway proteins. B) free energy of binding ΔG(bound-unbound) was determined from the free energy profiles for folate pathway dimers computed using all-atom metadynamics simulations. C) Simulations and experiment predict similar interaction surface for FolA with FolK. Residues of FolA involved in interactions with Folk based on experimental data (marked as red beads in left structure) and metadynamics simulations (marked as cyan beads in right structure). The simulation predicted surface and the experimentally detected interaction surfaces are virtually identical (p-value of* $3x10^{-8}$ *for the null hypothesis that experimental and predicted interaction surfaces have random overlap, see Methods for detail)*

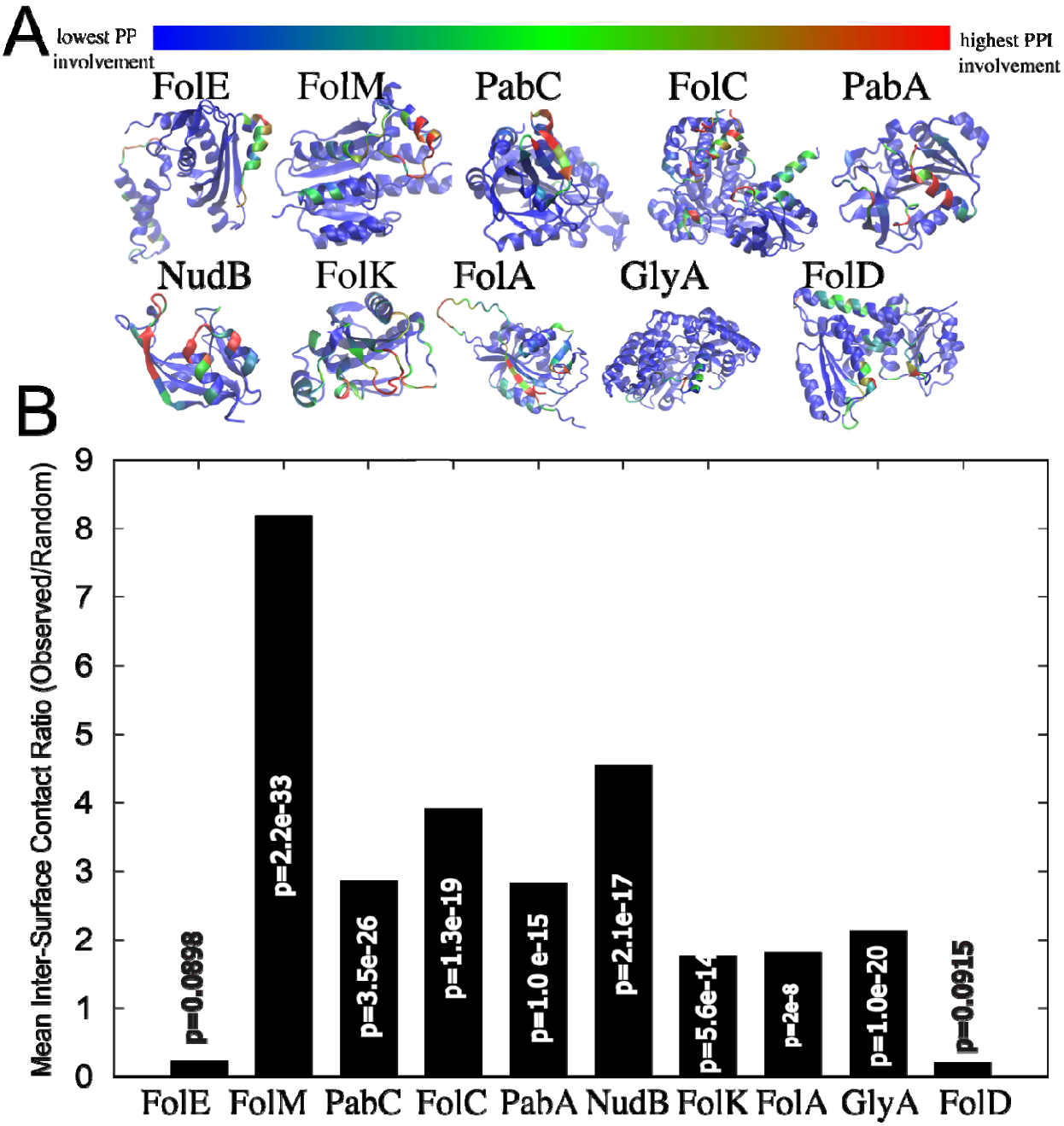

Figure 6. ***Folate pathway proteins use dedicated conserved surfaces to interact with all their metabolon partners.*** *A) Representation of amino acid residues that are most frequently involved in hetero-dimeric protein-protein interactions in the folate pathway. 10 folate pathway enzymes are represented in the figure, and the most frequently observed residues in PPIs are represented by red color. Green colored regions show intermediate involvement while regions shaded blue show negligible involvement in heterodimeric interactions within the folate pathways. B) The ratio of observed inter-surface contacts to inter-surface contacts expected for randomly occurring surfaces. This ratio is*

presented for all 10 folate-pathway proteins. P-values vs random overlap null model (see Methods) is shown for each protein.

