

A cellular selection identifies elongated flavodoxins that support electron transfer to sulfite reductase

Albert Truong^{1,2}, Dru Myerscough², Ian Campbell², Joshua Atkinson²,
and Jonathan J. Silberg^{1,3,4*}

Author affiliations:

1. Biochemistry and Cell Biology Graduate Program, Rice University, 6100 Main Street, MS-180, Houston, Texas 77005
2. Department of Biosciences, Rice University, 6100 Main Street, MS-140, Houston, TX, 77005
3. Department of Bioengineering, Rice University, 6100 Main Street, MS-142, Houston, TX, 77005
4. Department of Chemical and Biomolecular Engineering, Rice University, 6100 Main Street, MS-362, Houston, TX, 77005

*Address correspondence to: Jonathan J. Silberg
Biosciences Department
6100 Main Street
Houston, TX 77005 USA
Tel: 713-348-3849
Email: joff@rice.edu

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](https://doi.org/10.1002/pro.4746). Please cite this article as doi: [10.1002/pro.4746](https://doi.org/10.1002/pro.4746) © 2023 The Protein Society
Received: Mar 25, 2023; Revised: Jul 17, 2023; Accepted: Aug 04, 2023

ABSTRACT

Flavodoxins (Flds) mediate the flux of electrons between oxidoreductases in diverse metabolic pathways. To investigate whether Flds can support electron transfer to a sulfite reductase (SIR) that evolved to couple with a ferredoxin, we evaluated the ability of Flds to transfer electrons from a ferredoxin-NADP reductase (FNR) to a ferredoxin-dependent SIR using growth complementation of an *Escherichia coli* strain with a sulfur metabolism defect. We show that Flds from cyanobacteria complement this growth defect when coexpressed with an FNR and an SIR that evolved to couple with a plant ferredoxin. When we evaluated the effect of peptide insertion on Fld-mediated electron transfer, we observed a sensitivity to insertions within regions predicted to be proximal to the cofactor and partner binding sites, while a high insertion tolerance was detected within loops distal from the cofactor and within regions of helices and sheets that are proximal to those loops. Bioinformatic analysis showed that natural Fld sequence variability predicts a large fraction of the motifs that tolerate insertion of the octapeptide SGRPGSLS. These results represent the first evidence that Flds can support electron transfer to assimilatory SIRs, and they suggest that the pattern of insertion tolerance is influenced by interactions with oxidoreductase partners.

Keywords: cyanobacteria, deep sequencing, electron transfer, ferredoxin, flavodoxin, mutagenesis, protein design, selection, sulfite reductase, synthetic biology

INTRODUCTION

Flavodoxins (Flds) are low potential electron carriers that use a flavin mononucleotide cofactor to transfer electrons between partner oxidoreductases¹. Biochemical and genetic studies have shown that Flds couple with diverse oxidoreductases possessing cellular roles in glycolysis^{2,3}, photosynthesis⁴, hydrogen metabolism⁵, amino acid synthesis^{6,7}, nucleotide metabolism⁸, steroidogenesis^{9,10}, isoprenoid biosynthesis¹¹, redox homeostasis^{12,13}, lipid synthesis¹⁴, nitrogen assimilation^{6,15–17}, and sulfur metabolism^{18,19}. In some organisms, Fld deletions result in growth defects, indicating the essential role of Flds in supporting their metabolism²⁰. In addition, bioinformatic studies have revealed that some organisms have genomes with as many as ten Fld paralogs²¹. Some Flds efficiently deliver electrons to many acceptor proteins, including non-natural partners^{9,10,22–24}. Other Flds appear to have evolved structures that enable discrimination of partner proteins^{10,16,25–28}. Biochemical studies have shown that Flds can elongate a central beta sheet to achieve this specificity²⁹. In addition, bioinformatics has shown that this insertion can be used to differentiate single-domain Flds as either short chain, with an average length of ~145 residues, or long chain, with an average length of ~175 residues²¹. While these Fld types are typically differentiated by insertion within a single loop in the structure, it remains unclear where Fld paralogs can evolve structures through elongation to support coupling between different oxidoreductases.

Many genomes encoding Flds have ferredoxin protein electron carriers²¹, which also function as cellular electron transfer (ET) hubs. The relative abundances of Fld and ferredoxin electron carriers can vary widely. Gammaproteobacteria frequently have

multiple Fld and ferredoxin paralogs, while microbes from other taxonomic groups typically present higher abundances of ferredoxin electron carriers²¹. Both Flds and ferredoxins present low midpoint reduction potentials. Fld redox potentials range from -230 to -530 mV^{30,31} versus a standard hydrogen electrode, while 2Fe-2S and 4Fe-4S ferredoxins present potentials that range from -150 to -500 mV and -200 to -650 mV, respectively³². In many organisms, Fld and ferredoxin expression is controlled by environmental conditions^{33,34}, such as oxidative stress, salt stress, heavy metal toxicity, mineral availability, and light. Also, there is evidence that Fld and ferredoxin expression levels are strongly coupled to iron availability^{1,35,36}, with Fld levels being elevated under low iron availability conditions. This trend is thought to arise as iron limitation decreases the availability of substrates required for the biogenesis of iron-sulfur cluster cofactors on ferredoxins¹. While Flds and ferredoxins support ET to an overlapping set of almost twenty partner proteins, ferredoxins have been shown to couple with a much larger set (>80) of partner oxidoreductases^{1,32}. The extent to which Flds can support ET to many of these ferredoxin-partner proteins in natural or synthetic cellular systems remains unclear.

Cellular assays for ET have been used to probe sequence-structure-function relationships in ferredoxins³². These studies have revealed the utility of coupling the growth of a bacterium to ET mediated by a cytosolic protein electron carrier. This approach has been used to analyze how the sequence of 2Fe-2S ferredoxins from different organisms affects ET between a ferredoxin-NADP reductase (FNR) to a ferredoxin-dependent sulfite reductase (SIR) from plants³⁷. In addition, this assay has been used to explore ideas about ferredoxin evolution through protein engineering. Symmetrical ferredoxins designed computationally have been shown to support ET

through this synthetic pathway³⁸, providing evidence that 4Fe-4S ferredoxins could have evolved through a gene duplication of smaller peptides. Further, this cellular assay has been used for protein engineering to create allosteric electrical switches^{39,40}. Synthetic ferredoxins containing an inserted ligand-binding domain have been discovered that present ET in cells which is activated by endocrine disruptors⁴¹. While these studies have shown the power of using a cellular assay to engineer ferredoxins for bioelectronics⁴², this approach has not yet been applied to Flds.

Flds and ferredoxins both support ET to nitrite reductases⁴³, which share structural similarity with SIRs⁴⁴. This observation led us to investigate whether Flds could support ET to SIR. Through genome mining, a prior study identified cyanobacteria and algae encoding FNRs, ferredoxins, Flds, and SIRs⁴⁵, including *Acaryochloris*, *Anabaena*, *Crocospaera*, *Gloeotheca*, *Nostoc*, *Ostreococcus*, *Prochlorococcus*, *Synechococcus*, *Thalassiosira*, and *Trichodesmium*. This finding suggested that Flds may be capable of supporting ET from FNR to SIR, like ferredoxins. Herein, we show that cyanobacterial Flds support ET between a plant FNR and SIR using a cellular selection (Figure 1a) that requires this ET to complement an *Escherichia coli* growth defect⁴⁶. Since some Fld paralogs evolved elongated structures to support recognition of partner proteins²⁹, which are differentiated by an insertion that splits a central Fld beta sheet, we used this selection to investigate where a cyanobacterial Fld tolerates peptide insertions. We used this type of mutational lesion because we hypothesized that insertion tolerance would indicate the permissiveness of the protein fold to topological alterations or perturbations. By comparing peptide-insertion sensitivity profiles with structural models of Fld and Fld-partner complexes, we show that the pattern of Fld peptide-insertion sensitivity correlates

with proximity to the cofactor binding site and predicted partner interfaces. These studies provide fundamental insight into the ways that Fld evolution could affect oxidoreductase interactions and implicate peptide-insertion profiling as a strategy to assess the quality of predicted oxidoreductase binding interfaces. They also identify new Fld-partner proteins that can be used as living electronic components for bioelectronics and synthetic biology^{41,47}.