High-throughput computational approach reveals that enzymes use similar interfaces to interact with multiple proteins.

Mutagenesis-based library screening experiments so far reveal that the FolA protein interacts with two distinct partners, FolK and MetF with remarkably similar interaction interfaces. However, due to experimental limitations, the interaction interface study could only be performed for a limited number of interaction partners. To study whether multiple folate pathway proteins use similar interface to interact with each of their PPI partners, we employed a high-throughput computational approach wherein we first use AlfaFold2 to predict dimeric structures, for all pairwise interactions between different members of the folate pathway (Fig. 5A). Subsequently, we performed metadynamics simulations to further refine these dimeric structures and their corresponding interaction interfaces (Fig. 5A). To elucidate the most stable PPIs, we identify pairwise inter-chain residue-level contacts for all structures that correspond to free energy minima identified in the metadynamics simulations (Fig. 5A). We also compute the binding free energies (ΔGbound-unbound) from the free energy profiles computed using our metadynamics simulations. In Fig.5B, we present the results for these ΔG values in the form of a computational PPI map of the folate pathway. Interestingly, the folate pathway enzymes show significant propensity to engage in protein-protein interactions, with multiple partners.

In Fig.5C, we then compare the extent of overlap in predicted interaction interfaces of the FolA (from metadynamics simulations) for the FolA-FolK pair and compare it to the experimentally determined interface. As evident from Fig.5A and B, the simulations predictions agree with the experimentally predicted interaction interfaces, suggesting that simulations can be used to screen interaction interfaces in a high-throughput fashion. The overlap between the experimental and simulation predicted surfaces is statistically significant with a p value of $3x10^{-8}$, under null hypothesis that computation and experiment pick two unrelated surfaces on FolA. (see Methods)

We further calculate the frequency of involvement of every amino acid residue in each of the 10 proteins of the folate pathway, in heteromeric PPIs across different partners within the pathway. FolE, FolM, PabC, FolC, FabA interact with 4, 5, 5, 7 and 8 partners out of 10, respectively. On the other hand, NudB, FolK, FolA, GlyA interact with 9, 9, 9, 5 of the 10 possible partners,

respectively. In Fig.6A, we represent the protein structures with residues that are frequently involved in PPI across different partners shaded in red, those with moderate occurrence shaded in green and the residues least involved in PPIs shaded blue. As seen from Fig.6A, the residues most frequently involved in hetero-dimeric protein-protein contacts across all their PPI partners are clustered in a small region of the protein, as evident from large patches of non-interacting residues shaded in blue.

To check if there is a non-random overlap of interaction surfaces across different interaction partners, we employ the statistical analysis (see Methods section for details) which provides p-value of observed overlaps between interaction surfaces with different interaction partners against the null hypothesis that a protein utilizes unrelated surfaces to bind to different interaction partners in the folate metabolon (also see Supplementary Fig. S9). In Fig.6B, we present the results of this analysis in the form of the ratio of mean inter-surface contacts for the observed surfaces (from metadynamics simulations) versus that of control pairs pair of randomly drawn surfaces on the protein. A high ratio in Fig.6B suggests that the observed surfaces share a high degree of overlap compared to what would be expected for randomly chosen surfaces on the protein. As evident from the p-values shown in Fig.6B, we observe a statistically significant overlap in interaction surfaces across different interactions for all proteins of the folate pathway, except FolE and FolD. The corresponding Z-scores are presented in Supplementary Fig.S9. Interestingly, despite being involved in interactions across multiple different partners, the interaction interfaces appear remarkably clustered on one interface of the protein, across several different folate-pathway members. The metadynamics simulations, therefore, suggest that these proteins could exhibit a great deal of promiscuity in PPIs, with the same interface being involved in interactions across multiple different partners.