RESULTS

Flavodoxins support ET to SIRs. In many microbes, sulfur assimilation from sulfite requires a three-component electron transport chain made up of an electron donor (FNR) that draws reducing equivalents from NADPH, a ferredoxin electron carrier, and a sulfite reductase (SIR) that catalyzes the six-electron reduction of sulfite to sulfide^{48,49}. Because a prior bioinformatic study found that some cyanobacteria code for this pathway and Flds⁴⁵, we sought to investigate whether Flds could support ET from FNR to SIR. To do this, we used an *Escherichia coli* strain (EW11) whose growth has been engineered to be dependent upon ferredoxin-mediated ET from FNR to SIR⁴⁶. This strain represents a simple approach to assess whether Flds can support electron transfer between FNR to SIR that evolved to couple with ferredoxins. Since prior bioinformatics showed that *Synechocystis* sp. PCC6803 and *Nostoc* sp. PCC7120 each encode a single Fld, FNR, and SIR⁴⁵, we first investigated whether Flds from these microbes (sFld1 and nFld, respectively) could support ET in this pathway (Figure 1a). These Flds, which exhibit 67% identity, were expressed using an anhydrotetracycline (aTc) inducible promoter (Figure S1a) within *E. coli* EW11 also expressing *Zea mays* FNR and SIR using constitutive promoters (Figure S1b)³⁹. To identify optimal assay conditions, we evaluated the growth of cells expressing *Mastigocladus laminosus* ferredoxin and *Z. mays* FNR and SIR, which complements *E. coli* EW11 growth³⁹. When using 96-well plates for this cellular assay, we found that it was critical to use high shaking speeds for robust growth complementation (Figure S1c-e). Under these conditions, Flds complemented cell growth after 48 hours (Figure 1b). In non-selective growth medium, Fld expression had no significant effect on cell growth (Figure S2). These results show that Flds can support

cellular ET from an FNR to a SIR in a synthetic pathway. *Synechocystis* Fld was chosen for all subsequent analysis because it presented robust complementation in our assay.

Evaluating Fld mutation tolerance. Structural studies have shown that some Flds have evolved elongated structures to support interactions with partner oxidoreductases²¹. While rational design studies have shown that loop removal can disrupt partner binding²⁹, there have been no studies examining how peptide insertion affects Fld ET. Also, data on the effect of insertions on Fld activity may be useful in guiding future engineering of Fld switches through domain insertion³⁹. To test this idea, we characterized the effects of random peptide insertion on the ability of sFld1 to mediate ET from FNR to SIR. A combinatorial library was built by inserting the octapeptide SGRPGSLS at every backbone location, and the resulting library was selected for Fld-insertion variants that transfer electrons from FNR to SIR using growth complementation of *E. coli* EW11. This peptide was chosen because libraries containing this peptide are easy to generate⁵⁰. To allow comparison of the ET activity of insertion variants with native Fld1, we created a plasmid for expressing sFld1 with synonymous mutations, designated sFld2. *E. coli* EW11 expressing sFld1 and sFld2 presented similar growth complementation (Figure 1b). This observation indicates that the barcoded sFld2 could be embedded in our library selection to calibrate the activities of other Fld mutants.

After building a library of plasmids that express different peptide-insertion variants (Figure S3), we analyzed the library sequence diversity. All of the variants were present in three different sequencing experiments (Figure S4a), with similar average abundances across all variants (Figure 2). To establish which Fld variants support cellular ET, we computed *enrichment values* for each variant using the Enrich2 workflow⁵¹ defined as the

$\log_2(\text{selected:naive ratio})$ relative to the parental sFld2. Analysis of the distribution of enrichment values revealed two major clusters of phenotypes, with some variants of intermediate enrichment (Figure 3b). A fit of this enrichment data to a three-component gaussian mixture model yielded mean values for the largest modes, -5.3 and -0.8, respectively. To determine which variants are non-functional, we evaluated the growth complementation of seventeen peptide-insertion variants (Table S1). As controls, we evaluated cells transformed with a vector that expresses native Fld and a vector that expresses an inactive protein electron carrier. This inactive protein was a ferredoxin mutant (Fd-C42A), which does not complement growth^{39,40}. After 48 hours (Figure 3c), cells expressing individual peptide-insertion variants presented growth that corresponds with fitness scores. These results provide evidence that variants in the peak with low enrichment values are non-functional and that Fld variants with higher fitness scores support cellular ET. We posit that the bimodal trend in growth complementation arises because Fld is not rate limiting for growth in our assay until it loses a large fraction of its ET activity.

We next created a Fld fitness profile where the enrichment scores were scaled from zero to one, where a value of one represents native Fld activity and a value of zero represents variants with fitness that cannot be distinguished from an empty vector (Figure 4A). This profile shows which variants support ET from FNR to SIR. In total, five different Fld motifs tolerated peptide insertion without disrupting Fld-mediated ET from FNR to SIR, including residues spanning from 26 to 31, 39 to 44, 72 to 81, 122 to 138, and 164 to 170. The motif spanning residues 122 to 138 represents the loop that is used to differentiate short- and long-chain Flds²⁹. While most other insertion sites presented a

high sensitivity to peptide insertion, a handful of sites had intermediate fitness values, including those with peptide insertion after residues 3 to 5, 32, 37 to 38, 45-47, 71, 73, 110-112, 121, 150, and 161. These findings show that peptide-insertion sensitivity varies with location in Fld primary structure.

Mapping insertion sensitivity onto Fld structure. To investigate how our peptide insertion profile corresponds with the distribution of insertions observed in natural Flds, we evaluated where insertions appear in Fld homologs relative to each native position within *Synechocystis* Fld. To do this, we performed structural alignments of *Synechocystis* Fld against >2000 structures from three databases, including the PDB, AlphaFoldDB, and ESMfold. A heat map showing where insertions of different sizes are observed in this dataset reveals hot spots that correlate with the results of our selection (Figure 4B). Each of the Fld motifs with high octapeptide insertion tolerance in our selection (26 to 31, 39 to 44, 72 to 81, 122 to 138, and 164 to 170) presented high frequencies of peptide insertions in native Fld homologs. Some other regions, such as residues 85 to 110, display primarily small insertions in homologs, but do not tolerate octapeptide insertion in the selection. A comparison of the enrichments of the sFld insertion variants with log-frequencies of octapeptide insertions in our structural alignment yielded a Pearson correlation of 0.53 (Figure S5). Most sites with a high octapeptide insertion frequency in natural Flds also presented high insertion tolerance in our experiment. However, a subset of the sites with low frequencies of insertions in natural Flds displayed high insertion tolerance. This comparison shows that natural sequence variability predicts some but not all regions of peptide insertion tolerance in a cellular selection experiment.

To better understand the structural and functional features that influenced insertion sensitivity in our selection, we compared our results with a sFld1 structural model from AlphaFoldDB⁵². A structural alignment of this Fld with a homolog having 72% sequence identity, *Synechococcus elongatus* Fld⁵³, yielded a root mean square deviation of 0.27 Å and revealed similar backbone and side chain conformations of residues at the flavin mononucleotide binding site (Figure S6). This finding indicated that the cofactor binding interface of sFld can be inferred from the crystal structure. When we evaluated how Fld structural features relate to insertion sensitivity, we found that high-fitness variants had insertions in backbone locations that are distal from the flavin mononucleotide (Figure 5A). We observed a strong correlation ($r_s = 0.66$, $p < 10^{-4}$) between enrichment and sequence distance to residues that contact the flavin mononucleotide cofactor (Figure 5B). These findings show that Fld is most sensitive to insertions near the flavin mononucleotide binding site.

We next investigated how insertion sensitivity relates to secondary and tertiary structure. We calculated the number of insertion variants within each secondary structure that support ET (fitness > 0) versus do not support ET (fitness = 0). We observed differential insertion tolerance across secondary structure classes (Figure 5C), with similar proportions of functional variants resulting from peptide insertion in loops and helices and significantly fewer in beta sheets ($p = 0.003$), which are largely found within the core of the protein. We found that many of the enriched Fld variants with insertions in helices were near loops, so we also evaluated the enrichment of insertions in helices or sheets that were within two residues of a loop. Among the positions within helices that tolerated insertions, there was a significant enrichment of those proximal to loops ($p =$

0.04). While those positions within sheets that tolerated insertions were all near loops, this association was not statistically significant as there was a low number of enriched insertions in this secondary structure. We next evaluated the relationships between fitness and structural features proximal to the insertion site, including contact densities (Figure 6A-B) and residue depth (Figure 6C). We observed negative correlations between enrichment and intramolecular contact densities, defined as those residue-residue pairs within 8 Å ($r_s = -0.29$, $p = 10^{-4}$) and 14 Å ($r_s = -0.40$, $p < 10^{-4}$). We also observed a negative correlation with residue depth ($r_s = -0.20$, $p = 0.01$). A comparison of fitness with crystallographic B factors (Figure 6D) from the structure of a homologous Fld revealed a positive correlation between enrichment and B factors ($r_s = 0.34$, $p < 10^{-4}$). These findings show that native positions that tolerate peptide insertion tend to be less buried, make fewer residue-residue contacts, and have higher B factors.

To analyze how peptide-insertion sensitivity relates to partner oxidoreductase binding, we modeled the Fld-FNR and Fld-SIR complexes using AlphaFold-multimer^{54,55}. We aligned Fld (PDB ID 1czi) and *Z. mays* FNR (PDB ID 1jb9) crystal structures to the predicted complex to obtain the orientations of the flavin mononucleotide and flavin adenine dinucleotide cofactors. To determine if the predicted conformation is compatible with fast intermolecular ET, we compared the cofactor orientation to that observed in cytochrome P450 reductases (Figure S8), proteins with fused Fld and FNR domains⁵⁶. The predicted Fld-FNR binding mode recapitulated the side-to-side orientation of the cofactors observed in cytochrome P450 reductases with an interatomic distance of 2.4 Å, which is conducive to intermolecular ET. When we analyzed the association between peptide-insertion sensitivity and participation in the predicted Fld-FNR interface within the

modeled complex (Figure 7A), we found that positions making at least one contact with FNR were less likely to be enriched ($p < 10^{-4}$) (Figure 7B). Most Fld variants with insertions at positions making contacts with FNR were non-functional. This finding suggests that the pattern of peptide-insertion sensitivity may contain information about the Fld regions that mediate FNR binding.