Discussion

The one-carbon folate biosynthesis pathway comprising three intersecting and inter-dependent pathways (de novo purine, pyrimidine and methionine synthesis) is one of the central metabolic networks of the living cell. It has been long hypothesized that metabolon formation may be essential to this pathway to protect labile folate co-factors that are easily prone to oxidative damage (Stover and Field, 2011). In the absence of reducing agents, folate derivatives were

shown to have a half-life of a few minutes in vitro (Zheng and Cantley, 2019), however they appear much long-lived in *in vivo* studies (Krumdieck et al., 1978; von der Porten et al., 1992). Moreover, the fact that several mammalian proteins in the 1-carbon metabolism pathway had more than one binding sites for folate co-factors indicated that substrate channeling was potentially inherent to this pathway (Schirch and Strong, 1989). In our earlier work, immunoprecipitation followed by mass spectrometry showed extensive albeit weak and transient protein-protein interactions between DHFR and its functional neighbors (Bhattacharyya et al., 2016). In this work, we systematically delineate the extent of protein-protein interactions in the entire *E. coli* 1-carbon metabolism pathway, using a bimolecular fluorescence complementation assay that can capture transient interactions in vivo. We identified two clusters of proteins in the folate and purine biosynthesis pathways respectively that had very strong intra-cluster as well as inter-cluster interactions.

A remarkable finding of our study was that the extent of interaction was the strongest among members of the folate biosynthesis pathway followed by those in the purine biosynthesis pathway. Based on our data, there was no statistically significant interaction among the pyrimidine biosynthesis pathway proteins, though in several cases, they formed strong homomers (along the diagonal in Fig.2A). This trend in interactions aligns quite strikingly with the order of stabilities of the metabolites in the respective pathways as folate metabolites (5,10-methelene THF, 5-methyl THF, N-formyl THF) are considered the most labile (Zheng and Cantley, 2019). On the other hand, pyrimidine pathway intermediates are mostly stable metabolites in the form of mono-phosphates or tri-phosphates (OMP, UMP, UTP, CTP, dTMP, etc.) and presumably do not need metabolon like networks to protect them in the cytosol, while purine pathway intermediates are less stable (GAR, FGAR, AIR, NCAIR, SAICAR), as they are often difficult to quantify from the cell lysate using metabolomics (Bhattacharyya et al., 2021; Madrova et al., 2018). This observation may also hint more generally towards the possible evolution of PPI networks and metabolons in enzymatic networks to protect labile metabolites. Spatial clustering of functionally related enzymes into a dense cluster could also result in significant increase in metabolic flux compared to the bulk reaction scenario, as previously demonstrated by mechanistic models (Ranganathan et al., 2023) (Castellana et al., 2014)

Though much evidence is present in literature about the existence of metabolons in various organisms, very little information is available about structural architecture of these complexes (Bulutoglu et al., 2016; Skalidis et al., 2023; Wu and Minteer, 2015), particularly the nature of these PPI interfaces and their binding affinities. Some interesting questions are: do multiple proteins in a metabolon associate together, or do binary interactions occur at a time? How do these weak interaction interfaces differ from obligate PPI? Are active sites of the enzymes predominantly involved in the interaction interface? A key highlight of the current study is that it provides structural insights into the structural detail of the 1-carbon metabolon using a combination of high-throughput mutagenesis and screening experiments followed by meta-dynamics simulations in combination with AlphaFold2 prediction of initial complexes. Not only did the predicted interfaces agree with experimental findings, but they also shed light on some remarkable and unanswered aspects of the extended folate pathway metabolon. First, it shows that proteins that interact with multiple partners in the pathway almost always use the same interface for interaction with multiple metabolon partners, indicating a high degree of promiscuity at the interaction interface. This also suggests that monomeric enzymes are presumably involved in only binary interactions at a time. However, for oligomers, multiple subunits can associate with different proteins, and hence higher order complexes may form, though we do not have such evidence so far from our present study. Second, our data on DHFR also shows that the interface most used by the enzyme is far away from the active site. Proteins that interact with multiple partners (the so-called hub proteins) can be of two classes: date hubs (singlish hubs) wherein pairwise interactions happen one at a time, or party hubs (multi-hubs), which can engage multiple partners at the same time (Han et al., 2004; Peleg et al., 2014). Our findings suggest that promiscuous hub proteins in the 1-carbon metabolism pathway behave more like date hubs. Whether this is a more general property of metabolons remains to be established, and future work will address this. Nevertheless, one might surmise as to why evolution might favor singlish hubs over multi-hubs, as date hubs ensure that one interaction must have to dissociate for another new interaction to happen, and therefore usage of the same PPI interface might be important for the dynamic assembly and disassembly of interactions that is so fundamental to the concept of metabolons. Shared interfaces might also pose a natural limit on the size of the metabolon, which otherwise can form giant assemblies that completely

sequester proteins. It can also act as a density modulator that can increase metabolic fluxes in long biochemical pathways (Ranganathan et al., 2023).

Our finding about the *E. coli* 1-carbon metabolon is one of the few examples of metabolons in prokaryotic cells. Much work in the recent past have shown that biochemical reactions do not take place in an unorganized manner in the cytosol of bacteria, rather there is a considerable degree of self-organization inside them. Other examples of such supramolecular enzyme clusters in bacteria include the TCA cycle pathway in *Bacillus subtilis* (Meyer et al., 2011) and the glycolysis pathway in *E. coli* (Mowbray and Moses, 1976).

Our previous work on *E. coli* DHFR (Bhattacharyya et al., 2016) suggested that one of the most important fitness consequences of metabolons is that when individual components are over-expressed, it sequesters others in the complex forming near permanent interactions, thereby causing toxicity to the cells. A more detailed study on the propensity of a protein to be involved in PPI and its over-expression toxicity on a much larger dataset of proteins could shed light on sequestration of partner proteins through interaction as a general mechanism of toxicity in metabolic enzymes.

One might argue that our study merely shows interactions among multiple components of the pathway and did not present evidence of substrate channeling. Zhang and Fernie (Zhang and Fernie, 2021) argue that only enzyme assemblies that are shown to both physically and functionally interact (channel metabolites) should be termed metabolons. However, considering that we observe the strongest interaction among enzymes with the most labile and short-lived substrates (folate pathway), and no interaction among the pyrimidine pathway with the most stable intermediates, suggest that substrate channeling must have been an evolutionary driving force to form transient complexes to protect unstable ligands from degradation by limiting their diffusion paths in the cytosol. Future work with more detailed insight into the effect of clustering on metabolic fluxes will decisively answer the outstanding questions concerning functional end evolutionary significance of metabolons.

Methods:

Cloning and transformation of split YFP plasmids

The genes corresponding to 41 proteins in the folate pathway were amplified from the *E. coli* genome using primers containing BamHI and XbaI sites. The genes were cloned in two different formats in two mutually compatible plasmids: as a fusion to the N-terminus of YFP under the control of IPTG inducible tac promoter, and as a fusion to the C-terminus of YFP under control of an arabinose inducible pBAD promoter. The YFP used in this work was the monomeric version of enhanced YFP (EYFP, truncated at 155/156). This is brighter and matures faster than regular YFP, but still requires incubation at 30C for full maturation.

Out of the 41 proteins, 35 proteins were successfully cloned as fusions. NYFP fusion plasmids were transformed into BW27783 cells by TSS method (Chung et al., 1989). This method enables transformation by simple incubation at 4C without any heat shock, and thereby allows high-throughput transformation in multi-well plates. Subsequently, cells transformed with NYFP-fusion plasmids were grown, and were made competent by the TSS method. For NYFP-fusion plasmid, 35 CYFP-fusion plasmids were transformed, resulting in a 35x35 matrix of all possible combinations.

Flow cytometry to detect interaction by split YFP assay

For FACS analysis, BW27783 cells transformed with both NYFP and CYFP fusion proteins were inoculated from glycerol stocks into 200ul of supplemented M9 medium (containing Ampicillin and Chloramphenicol) in deep well plates and grown at 37C overnight. Next day, the cultures were diluted 1:100 in 150ul of supplemented M9 medium containing antibiotics and inducers (0.2% arabinose and 1mM IPTG), grown at 30C for 4-5 hours, following which YFP fluorescence was detected by FACS. For analysis, the BD LSR Fortessa was used along with the HTS (high throughput) module. For each well of the 96 well plate, 30,000 cells were analyzed, and the mean fluorescence intensity was calculated. Those wells for which less than 10,000 cells could be analyzed were discarded.