A Fld-SIR structural model was also created using AlphaFold-multimer. To assess whether the predicted complex is compatible with efficient intercofactor ET, we aligned Fld (PDB ID 1czi) and *Z. mays* SIR (PDB ID 5h92) crystal structures with our Fld-SIR model (Figure S9). The intercofactor distance for the Fld-SIR complex was 8.3 Å, which is slightly lower than the intercofactor distance (~12 Å) observed in the crystal structure of SIR with its native ferredoxin partner⁴⁸. A comparison of the intercofactor distances and orientations in the Fld-SIR and ferredoxin-SIR complexes is provided in Figure S10. This comparison shows that the Fld-SIR distance is ≤ 9 Å, which is expected to be compatible with intermolecular ET like ferredoxin-SIR. We next analyzed the association between peptide-insertion sensitivity and participation in the Fld-SIR interface in the predicted complex (Figure 7C). As observed with the Fld-FNR complex, positions making at least one contact with FNR were less likely to be enriched ($p = 0.009$) (Figure 7D). Most variants having insertions at Fld locations making contacts with FNR were non-functional. This finding supports the idea that the pattern of insertion sensitivity may contain information about the Fld regions that bind both partner oxidoreductases.

To better understand the relative importance of the observed correlations, logistic regression was used to analyze the correlations between Fld structural features and peptide-insertion tolerance. This analysis revealed that highly enriched insertion mutants

could be distinguished from unenriched mutants based on long-range contact density and sequence distance to contacts with flavin mononucleotide with an area under the receiver operating characteristic of 0.94 (Figure S11). As these features describe general trends based on the wild-type structure, it is likely that other biophysical parameters also influenced the fitness of each mutant. As many of the Fld residues proximal to the flavin mononucleotide also participate in SIR and FNR binding, the correlation with sequence distance to the cofactor may also reflect the importance of FNR and SIR interactions.

DISCUSSION

Our results herein extend the list of Fld-partner oxidoreductases to include SIRs. They also show the utility of synthetic ET pathways for rapidly assessing the coupling of protein electron carriers with different partner oxidoreductases³². The finding that cyanobacterial Flds can mediate ET from plant FNR to SIR implicates a role for this three-component ET pathway in microbes whose genomes encode all three of these oxidoreductases, although further studies will be required to directly demonstrate a role for Flds in mediating ET between FNR and SIR in nature. Prior bioinformatic analysis revealed that a majority of the microbes containing FNR, Flds and ferredoxin-dependent SIR are marine cyanobacteria, such as *Prochlorococcus* and *Synechocystis*⁴⁵. In the future, one way to quickly test ET between these proteins will be to assess whether cyanobacterial FNR, Fld, and SIR from the same organisms can support growth complementation in the cellular assay described herein.

The finding that insertion sensitivity depends on proximity to cofactor and partner binding interfaces illustrates how peptide-insertion tolerance could be used to corroborate predicted binding interfaces generated by models like AlphaFold-multimer^{54,55}. Peptide-insertion sensitivity correlated strongly with two structural parameters, the density of intramolecular contacts and the proximity to cofactor. Prior studies have probed protein structure by inserting peptides with a range of lengths^{57–61}. These studies found that peptide insertion can provide insight into molecular interactions⁵⁷, conditional phenotypes for genetic studies⁵⁸, and sites compatible with the insertion of affinity tags⁵⁹. A major limitation of these prior studies is the use of transposon mutagenesis to generate sequence diversity, which results in non-uniform variant abundances within naive

libraries⁶². Our study leveraged a recently described algorithm for domain insertion to create more uniform libraries⁶³, which enabled us to create a comprehensive fitness map. By scoring the effects of peptide insertion on cellular function, we observed strong correlations between insertion sensitivity and intramolecular and intermolecular interactions. Further studies will be required to establish the mechanisms by which different sequence changes affect function, *e.g.*, direct measurements of folding, cofactor binding, electrochemical midpoint potential, and partner-binding.

Insertions and deletions contribute to protein evolution, but it remains unclear how structural and functional features constrain changes in sequence length⁶⁴. Similar to a prior study that examined protein tolerance to single amino acid insertions⁶⁵, we found that octapeptide insertion tolerance correlates with the secondary structure targeted by insertion, with loop tolerance being greater than that of helices and helices having a greater tolerance than beta strands. Because our analysis was focused on a single protein fold, we cannot conclude whether the observed sensitivity to insertions is intrinsic to beta strands or due to the location of the Fld beta sheets, which are mostly buried. Past studies of Fld folding pathways have shown that the beta sheets are important components of the folding nucleus, which may contribute to their sensitivity to insertions⁶⁶. Our study extends these findings by revealing correlations between insertion tolerance and proximity to cofactor and partner protein binding sites in an oxidoreductase. The former correlation is similar to that reported in a recent study of peptide-insertion tolerance within the membrane protein wire MtrA⁶¹. Interestingly, many of the regions that tolerate insertion in the *Synechocystis* Fld correspond with insertions of varying lengths observed across $>10^3$ Fld homologs. In general, positions that tolerate larger insertions in nature

are a subset of the positions that tolerate smaller insertions. This trend extends to large insertions over 10 residues, which include elements that can present defined tertiary structures (*data not shown*), suggesting that the empirical peptide insertion profile for sFId will guide further engineering efforts. Taken together, our results support the idea that tolerance to insertion is constrained by features of both the protein fold and functional motifs, requiring consideration of non-local intramolecular interactions and intermolecular interactions.

Our results illustrate how insertional mutagenesis can be used to probe how different protein features constrain their evolution through elongation⁶⁷. It is unclear how the identity of the protein homolog targeted for random peptide insertion will affect insertion tolerance, as well as the partner oxidoreductases used to select for coupling. Prior studies varying the stability of the protein homolog targeted for mutagenesis have revealed that sensitivity to mutational lesions decreases as the targeted protein increases in thermostability^{68–70}. Peptide insertion profiles created with multiple homologs will be useful for establishing how topology, stability, and local energetics govern mutational tolerance⁷¹. Also, it will be interesting to use this approach with peptides of varying sequences and lengths to generate larger data sets as a means to develop rules for domain insertion⁷². Emerging machine learning approaches can be applied to these data sets in tandem with metagenomic data to predict design rules for elongated proteins^{73,74}. Such rules will be critical for rationally inserting larger polypeptides as a means to create protein switches for synthetic biology⁷⁵, such as polypeptides encoding ligand-binding domains. The creation of such allosteric switches within protein electron carriers is

needed to dynamically regulate electron flow in living sensors created for metabolic flux control and sensing applications^{39,41}.

METHODS

Materials. Chemicals were from VWR, MilliporeSigma, Fisher, Apex Biosciences, Research Products International, or BD Biosciences. Enzymes and molecular biology kits were from Zymo Research, Qiagen, and New England Biolabs.

Strains. *E. coli* XL1-Blue (Agilent, Inc) and Turbo Competent *E. coli* (New England Biolabs) were used for plasmid construction, *E. coli* 10G (Lucigen) was used for library construction, and *E. coli* EW11 was used for growth complementation⁴⁶. For molecular biology, cells were grown at 37 °C while shaking at 250 rpm in lysogeny broth (LB) pH 7. Cells were transformed using electroporation (1 pulse, ~5 ms, 1.8 kV, 1 mm gap electroporation cuvette). Following transformation, cells were allowed to recover in super optimal broth (SOB) pH 7 for 1 hour at 37 °C, which contained 5 g/L yeast extract, 20 g/L tryptone, 10 mM sodium chloride, 2 mM potassium chloride, and 20 mM magnesium sulfate.

Plasmids. Table S2 lists the plasmids used in this study. *Zea mays* FNR and SIR were constitutively expressed using pSAC01³⁹. The control vectors expressing *Mastigocladus laminosus* ferredoxin (pFd007) and an inactive C42A mutant of this ferredoxin^{39,40}. Plasmids for expressing sFid1 and nFid under control of aTc-inducible promoter (pAG034 and pAG036, respectively) were created by PCR amplifying commercially synthesized genes (Integrated DNA Technologies, Inc.) and cloning them into pFd007³⁹, which contains a ColE1 origin, chloramphenicol resistance gene, and a synthetic translation initiation region. A plasmid for expressing sFid2, which encoded sFid1 with a barcode, was created with synonymous mutations in the codons for residues 30 and 31 (designated pAT001); these codons were mutated from AGTGTG to TCCGTT. To create

plasmids for assessing the growth complementation of individual peptide-insertion mutants, pAG034 was PCR amplified using primers that code for insertion and circularized using Golden Gate⁷⁶. Mutant plasmids were also isolated from the naive peptide-insertion library. All plasmids were sequence verified.