Library generation

Based on the structure of DHFR (PDB: 7DFR), we selected all surface exposed residues of DHFR (accessibility > 80%), excluding its active site and those contacting active site residues

within 5A radius. There were 85 such positions in DHFR that satisfied the above criteria, and at each of these positions, a VDS library was created. VDS codes for a combination of [ACG][AGT][GC] bases, which in total code for 1530 codons and 1105 amino acids. These codons code for all different amino acid types (hydrophobic, polar, charged), however does not incorporate large hydrophobics (F, W or Y) or proline, and does not incorporate any stop codon. The mutagenesis was carried out on the NYFP-DHFR plasmid in a 96-well PCR plate using a megaprimer based method, following which the reactions from individual wells were combined, digested overnight using DpnI to remove the WT plasmid and transformed into DH5-alpha competent cells. A small amount was plated for single colonies, while the rest of the culture was grown overnight in the presence of antibiotics. The plasmid isolated from single clones were sequenced to confirm presence of mutations. Following this, the library plasmid was isolated from the culture using miniprep.

Sorting of clones

The NYFP-DHFR VDS library plasmid pool was transformed into the background of BW27783 cells containing CYFP-MetF/CYFP-FolK plasmids using a cell: plasmid ratio of 100:1 to minimize chances of multiple plasmids transformed into a single cell. The cultures were grown overnight at 37C, and next day they were diluted 1:100 into fresh supplemented M9 medium containing antibiotics and inducers (0.2% arabinose and 1mM IPTG). Using the BD FACS Aria Cell Sorter, single clones from the top and bottom 5% of the fluorescence intensity distribution of both cultures were sorted into wells of a 96-well plate that contained fresh supplemented M9 medium without any antibiotics. The single clones were allowed to grow for 3 days, following which they were diluted in medium containing antibiotics and allowed to grow overnight. Next day, after growth in the presence of inducers, fluorescence intensity was measured on the Fortessa analyzer, and those that had substantial change in fluorescence compared to the WT culture were sent for sequencing to identify the mutation.

Raman Spectroscopy:

To better understand any plausible interaction between DHFR and HPPK we performed Raman spectroscopy. As a vibrational spectroscopy technique Raman spectroscopy provides a host of structural information and hence has been in use for mechanistically probing conformational changes in proteins (Benevides et al., 2003; Rygula et al., 2013; Tuma, 2005). As compared to

other spectroscopic techniques Raman spectroscopy is essentially a non-destructive technique and is known for its high sensitivity. Raman spectroscopy is an informative technique in probing conformational changes upon binding interactions in protein (Lippert et al., 1976). Conformational changes primarily impact the Amide I range (1600 to 1700 cm^{-1}) which arise due to C=O stretching vibrations of the peptide bonds in turn modulated by the secondary structural elements viz. helices, sheets, turns etc. We probed into the secondary structure of DHFR and HPPK using Raman spectroscopy with an aim to investigate the solution status after mixing. Any interaction upon mixing would essentially give rise to new spectral features which upon careful examination help in mechanistic understanding of conformational changes upon binding. Control experiments were run in presence of ADK (detailed in the results section) and protein concentration in all the acquisition was kept at 15 µM.

The Raman spectroscopy studies utilized a Horiba XploRa confocal Raman microscope, featuring a thermoelectrically cooled detector at -70°C and a 1200 gr/mm grating optimized for 750nm. A 785 nm solid-state laser was employed as the excitation source, with each measurement lasting 180 seconds. To ensure accuracy, four spectra were gathered for each sample, facilitating cosmic ray removal. Spectral smoothing was achieved using the Horiba denoise algorithm, and a polynomial fit was applied for fluorescent baseline correction prior to peak deconvolution in Labspec 6. The 785 nm laser's power was maintained at ~41 mW to avoid photo-bleaching or thermal damage, given the excitation wavelength's significant separation from protein absorption bands. Settings included a 200 µm spectrometer slit, a 500 µm confocal aperture, and calibration against a 520.7 cm^{-1} silicon reference. Experiments used 20 µM protein solutions in sodium phosphate buffer, with a 20µl volume, alongside buffer blanks for thorough comparison. Spectral data were normalized against the distinct 330 cm^{-1} peak.

Metadynamics Simulations.

35 enzymes corresponding to the Folate, Purine and Pyrimidine pathways were used in the computational study. These correspond to the same set of genes listed in the experimental PPI matrix shown in Fig.2. We use AlphaFold 2.0 multimer build to predict the pairwise dimeric structures for all intra-pathway pairs within the 3 pathways. We also randomly sample 55 dimers each corresponding to the Folate-Purine and Purine-Pyrimidine interactions to estimate inter-pathway PPIs.

To account for dynamicity in interactions and compute the stability of the AlphaFold predicted dimers, we employed metadynamics simulations wherein the distance between two chains in the dimer is used as the collective variable of interest (Fig. 5A). We then compute the free energy landscape as a function of these inter-chain distances for every folate dimer under study (Fig.5A). Minima at shorter distances (< interaction radii of the pair) corresponds to a favorable dimer. We then integrated over all bound and unbound structures in the landscape to compute the binding free energy ($\Delta G_{bound\text{-}unbound}$ in Fig.5B). All simulations were performed using NAMD2.13 with the CHARMM36 forcefield. Interaction cutoffs of 12 A and 13.5 A were set for van der Waals and electrostatic interactions. A simulation timestep of 2 fs was used for the study. Meta dynamics simulations were run till the PMFs exhibited convergence.

Statistical test to check for non-random overlap of interaction surfaces.

To check if there is a non-random overlap of interaction surfaces across different interaction partners, we define an order parameter -- inter-surface Cα- Cα contacts as a measure of shared interaction surfaces. To compute this quantity, we define for each folate protein an interaction surface --a list of interacting residues -- with every other protein that it interacts with in the folate pathway. To draw this residue list, we use structures that populate the free energy minima in our metadynamics simulations or revealed by high throughput mutagenesis experiment. For instance, in metadynamics simulations FolA shows stable interactions with 4 other partners. Therefore, folA has 4 potentially distinct interaction surfaces for each of these different partners. We then compute Ca-Ca contacts for the set of FolA residues that define each of the 4 surfaces (across all possible surface pairs) and then compute a quantity -- $MISC^O$ - Mean Inter-surface Contacts derived from actual PPI surfaces obtained from mutational mapping or metadynamics simulations -- for each of the 10 folate proteins.

$$MISC^O = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{C^O\left(S_i, S_j\right)}{n(n-1)/2} \qquad (1)$$

Here superscript O refers to observed (as opposed to random control) surfaces. n is the number of detected interaction surfaces for the protein. S_i and S_j represent the list of residues that make

up the i and j'th surfaces, respectively. The function $C^O\left(S_i, S_j\right)$ represents the normalized number of pairwise C_α- C_α contacts between residues that make up the i and jth surfaces.

$$C^O\left(S_i, S_j\right) = \frac{\sum_{s=1}^{m_i}\sum_{p=1}^{m_j}\Delta\left(\left|\vec{r}_s^{\,i} - \vec{r}_p^{\,j}\right|\right)}{m_i m_j} \qquad (2)$$

Here, m_i is the number of residues belonging to i-th surface, $\left\{\vec{r}_s^{\,i}\right\}$ represents set of coordinates of all C_α atoms belonging to surface i and

$$\Delta\left(\left|\vec{r}_s^{\,i} - \vec{r}_p^{\,j}\right|\right) = \begin{cases} 1 \text{ if } \left|\vec{r}_s^{\,i} - \vec{r}_p^{\,j}\right| \le 7\text{A} \\ 0 \text{ if } \left|\vec{r}_s^{\,i} - \vec{r}_p^{\,j}\right| > 7\text{A} \end{cases}$$

Is a contact function that defines contact between C_α atoms belonging to different surfaces with 7A cutoff.

If the interaction surfaces for different partners exactly overlap, the mean inter-surface contacts (observed) would be high. On the other hand, if these surfaces were distinct and non-overlapping, this quantity would be very small. We compute these mean inter-surface contacts (observed) for each of the 10 enzymes of the folate pathway.

To determine the statistical significance of the overlap between different interacting surfaces we propose a null hypothesis that the interaction surfaces are randomly distributed on the surface of a protein in question. In other words, we test whether the mean inter-surface contacts (observed) are significantly higher than the analogous quantity $MISC^R$ – Mean Inter-surface Contacts (Random) -- for a set of 100 randomly drawn surfaces for every folate protein under study.

$$MISC^R = \sum_{i=1}^{n_R}\sum_{j=i+1}^{n_R}\frac{C^R\left(S_i, S_j\right)}{n_R(n_R-1)/2} \qquad (3)$$

Where $n_R = 50$ the number of randomly generated surfaces The function $C^R\left(S_i, S_j\right)$ represents the normalized number of pair-wise C_α- C_α contacts between the i and jth randomly generated

surfaces, defined analogously to Eq.2 but using C_α coordinates of randomly generated surfaces (see below for the algorithm that generated random surfaces). The mean inter-surface contacts (random) are then compared to the inter-surface contacts observed in experiment and simulations to assess the non-random nature of their overlap. 10,000 such random sets are generated, and the distribution of the inter-surface contacts in the random control is then computed (Supplementary Fig. S10). Our null hypothesis is that there is no significant difference between the observed mean normalized contacts and what would be expected by random chance. In other words, any observed difference is due to random variation per the null hypothesis.