Growth complementation. ET from FNR to SIR was evaluated using growth complementation of *E. coli* EW11 as previously described⁴⁶. To perform growth complementation, cells were transformed with pSAC01³⁹ and plasmids expressing protein electron carriers. Individual colonies were grown in a modified m9 minimal medium (m9c) that includes both sulfur-containing amino acids (80 mg/L each), as well as sodium phosphate heptahydrate, dibasic (6.8 g/L), potassium phosphate, monobasic (3 g/L), sodium chloride (0.5 g/L), 2% glucose, ammonium chloride (1 g/L), calcium chloride (11 mg/L), magnesium sulfate (0.24 g/L), ferric citrate (0.12 g/L), p-aminobenzoic acid (2 mg/L), inositol (20 mg/L), adenine (5 mg/L), uracil (20 mg/L), tryptophan (40 mg/L), tyrosine (1.2 mg/L), and the remaining 16 amino acids (80 mg/L each). To evaluate complementation, cells were transferred to a m9 minimal medium that is selective (m9sa). This medium is identical to m9c other than the lack of sulfur-containing amino acids, with magnesium sulfate (0.24 mg/L) as the only S source.

Library synthesis. The peptide-insertion library was created as described previously⁵⁰. In brief, the sFid1 gene was computationally fragmented into four tiles, and variants of each tile were created that contained an 8-codon insertion between every codon in the tile. Tile sequences, as well as primer sequences for cloning, were generated and synthesized by Twist Biosciences and Integrated DNA Technologies, respectively. Each tile was PCR amplified using Q5 DNA polymerase, the amplicons were gel purified, and

Golden Gate cloning with BsmBI was used to insert each tile into pAG034, a vector with a chloramphenicol resistance marker that expresses sFId1 under the control of an aTc-inducible promoter. Each vector ensemble was transformed into *E. coli* 10G and plated on LB-agar medium containing chloramphenicol (34 ng/mL). After overnight incubation at 37 °C, colony forming units (CFU) were quantified. Each vector ensemble, which contained ~50 variants, yielded ≥3600 CFU. Assuming that ~50% of the synthesized oligo library sequences contain errors and sampling replacement, this colony count indicates that >99% of our variants were sampled at this step⁷⁷. To generate the naive library, colonies were harvested from plates, pooled, and purified using a miniprep kit (Qiagen, Inc.). Library sequence diversity was evaluated using AmpliconEZ sequencing (Genewiz, Inc.).

Library Selection. Electrocompetent *E. coli* EW11 containing pSAC01 were transformed by electroporation with a mixture of the naive library (22.3 fmols) and the barcoded sFId2 vector pAT001 (1.2 fmols). Following transformation, cells were grown in SOB while shaking at 250 rpm for 1 hour at 37 °C. The culture was split into three equal volumes, which were plated on LB-agar medium containing chloramphenicol and streptomycin, 34 ng/mL and 100 ng/mL, respectively. Plates were incubated overnight at 37 °C, and CFU were visually counted (29,500 CFU). To prepare glycerol stocks of resulting transformants for this unselected library, plates were scraped, and a cell pellet was obtained by centrifugation for 5 minutes at 4000g. The cells were washed twice by resuspending the pellet in 50% glycerol followed by centrifugation. Washed cells were resuspended in 50% glycerol, divided into aliquots, and stored at -80 °C. Prior to use, the CFU/mL of the cell library was calculated by plating serial dilutions of library stock on LB-agar plates

containing 34 ng/mL chloramphenicol and 100 ng/mL streptomycin. To select the library for Flds that mediate ET from FNR to SIR, *E. coli* EW11 transformed with the library (5×10^6 CFUs) was used to inoculate three flasks of m9sa medium (30 mL) containing 34 ng/mL chloramphenicol, 100 ng/mL streptomycin, and 100 ng/mL anhydrotetracycline. The number of CFUs used to inoculate the media was calculated by counting colonies on LB-agar medium containing antibiotics following serial dilutions of the inoculum. The m9sa cultures were then incubated at 37°C while shaking at 300 rpm. After 48 hours, the selected cultures grew to 3.12×10^8 , 9.8×10^8 , and 3.6×10^8 CFU/mL. Plasmids were isolated from the inoculum and the m9sa cultures following selection using a miniprep kit (Qiagen, Inc.).

Sequencing. Prior to deep sequencing, each half of the Fld gene in the naive, transformed, and selected libraries was PCR amplified using Q5 DNA polymerase. This PCR amplification added unique sequencing adaptors to the ends of each amplicon (Table S3). To sequence the first half of the Fld gene, adaptors were used that added 42 base pairs prior to the start codon and after base pair 297 in the coding sequence. To sequence the second half, adaptors were added prior to base pair 188 of the coding sequence and after a location 39 base pairs following the stop codon. All of these amplicons were sequenced using the Genewiz AmpliconEZ service. Individual plasmids were sequence verified using Sanger sequencing (Genewiz, Inc.). Raw sequence reads, frequency values, and statistical analysis are provided as Supplementary Data. For each naive library sequencing experiment ($n = 3$), the coefficient of variance (CV) for all variants in a sequencing run was calculated. The CVs for each naive library experiment varied from 0.34 to 0.43. We transformed this library into *E. coli* EW11 with sFld2 and analyzed

the average sequence diversity before and after selection using three different measurements (Figure S4b-c). Following transformation into this strain, a narrow range of CVs (0.41 to 0.48) was observed for the variant frequencies. The average mutant frequencies in the naive and unselected libraries presented a linear correlation with an $R^2 = 0.82$ (Figure S12). With the selected libraries, a larger range of CVs (1.36 to 1.42) was observed within each sequencing run. Taken together, these results show that the sequence diversity in our library was not affected by transformation into *E. coli* EW11, while the selection for growth complementation enriched a subset of variants.

Peptide-insertion profile. The paired reads representing the forward and reverse sequencing data were merged using the BBmerge.sh script⁷⁸. Sequencing reads that were unable to be merged were not used in calculations of enrichment. In cases where >90% of reads were unmergeable, the full data set was not used for enrichment calculations. To identify sequences containing an inserted peptide, we used a previously described Python script (github.com/SavageLab/dipseq) developed for domain insertion analysis⁷⁹. To eliminate sequences that contained out-of-frame insertions and to graph the abundances of each unique in-frame insertion variant, we used a python script (<https://github.com/SilbergLabRice/dipseqplotter>). Variant counts were then used to calculate enrichment scores relative to sFId2 using the Enrich2 method⁵¹. Only the first naïve technical replicate was used for calculating enrichment values, as the other two replicates had much lower sequencing depth, though similar frequencies were observed for each variant across the technical replicates. When mutants had an enrichment score below -3.87, they were designated inactive. This threshold was obtained by fitting a gaussian mixture model to the distribution of log-enrichment scores. All variants <5

standard deviations higher than the mean of the low enrichment peak (mean = -5.3) were designated as non-functional. Using this scoring, fitness scores ranging from 0 to 1 were assigned, where 1 represents the parental Fld enrichment and 0 represents Fld variants that do not present cellular ET.

Structural modeling. A model of the sFld1 (IsiB) structure from AlphaFoldDB (entry P27319) was used for structural analysis of free sFld1⁵². All residues except the first and last two residues had confidence metrics (pLDDTs) above 90, and the lowest pLDDT was >70. These high values led us to use all modeled positions for further analysis. The sFld1 exhibits 72% identity to *Synechococcus elongatus* IsiB, a structurally characterized homolog, protein database (PDB) ID 1czi⁵³, enabling identification of residues at the conserved flavin mononucleotide binding site. AlphaFold-multimer was used to model the Fld-FNR or Fld-SIR complexes^{54,55}. Target sequences were obtained from Uniprot for the sFld1 IsiB (P27319), *Z. mays* chloroplastic SIR (O23813) and *Z. mays* root FNR (Q41736). To mirror the partner-protein sequences used in our cellular assay, the N-terminal ten residues of the FNR were omitted and a glycine residue was added between Met11 and Ser12⁴⁶. Also, the N-terminal chloroplast localization sequence was omitted when modeling SIR. To obtain models for the Fld-FNR and Fld-SIR complexes, we used the ColabFold⁸⁰ (version 1.5.2) implementation of AlphaFold-multimer (version 2.3.0). The multiple sequence alignments of Fld, FNR, and SIR homologs that were required for AlphaFold analysis were obtained automatically from ColabFoldDB⁸⁰ using the mmSEQs2 webserver^{81,82}. One random seed and six recycles were used for the Fld-FNR complex. Sixteen seeds and six recycles were used for the Fld-SIR complex. According to the default behavior of AlphaFold/ColabFold, sequences from the same organism were

paired in the multiple sequence alignment, and unpaired sequences were also included. Templates were included from the PDB70 version 13Mar22^{83,84}. Cofactors were placed in each predicted structure by backbone alignment with structures of Fld PDB ID 1czi⁵³, SIR PDB ID 5h92⁴⁸, and FNR PDB ID 1jb9⁸⁵.

The confidence metrics for each AlphaFold protein complex prediction are provided in Table S4. For the Fld-FNR complex, the predicted binding mode had a predicted template modeling (pTM) score of 0.92 and interface pTM of 0.87. Predicted aligned error (PAE) and predicted local-distance difference test (pLDDT) plots are included in Figure S10. We note that the correspondence in cofactor geometry with cytochrome P450 reductases did not arise from the use of templates, as AlphaFold-multimer does not use template information for predicting relative orientations of separate proteins⁵⁵. The intercofactor distances and orientations of the Fld and FNR in this binding mode are similar to those for a pair of homologues using conventional docking approaches⁸⁶.

For the Fld-SIR complex, a single binding mode was predicted with a pTM score of 0.94 and interface pTM (ipTM) of 0.87. Given the high multiple sequence alignment depth, this uncertainty is thought to arise from the shallow, smooth, and highly-charged interface of the *Z. mays* SIR, which is compatible with multiple binding conformations. Ensemble binding has been observed with the *Z. mays* SIR and its native ferredoxin partner, which binds in at least three conformations that primarily vary by rigid-body rotation along the inter-cofactor axis⁴⁸.

Structural calculations. To quantify intramolecular contact density, residue-residue interactions were identified using distance cutoffs of 8 and 14 Å, which are commonly

used to consider medium-range contacts alone and medium-range and long-range contacts together^{87,88}. Intermolecular contacts were considered separately to identify residues at the predicted interfaces with FNR and SIR. Residue depth was calculated using the MSMS package as the average depth over all atoms in a residue⁸⁹. Local flexibility was inferred from the C α B factors from a related Fld crystal structure⁵³.

Structural alignments. To generate a profile of insertions observed in homologous Flds, we analyzed Fld structures in the PDB and predicted by AlphaFold and ESMfold⁹⁰. As tertiary structure is more conserved than primary structure⁹¹, structural homology searches can be more sensitive than sequence-based searches since they directly compare the proximity of aligned residues in space. We performed a structural homology search for the *Synechocystis* Fld structure from AlphaFoldDB against all structures in the PDB as well as clustered versions of AlphaFoldDB and the Mgnify ESM database using the Foldseek webserver^{90,92,93}. Foldseek encodes protein structure in a sequence representation, which enables very fast alignment using MMseqs, followed by local sequence alignment using the Gotoh-Smith-Waterman algorithm. This approach enabled comparison of large numbers of Fld homologs from structure databases. Targets that were less than 120 residues in length, aligned to less than 120 residues of the *Synechocystis* Fld, or had bit-score probabilities less than 0.99 were omitted. As the PDB contains many redundant structures, e.g., point mutants, only a single structure from each taxonomic identifier was used. PDB structures lacking taxonomic identifiers were also omitted. In total, this yielded a dataset comprising 2417 alignments. This dataset included both short and long flavodoxins (Figure S13A), as well as Flds that are fused to other domains. The insertion length distribution follows an exponential-like or Zipfian

distribution (Figure S13B), as observed in previous bioinformatic studies across protein families^{94–96}. From these alignments, insertions were identified by gaps in the *Synechocystis* Fld sequence, and insertion length was counted as the length of a gap following the last aligned residue. Deletions were not considered. Insertion frequency was calculated by dividing the number of insertions observed at a position in the alignment by the coverage (number of sequences in the alignment) at that position (Figure S13C). Because this approach does not infer phylogenetic relationships, insertions identified represent apparent differences in structure rather than evolutionary events.

Statistics. Growth complementation data, which are plotted as mean and standard deviation of three biological replicates, were analyzed by Dunnett's test versus cells expressing the inactive ferredoxin or by two-tailed Welch's t-test. Library selection data is presented as the mean and standard deviation of three biological (selective library) or technical (naive and non-selected library) replicates. A three-component gaussian mixture model was fit to the enrichment data using the Scikit-learn library version 1.2.2 for Python.⁹⁷ To quantify associations between enrichment classification and secondary structure class and participation in the predicted interfaces with FNR and SIR, we used Fisher's exact test implemented in scipy version 1.10.1. The scipy package was also used to compute Spearman correlations between structural features and enrichment values, and two-sided p values were calculated using the permutation test with 100000 resamples. To analyze the correlations between structural features and peptide insertion tolerance, logistic regression was performed. This analysis used the statsmodels library version 0.13.5 for Python. As many of the structural features (*i.e.*, SASA, residue depth, contact density) were highly correlated, only features with Pearson and Spearman

correlation coefficients less than 0.4 were considered. When several metrics were correlated, the metric with the highest Pearson correlation with enrichment was chosen to be included in the regression analysis. This process of elimination narrowed the features to contact density and sequence distance to flavin mononucleotide. Only insertion variants that were classified as highly enriched (WT-like fitness) or unenriched based on the gaussian mixture model were included in the regression, as those variants with intermediate fitness enrichment were expected to be determined by more subtle biophysical factors. Table S5 shows validation statistics and regression coefficients for the logistic regression model.

SUPPLEMENTAL MATERIAL

Supplemental Figures S1-S13 and Tables S1-S5 are provided as a single PDF file. Also, sequencing counts and frequencies are provided as a single XLSX file.

ACKNOWLEDGEMENTS

We are grateful for support from the Office of Basic Energy Sciences of the U.S. Department of Energy grant DE-SC0014462 (to JJS), Office of Naval Research grant N00014-20-1-2274 (to JJS), NASA Astrobiology Institute grant 80NSSC18M0093 (to JJS, and NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2010604 (to JTA). Research was also sponsored by the Army Research Office and was accomplished under Grant Number W911NF-22-1-0239. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. We also thank Anna Guseva for sharing Fld plasmids. *E. coli* EW11 was a gift from P. Silver (Harvard University).

REFERENCES

1. Pierella Karlusich JJ, Lodeyro AF, Carrillo N (2014) The long goodbye: the rise and fall of flavodoxin during plant evolution. *J. Exp. Bot.* 65:5161–5178.
2. Nakayama T, Yonekura S-I, Yonei S, Zhang-Akiyama Q-M (2013) *Escherichia coli* pyruvate:flavodoxin oxidoreductase, YdbK - regulation of expression and biological roles in protection against oxidative stress. *Genes Genet. Syst.* 88:175–188.
3. Crain AV, Broderick JB (2013) Flavodoxin cofactor binding induces structural changes that are required for protein-protein interactions with NADP(+) oxidoreductase and pyruvate formate-lyase activating enzyme. *Biochim. Biophys. Acta* 1834:2512–2519.
4. Nogués I, Hervás M, Peregrina JR, Navarro JA, de la Rosa MA, Gómez-Moreno C, Medina M (2005) Anabaena flavodoxin as an electron carrier from photosystem I to ferredoxin-NADP⁺ reductase. Role of flavodoxin residues in protein-protein interaction and electron transfer. *Biochemistry* 44:97–104.
5. Gutekunst K, Chen X, Schreiber K, Kaspar U, Makam S, Appel J (2014) The Bidirectional NiFe-hydrogenase in *Synechocystis* sp PCC 6803 Is Reduced by Flavodoxin and Ferredoxin and Is Essential under Mixotrophic, Nitrate-limiting Conditions. *J. Biol. Chem.* 289:1930–1937.
6. Vígara AJ, Inda LA, Vega JM, Gómez-Moreno C, Peleato ML (1998) Flavodoxin as an Electronic Donor in Photosynthetic Inorganic Nitrogen Assimilation by Iron-deficient *Chlorella fusca* Cells. *Photochem. Photobiol.* 67:446–449.
7. Hall DA, Jordan-Starck TC, Loo RO, Ludwig ML, Matthews RG (2000) Interaction of flavodoxin with cobalamin-dependent methionine synthase. *Biochemistry* 39:10711–10719.
8. Lofstad M, Gudim I, Hammerstad M, Røhr ÅK, Hersleth H-P (2016) Activation of the Class Ib Ribonucleotide Reductase by a Flavodoxin Reductase in *Bacillus cereus*. *Biochemistry* 55:4998–5001.
9. Jenkins CM, Waterman MR (1994) Flavodoxin and NADPH-flavodoxin reductase from *Escherichia coli* support bovine cytochrome P450c17 hydroxylase activities. *J. Biol. Chem.* 269:27401–27408.
10. Jenkins CM, Genzor CG, Fillat MF, Waterman MR, Gómez-Moreno C (1997) Negatively charged anabaena flavodoxin residues (Asp144 and Glu145) are important for reconstitution of cytochrome P450 17 α -hydroxylase activity. *J. Biol. Chem.* 272:22509–22513.
11. Puan K-J, Wang H, Dairi T, Kuzuyama T, Morita CT (2005) *fldA* is an essential gene required in the 2-C-methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis. *FEBS Lett.* 579:3802–3806.
12. Nogués I, Martínez-Júlvez M, Navarro JA, Hervás M, Armenteros L, de la Rosa MA, Brodie TB, Hurley JK, Tollin G, Gómez-Moreno C, et al. (2003) Role of hydrophobic interactions in the flavodoxin mediated electron transfer from photosystem I to ferredoxin-NADP⁺ reductase in *Anabaena* PCC 7119. *Biochemistry* 42:2036–2045.
13. St Maurice M, Cremades N, Croxen MA, Sisson G, Sancho J, Hoffman PS (2007) Flavodoxin:quinone reductase (FqrB): a redox partner of pyruvate:ferredoxin oxidoreductase that reversibly couples pyruvate oxidation to NADPH production in *Helicobacter pylori* and *Campylobacter jejuni*. *J. Bacteriol.* 189:4764–4773.