To determine the significance of the observed test statistic, we perform a t-test for statistical significance. Here, the Z score is calculated using the following equation,

$$Z = \frac{MISC^O - MISC^R}{D} \qquad (4)$$

Here, the denominator D refers to the variation in the inter-surface contacts, for the randomly generated surfaces.

$$D = \sigma / \sqrt{n} \qquad (5)$$

Here, σ refers to the standard deviation in the normalized pair-wise contact number (equation 2) for the randomly generated surfaces whereas 'n' refers to the number of randomly generated sets used for the computation of the statistic. The Z score measures how many standard deviations from the mean the observed value lies and determines p-value for the underlying t-statistic for the random null model We set a significance level (e.g., 0.05), which represents the threshold for statistical significance. If the p-value is less than this significance level, we reject the null hypothesis. Otherwise, the null hypothesis is accepted.

Protocol for generating random surfaces. To ensure that the difference in inter-surface contacts for the observed and randomly generated surfaces are not merely due to differences in the surface sizes, we followed the following protocol while generating random surfaces.

i) For each of the 10 folate pathway proteins under study, we compute the radius of gyration of the residues that make up the 'n' interaction surfaces for each protein. The mean radius of gyration for the surface is inferred from these values. This is used as the radius of an idealized spherical interaction surface.

ii) To draw the random surfaces, we define a list of solvent accessible residues for every folate protein based on DSSP calculations. All residues with relative solvent exposed surface area > 0.25 were considered exposed and were considered while defining random surfaces. For every random surface, we pick a residue from this exposed residue list and draw a probe radius (computed in step (i)) around it to define a set of exposed residues that make up the random surface.

iii) To account for the irregularities in protein surfaces, and the accompanying variability in the number of C_α atoms that make up each surface, we employ a normalized contact number to compute statistical differences. The raw contacts are normalized by m_i*m_j, where mi and mj refer to the number of C_α atoms that are part of the ith and j^{th} random surfaces (see Eq. 2)

References:

An, S., Kumar, R., Sheets, E.D., and Benkovic, S.J. (2008). Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. Science *320*, 103-106.

Benevides, J.M., Overman, S.A., and Thomas Jr, G.J. (2003). Raman spectroscopy of proteins. Current protocols in protein science *33*, 17.18. 11-17.18. 35.

Bhattacharyya, S., Bershtein, S., Adkar, B.V., Woodard, J., and Shakhnovich, E.I. (2021). Metabolic response to point mutations reveals principles of modulation of in vivo enzyme activity and phenotype. Mol Syst Biol *17*, e10200.

Bhattacharyya, S., Bershtein, S., Yan, J., Argun, T., Gilson, A.I., Trauger, S.A., and Shakhnovich, E.I. (2016). Transient protein-protein interactions perturb E. coli metabolome and cause gene dosage toxicity. Elife *5*, e20309.

Bulutoglu, B., Garcia, K.E., Wu, F., Minteer, S.D., and Banta, S. (2016). Direct Evidence for Metabolon Formation and Substrate Channeling in Recombinant TCA Cycle Enzymes. ACS Chem Biol *11*, 2847-2853.

Castellana, M., Wilson, M.Z., Xu, Y., Joshi, P., Cristea, I.M., Rabinowitz, J.D., Gitai, Z., and Wingreen, N.S. (2014). Enzyme clustering accelerates processing of intermediates through metabolic channeling. Nat Biotechnol *32*, 1011-1018.

Chung, C.T., Niemela, S.L., and Miller, R.H. (1989). One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. Proc Natl Acad Sci U S A *86*, 2172-2175.

Deng, Y., Gam, J., French, J.B., Zhao, H., An, S., and Benkovic, S.J. (2012). Mapping protein-protein proximity in the purinosome. J Biol Chem *287*, 36201-36207.