14. Chazarreta-Cifre L, Martiarena L, de Mendoza D, Altabe SG (2011) Role of ferredoxin and flavodoxins in *Bacillus subtilis* fatty acid desaturation. *J. Bacteriol.* 193:4043–4048.
15. Pence N, Tokmina-Lukaszewska M, Yang Z-Y, Ledbetter RN, Seefeldt LC, Bothner B, Peters JW (2017) Unraveling the interactions of the physiological reductant flavodoxin with the different conformations of the Fe protein in the nitrogenase cycle. *J. Biol. Chem.* 292:15661–15669.
16. Gangeswaran R, Eady RR (1996) Flavodoxin 1 of *Azotobacter vinelandii*: characterization and role in electron donation to purified assimilatory nitrate reductase. *Biochem. J.* 317 (Pt 1):103–108.
17. Poudel S, Colman DR, Fixen KR, Ledbetter RN, Zheng Y, Pence N, Seefeldt LC, Peters JW, Harwood CS, Boyd ES (2018) Electron Transfer to Nitrogenase in Different Genomic and Metabolic Backgrounds. *J. Bacteriol.* 200:e00757-17.
18. Drake HL, Akagi JM (1977) Bisulfite reductase of *Desulfovibrio vulgaris*: explanation for product formation. *J. Bacteriol.* 132:139–143.
19. Birch OM, Hewitson KS, Fuhrmann M, Burgdorf K, Baldwin JE, Roach PL, Shaw NM (2000) MioC is an FMN-binding protein that is essential for *Escherichia coli* biotin synthase activity in vitro. *J. Biol. Chem.* 275:32277–32280.
20. Bluford J, Windham E, Truong A, Silberg JJ. Cellular strategies to study and engineer low-potential protein electron carriers. In: Furst A, editor. *Electron Transfer in Biomacromolecules*. De Gruyter; 2023. Available from: <https://www.degruyter.com/document/isbn/9781501516184/html>
21. Campbell IJ, Bennet GN, Silberg JJ (2019) Evolutionary Relationships Between Low Potential Ferredoxin and Flavodoxin Electron Carriers | Energy Research. *Front. Energy Res.* 7:1–18.
22. Mayhew SG, Massey V (1969) Purification and Characterization of Flavodoxin from *Peptostreptococcus elsdenii*. *J. Biol. Chem.* 244:794–802.
23. Nogués I, Tejero J, Hurley JK, Paladini D, Frago S, Tollin G, Mayhew SG, Gómez-Moreno C, Ceccarelli EA, Carrillo N, et al. (2004) Role of the C-terminal tyrosine of ferredoxin-nicotinamide adenine dinucleotide phosphate reductase in the electron transfer processes with its protein partners ferredoxin and flavodoxin. *Biochemistry* 43:6127–6137.
24. Tognetti VB, Palatnik JF, Fillat MF, Melzer M, Hajirezaei M-R, Valle EM, Carrillo N (2006) Functional replacement of ferredoxin by a cyanobacterial flavodoxin in tobacco confers broad-range stress tolerance. *Plant Cell* 18:2035–2050.
25. Wan JT, Jarrett JT (2002) Electron acceptor specificity of ferredoxin (flavodoxin):NADP⁺ oxidoreductase from *Escherichia coli*. *Arch. Biochem. Biophys.* 406:116–126.
26. Lawson RJ, von Wachenfeldt C, Haq I, Perkins J, Munro AW (2004) Expression and characterization of the two flavodoxin proteins of *Bacillus subtilis*, YkuN and YkuP: biophysical properties and interactions with cytochrome P450 *Biol. Biochemistry* 43:12390–12409.
27. Wang Z-Q, Lawson RJ, Buddha MR, Wei C-C, Crane BR, Munro AW, Stuehr DJ (2007) Bacterial flavodoxins support nitric oxide production by *Bacillus subtilis* nitric-oxide synthase. *J. Biol. Chem.* 282:2196–2202.

28. Boynton TO, Daugherty LE, Dailey TA, Dailey HA (2009) Identification of *Escherichia coli* HemG as a novel, menadione-dependent flavodoxin with protoporphyrinogen oxidase activity. *Biochemistry* 48:6705–6711.
29. López-Llano J, Maldonado S, Bueno M, Lostao A, Ángeles-Jiménez M, Lillo MP, Sancho J (2004) The Long and Short Flavodoxins: I. THE ROLE OF THE DIFFERENTIATING LOOP IN APOFLAVODOXIN STRUCTURE AND FMN BINDING *. *J. Biol. Chem.* 279:47177–47183.
30. Deistung J, Thorneley RN (1986) Electron transfer to nitrogenase. Characterization of flavodoxin from *Azotobacter chroococcum* and comparison of its redox potentials with those of flavodoxins from *Azotobacter vinelandii* and *Klebsiella pneumoniae* (nifF-gene product). *Biochem. J.* 239:69–75.
31. Kimmich N, Das A, Sevrioukova I, Mehareenna Y, Sligar SG, Poulos TL (2007) Electron transfer between cytochrome P450cin and its FMN-containing redox partner, cindoxin. *J. Biol. Chem.* 282:27006–27011.
32. Atkinson JT, Campbell I, Bennett GN, Silberg JJ (2016) Cellular Assays for Ferredoxins: A Strategy for Understanding Electron Flow through Protein Carriers That Link Metabolic Pathways. *Biochemistry* 55:7047–7064.
33. Mazouni K, Domain F, Chauvat F, Cassier-Chauvat C (2003) Expression and regulation of the crucial plant-like ferredoxin of cyanobacteria. *Mol. Microbiol.* 49:1019–1029.
34. Cassier-Chauvat C, Chauvat F (2014) Function and Regulation of Ferredoxins in the Cyanobacterium, *Synechocystis* PCC6803: Recent Advances. *Life Basel Switz.* 4:666–680.
35. Singh AK, McIntyre LM, Sherman LA (2003) Microarray analysis of the genome-wide response to iron deficiency and iron reconstitution in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.* 132:1825–1839.
36. Bovy A, de Vrieze G, Lugones L, van Horssen P, van den Berg C, Borrias M, Weisbeek P (1993) Iron-dependent stability of the ferredoxin I transcripts from the cyanobacterial strains *Synechococcus* species PCC 7942 and *Anabaena* species PCC 7937. *Mol. Microbiol.* 7:429–439.
37. Campbell IJ, Olmos JL, Xu W, Kahanda D, Atkinson JT, Sparks ON, Miller MD, Phillips GN, Bennett GN, Silberg JJ (2020) *Prochlorococcus* phage ferredoxin: structural characterization and electron transfer to cyanobacterial sulfite reductases. *J. Biol. Chem.* 295:10610–10623.
38. Mutter AC, Tyryshkin AM, Campbell IJ, Poudel S, Bennett GN, Silberg JJ, Nanda V, Falkowski PG (2019) De novo design of symmetric ferredoxins that shuttle electrons in vivo. *Proc. Natl. Acad. Sci.* 116:14557–14562.
39. Atkinson JT, Campbell IJ, Thomas EE, Bonitatibus SC, Elliott SJ, Bennett GN, Silberg JJ (2019) Metalloprotein switches that display chemical-dependent electron transfer in cells. *Nat. Chem. Biol.* 15:189.
40. Wu B, Atkinson JT, Kahanda D, Bennett GN, Silberg JJ (2020) Combinatorial design of chemical-dependent protein switches for controlling intracellular electron transfer. *AIChE J.* 66:e16796.
41. Atkinson JT, Su L, Zhang X, Bennett GN, Silberg JJ, Ajo-Franklin CM (2022) Real-time bioelectronic sensing of environmental contaminants. *Nature* 611:548–553.

42. Tseng C-P, Silberg JJ, Bennett GN, Verduzco R (2020) 100th Anniversary of Macromolecular Science Viewpoint: Soft Materials for Microbial Bioelectronics. *ACS Macro Lett.* 9:1590–1603.
43. Arizmendi JM, Serra JL (1990) Purification and some properties of the nitrite reductase from the cyanobacterium *Phormidium laminosum*. *Biochim. Biophys. Acta* 1040:237–244.
44. Swamy U, Wang M, Tripathy JN, Kim S-K, Hirasawa M, Knaff DB, Allen JP (2005) Structure of spinach nitrite reductase: implications for multi-electron reactions by the iron-sulfur:siroheme cofactor. *Biochemistry* 44:16054–16063.
45. Guseva A (2020) Flavodoxin protein electron carriers: bioinformatic analysis and interactions with sulfite reductases. Available from: <https://scholarship.rice.edu/handle/1911/109196>
46. Barstow B, Agapakis CM, Boyle PM, Grandl G, Silver PA, Wintermute EH (2011) A synthetic system links FeFe-hydrogenases to essential *E. coli* sulfur metabolism. *J. Biol. Eng.* 5:7.
47. Atkinson JT, Chavez MS, Niman CM, El-Naggar MY (2023) Living electronics: A catalogue of engineered living electronic components. *Microb. Biotechnol.* 16:507–533.
48. Kim JY, Nakayama M, Toyota H, Kurisu G, Hase T (2016) Structural and mutational studies of an electron transfer complex of maize sulfite reductase and ferredoxin. *J. Biochem. (Tokyo)* 160:101–109.
49. Shinohara F, Kurisu G, Hanke G, Bowsher C, Hase T, Kimata-Arigo Y (2017) Structural basis for the isotype-specific interactions of ferredoxin and ferredoxin: NADP⁺ oxidoreductase: an evolutionary switch between photosynthetic and heterotrophic assimilation. *Photosynth. Res.* 134:281–289.
50. Coyote-Maestas W, Nedrud D, Okorafor S, He Y, Schmidt D (2020) Targeted insertional mutagenesis libraries for deep domain insertion profiling. *Nucleic Acids Res.* 48:e11–e11.
51. Rubin AF, Gelman H, Lucas N, Bajjalieh SM, Papenfuss AT, Speed TP, Fowler DM (2017) A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* 18:150.
52. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50:D439–D444.
53. Hoover DM, Drennan CL, Metzger AL, Osborne C, Weber CH, Patridge KA, Ludwig ML (1999) Comparisons of wild-type and mutant flavodoxins from *Anacystis nidulans*. Structural determinants of the redox potentials. *J. Mol. Biol.* 294:725–743.
54. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589.
55. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Židek A, Bates R, Blackwell S, Yim J, et al. (2021) Protein complex prediction with AlphaFold-Multimer. :2021.10.04.463034. Available from: <https://www.biorxiv.org/content/10.1101/2021.10.04.463034v1>
56. Wang M, Roberts DL, Paschke R, Shea TM, Masters BS, Kim JJ (1997) Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. *Proc. Natl. Acad. Sci. U. S. A.* 94:8411–8416.

57. Cao Y, Hallet B, Sherratt DJ, Hayes F (1997) Structure-function correlations in the XerD site-specific recombinase revealed by pentapeptide scanning mutagenesis. *J. Mol. Biol.* 274:39–53.
58. Poussu E, Vihinen M, Paulin L, Savilahti H (2004) Probing the alpha-complementing domain of *E. coli* beta-galactosidase with use of an insertional pentapeptide mutagenesis strategy based on Mu in vitro DNA transposition. *Proteins* 54:681–692.
59. Hoeller BM, Reiter B, Abad S, Graze I, Glieder A (2008) Random tag insertions by Transposon Integration mediated Mutagenesis (TIM). *J. Microbiol. Methods* 75:251–257.
60. Pajunen M, Poussu E, Turakainen H, Savilahti H (2009) Application of Mu in vitro transposition for high-precision mapping of protein-protein interfaces on a yeast two-hybrid platform. *Methods San Diego Calif* 49:255–262.
61. Campbell IJ, Atkinson JT, Carpenter MD, Myerscough D, Su L, Ajo-Franklin CM, Silberg JJ (2022) Determinants of Multiheme Cytochrome Extracellular Electron Transfer Uncovered by Systematic Peptide Insertion. *Biochemistry* 61:1337–1350.
62. Atkinson JT, Jones AM, Zhou Q, Silberg JJ (2018) Circular permutation profiling by deep sequencing libraries created using transposon mutagenesis. *Nucleic Acids Res.* 46:e76.
63. Coyote-Maestas W, He Y, Myers CL, Schmidt D (2019) Domain insertion permissibility-guided engineering of allostery in ion channels. *Nat. Commun.* 10:290.
64. Miton CM, Tokuriki N (2023) Insertions and Deletions (Indels): A Missing Piece of the Protein Engineering Jigsaw. *Biochemistry* 62:148–157.
65. Gonzalez CE, Roberts P, Ostermeier M (2019) Fitness Effects of Single Amino Acid Insertions and Deletions in TEM-1 β -Lactamase. *J. Mol. Biol.* 431:2320–2330.
66. Steensma E, van Mierlo CP (1998) Structural characterisation of apoflavodoxin shows that the location of the stable nucleus differs among proteins with a flavodoxin-like topology. *J. Mol. Biol.* 282:653–666.
67. Netzer WJ, Hartl FU (1997) Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388:343–349.
68. Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, Arnold FH (2005) Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. U. S. A.* 102:606–611.
69. Li Y, Drummond DA, Sawayama AM, Snow CD, Bloom JD, Arnold FH (2007) A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat. Biotechnol.* 25:1051–1056.
70. Segall-Shapiro TH, Nguyen PQ, Dos Santos ED, Subedi S, Judd J, Suh J, Silberg JJ (2011) Mesophilic and hyperthermophilic adenylate kinases differ in their tolerance to random fragmentation. *J. Mol. Biol.* 406:135–148.
71. Atkinson JT, Jones AM, Nanda V, Silberg JJ (2019) Protein tolerance to random circular permutation correlates with thermostability and local energetics of residue-residue contacts. *Protein Eng. Des. Sel. PEDS* 32:489–501.

72. Ostermeier M (2005) Engineering allosteric protein switches by domain insertion. *Protein Eng. Des. Sel. PEDS* 18:359–364.
73. Coyote-Maestas W, Nedrud D, Suma A, He Y, Matreyek KA, Fowler DM, Carnevale V, Myers CL, Schmidt D (2021) Probing ion channel functional architecture and domain recombination compatibility by massively parallel domain insertion profiling. *Nat. Commun.* 12:7114.
74. Fan X, Pan H, Tian A, Chung WK, Shen Y (2023) SHINE: protein language model-based pathogenicity prediction for short inframe insertion and deletion variants. *Brief. Bioinform.* 24:bbac584.
75. Del Valle I, Fulk EM, Kalvapalle P, Silberg JJ, Masiello CA, Stadler LB (2020) Translating New Synthetic Biology Advances for Biosensing Into the Earth and Environmental Sciences. *Front. Microbiol.* 11:618373.
76. Engler C, Kandzia R, Marillonnet S (2008) A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE* 3:e3647.
77. Bosley AD, Ostermeier M (2005) Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomol. Eng.* 22:57–61.
78. Bushnell B, Rood J, Singer E (2017) BBMerge - Accurate paired shotgun read merging via overlap. *PloS One* 12:e0185056.
79. Oakes BL, Nadler DC, Flamholz A, Fellmann C, Staahl BT, Doudna JA, Savage DF (2016) Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* 34:646–651.
80. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M (2022) ColabFold: making protein folding accessible to all. *Nat. Methods* 19:679–682.
81. Steinegger M, Söding J (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35:1026–1028.
82. Mirdita M, Steinegger M, Söding J (2019) MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinforma. Oxf. Engl.* 35:2856–2858.
83. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J (2019) HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* 20:473.
84. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
85. Aliverti A, Faber R, Finnerty CM, Ferioli C, Pandini V, Negri A, Karplus PA, Zanetti G (2001) Biochemical and crystallographic characterization of ferredoxin-NADP(+) reductase from nonphotosynthetic tissues. *Biochemistry* 40:14501–14508.
86. Medina M, Abagyan R, Gómez-Moreno C, Fernandez-Recio J (2008) Docking analysis of transient complexes: Interaction of ferredoxin-NADP+ reductase with ferredoxin and flavodoxin. *Proteins Struct. Funct. Bioinforma.* 72:848–862.
87. Kozma D, Simon I, Tusnády GE (2012) CMWeb: an interactive on-line tool for analysing residue-residue contacts and contact prediction methods. *Nucleic Acids Res.* 40:W329–333.

88. Nagarajan R, Archana A, Thangakani AM, Jemimah S, Velmurugan D, Gromiha MM (2016) PDBparam: Online Resource for Computing Structural Parameters of Proteins. *Bioinforma. Biol. Insights* 10:73–80.
89. Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 38:305–320.
90. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Costa A dos S, Fazel-Zarandi M, Sercu T, Candido S, et al. (2022) Language models of protein sequences at the scale of evolution enable accurate structure prediction. :2022.07.20.500902. Available from: <https://www.biorxiv.org/content/10.1101/2022.07.20.500902v1>
91. Illergård K, Ardell DH, Elofsson A (2009) Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. *Proteins* 77:499–508.
92. Richardson L, Allen B, Baldi G, Beracochea M, Bileschi ML, Burdett T, Burgin J, Caballero-Pérez J, Cochrane G, Colwell LJ, et al. (2023) MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Res.* 51:D753–D759.
93. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M (2023) Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.*:1–4.
94. Qian B, Goldstein RA (2001) Distribution of indel lengths. *Proteins Struct. Funct. Bioinforma.* 45:102–104.
95. Chang MSS, Benner SA (2004) Empirical analysis of protein insertions and deletions determining parameters for the correct placement of gaps in protein sequence alignments. *J. Mol. Biol.* 341:617–631.
96. Pang A, Smith AD, Nuin PA, Tillier ER (2005) SIMPROT: Using an empirically determined indel distribution in simulations of protein evolution. *BMC Bioinformatics* 6:236.
97. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. (2011) Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12:2825–2830.