French, J.B., Jones, S.A., Deng, H., Pedley, A.M., Kim, D., Chan, C.Y., Hu, H., Pugh, R.J., Zhao, H., Zhang, Y., *et al.* (2016). Spatial colocalization and functional link of purinosomes with mitochondria. Science *351*, 733-737.
Gallagher, W. (2009). FTIR analysis of protein structure. Course manual Chem *455*.
Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P., *et al.* (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature *430*, 88-93.
Kappock, T.J., Ealick, S.E., and Stubbe, J. (2000). Modular evolution of the purine biosynthetic pathway. Curr Opin Chem Biol *4*, 567-572.
Krumdieck, C.L., Fukushima, K., Fukushima, T., Shiota, T., and Butterworth, C.E., Jr. (1978). A long-term study of the excretion of folate and pterins in a human subject after ingestion of 14C folic acid, with observations on the effect of diphenylhydantoin administration. Am J Clin Nutr *31*, 88-93.
Lippert, J., Tyminski, D., and Desmeules, P. (1976). Determination of the secondary structure of proteins by laser Raman spectroscopy. Journal of the American Chemical Society *98*, 7075-7080.
Lukatsky, D.B., Shakhnovich, B.E., Mintseris, J., and Shakhnovich, E.I. (2007). Structural similarity enhances interaction propensity of proteins. J Mol Biol *365*, 1596-1606.
Madrova, L., Krijt, M., Baresova, V., Vaclavik, J., Friedecky, D., Dobesova, D., Souckova, O., Skopova, V., Adam, T., and Zikanova, M. (2018). Mass spectrometric analysis of purine de novo biosynthesis intermediates. PLoS One *13*, e0208947.
Meyer, F.M., Gerwig, J., Hammer, E., Herzberg, C., Commichau, F.M., Volker, U., and Stulke, J. (2011). Physical interactions between tricarboxylic acid cycle enzymes in Bacillus subtilis: evidence for a metabolon. Metab Eng *13*, 18-27.
Mowbray, J., and Moses, V. (1976). The tentative identification in Escherichia coli of a multienzyme complex with glycolytic activity. Eur J Biochem *66*, 25-36.
Peleg, O., Choi, J.M., and Shakhnovich, E.I. (2014). Evolution of specificity in protein-protein interactions. Biophys J *107*, 1686-1696.
Ranganathan, S., Liu, J., and Shakhnovich, E. (2023). Enzymatic metabolons dramatically enhance metabolic fluxes of low-efficiency biochemical reactions. Biophys J *122*, 4555-4566.
Robinson, J.B., Jr., and Srere, P.A. (1985). Organization of Krebs tricarboxylic acid cycle enzymes. Biochem Med *33*, 149-157.
Rygula, A., Majzner, K., Marzec, K.M., Kaczor, A., Pilarczyk, M., and Baranska, M. (2013). Raman spectroscopy of proteins: a review. Journal of Raman Spectroscopy *44*, 1061-1076.
Schirch, V., and Strong, W.B. (1989). Interaction of folylpolyglutamates with enzymes in one-carbon metabolism. Arch Biochem Biophys *269*, 371-380.
Skalidis, I., Kyrilis, F.L., Tuting, C., Hamdi, F., Trager, T.K., Belapure, J., Hause, G., Fratini, M., O'Reilly, F.J., Heilmann, I., *et al.* (2023). Structural analysis of an endogenous 4-megadalton succinyl-CoA-generating metabolon. Commun Biol *6*, 552.
Srere, P.A. (1987). Complexes of sequential metabolic enzymes. Annu Rev Biochem *56*, 89-124.
Stover, P.J., and Field, M.S. (2011). Trafficking of intracellular folates. Adv Nutr *2*, 325-331.
Sweetlove, L.J., and Fernie, A.R. (2018). The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation. Nat Commun *9*, 2136.
Tuma, R. (2005). Raman spectroscopy of proteins: from peptides to large assemblies. Journal of Raman Spectroscopy: An International Journal for Original Work in all Aspects of Raman Spectroscopy, Including Higher Order Processes, and also Brillouin and Rayleigh Scattering *36*, 307-319.

von der Porten, A.E., Gregory, J.F., 3rd, Toth, J.P., Cerda, J.J., Curry, S.H., and Bailey, L.B. (1992). In vivo folate kinetics during chronic supplementation of human subjects with deuterium-labeled folic acid. J Nutr *122*, 1293-1299.

Wu, F., and Minteer, S. (2015). Krebs cycle metabolon: structural evidence of substrate channeling revealed by cross-linking and mass spectrometry. Angew Chem Int Ed Engl *54*, 1851-1854.

Zhang, Y., and Fernie, A.R. (2021). Metabolons, enzyme-enzyme assemblies that mediate substrate channeling, and their roles in plant metabolism. Plant Commun *2*, 100081.

Zhang, Y., Morar, M., and Ealick, S.E. (2008). Structural biology of the purine biosynthetic pathway. Cell Mol Life Sci *65*, 3699-3724.

Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res *33*, 2302-2309.

Zheng, Y., and Cantley, L.C. (2019). Toward a better understanding of folate metabolism in health and disease. J Exp Med *216*, 253-266.