FIGURE LEGENDS

Figure 1. Flds transfer electrons from FNR to SIR. (A) The three-component ET pathway tested for growth complementation in *E. coli* EW11. (B) *Nostoc* PCC7120 Fld (nFld) and *Synechocystis* PCC6803 Fld (sFld1) both complement the growth of *E. coli* EW11 after 48 hours at 37°C when the inducer anhydrotetracycline (aTc) is added to growth medium having sulfate as the only sulfur source. As a positive control, we evaluated complementation when cells expressed *Mastigocladus laminosus* ferredoxin (Fd), which has been shown to couple with FNR and SIR. As a negative control, we evaluated complementation using a C42A mutant of this ferredoxin, which cannot coordinate an iron-sulfur cluster. Cells expressing Fd, nFld, sFld1, and a Fld from *Synechocystis* PCC6803 containing silent mutations in codons 30 and 31 (sFld2) all present significantly higher growth than this control ($p < 0.01$, Dunnett's test). Error bars represent the standard deviation from three biological replicates. *** $p < 0.001$, ** $0.001 < p < 0.01$.

Figure 2. Relative abundances of each peptide-insertion variant. (A) A vector library was built to have codons encoding an octapeptide inserted after each codon. Deep sequencing was used to quantify the relative abundances of each Fld variant in the (B) naive, (C) transformed, and (D) selected libraries. For the naive library, the bars represent the mean counts of each insertion variant from three technical replicates, while the selected library represents mean counts from three biological replicates.

Figure 3. Peptide-insertion profile calculated from sequencing data. (A) The relative enrichment of each insertion variant is mapped onto primary structure and compared with predicted secondary structure (top), contacts with flavin mononucleotide (FMN), and

intermolecular contacts made with the donor (FNR) and acceptor (SIR) partners. Pairs of residues were considered contacting if any constituent heavy atoms were within 8 Å. **(B)** The distribution of enrichment values fitted to a gaussian mixture model. Mutants with enrichments within five standard deviations of the low enrichment peak (blue) were designated depleted, *i.e.*, those with scores less than -3.87, while all others having greater values were designated as enriched. **(C)** Comparison of growth complementation of individual Fld variants after 48 hours with calculated enrichment values. OD bars represent standard deviation from three biological replicates, while enrichment error bars represent two standard errors calculated using Enrich2. Dashed lines indicate the enrichment value and growth complementation of the sFld2 positive control.

Figure 4. Insertion variant fitness comparison. **(A)** A fitness profile that scales the cellular activity of each insertion variant to that of native Fld, which we designate as having a fitness value of 1. The profile is compared with secondary structure (*top*), contacts with the flavin mononucleotide (FMN), and intermolecular contacts made with the donor (FNR) and acceptor (SIR) partners. **(B)** Insertions observed in homologous Flds relative to *Synechocystis* Fld. The frequencies of insertions having lengths of one to ten amino acids and greater than 10 amino acids are shown.

Figure 5. Insertion tolerance correlates with loop proximity and cofactor distance. **(A)** The *Synechocystis* Fld structure predicted by AlphaFold is colored by enrichment values obtained from the selection. The flavin mononucleotide cofactor was placed by backbone alignment with the crystal structure 1czi. **(B)** The relative enrichments of each variant plotted against the minimum sequence distance from the insertion to a Fld residue that interacts with the flavin mononucleotide. **(C)** The relative abundances of enriched

and depleted variants sorted by secondary structure location. Diagonal lines indicate variants having a peptide inserted within two residues of a loop.

Figure 6. Residues with more intramolecular contacts or greater burial generally have lower insertion tolerance. Relationships between insertion tolerance and residue-residue contact densities at **(A)** 8 Å and **(B)** 14 Å, as well as **(C)** residue depth and **(D)** B factors from the crystal structure of *S. elongatus* Fld (PDB ID 1czi).

Figure 7. Positions at the predicted interfaces with partner oxidoreductases do not tolerate insertions. **(A)** Binding mode predicted for the Fld-FNR complex. This binding mode recapitulates the interflavin geometry observed among cytochrome reductases. **(B)** Enrichment values for Fld insertion variants plotted against numbers of intermolecular residue-residue contacts with FNR at 8 Å. **(C)** The predicted binding mode for the Fld-SIR complex. **(D)** Enrichment values for Fld insertion variants plotted against numbers of intermolecular residue-residue contacts with SIR at 8 Å. For both complexes, contacts were counted for the residue preceding the backbone insertion site.

Figure 1

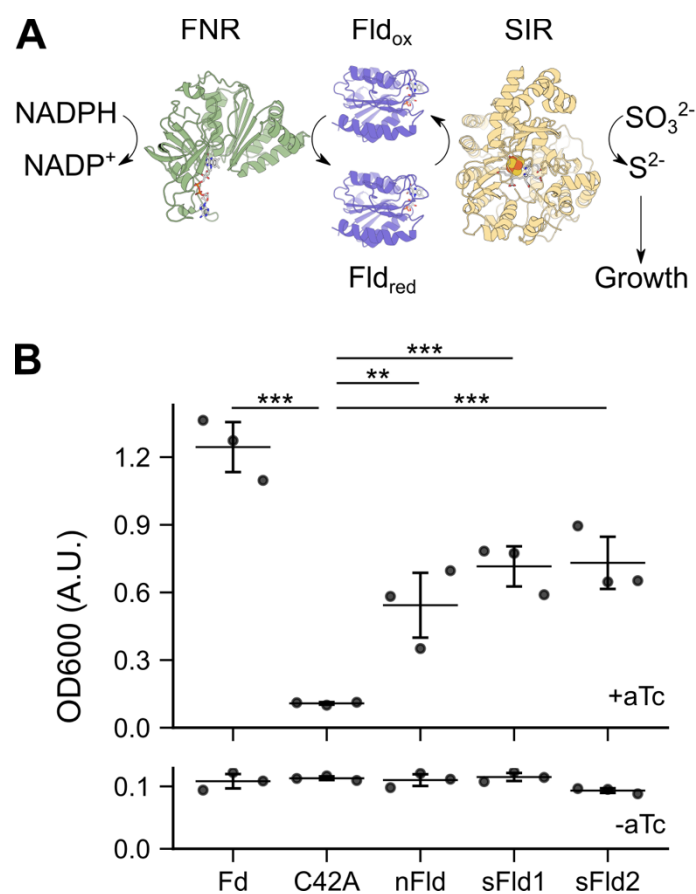


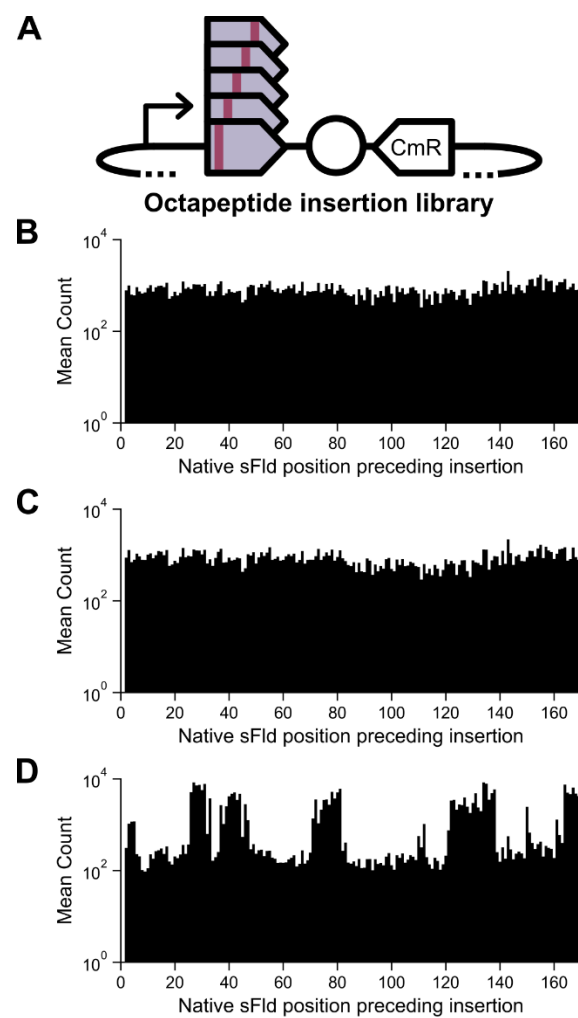
Figure 2

Figure 3

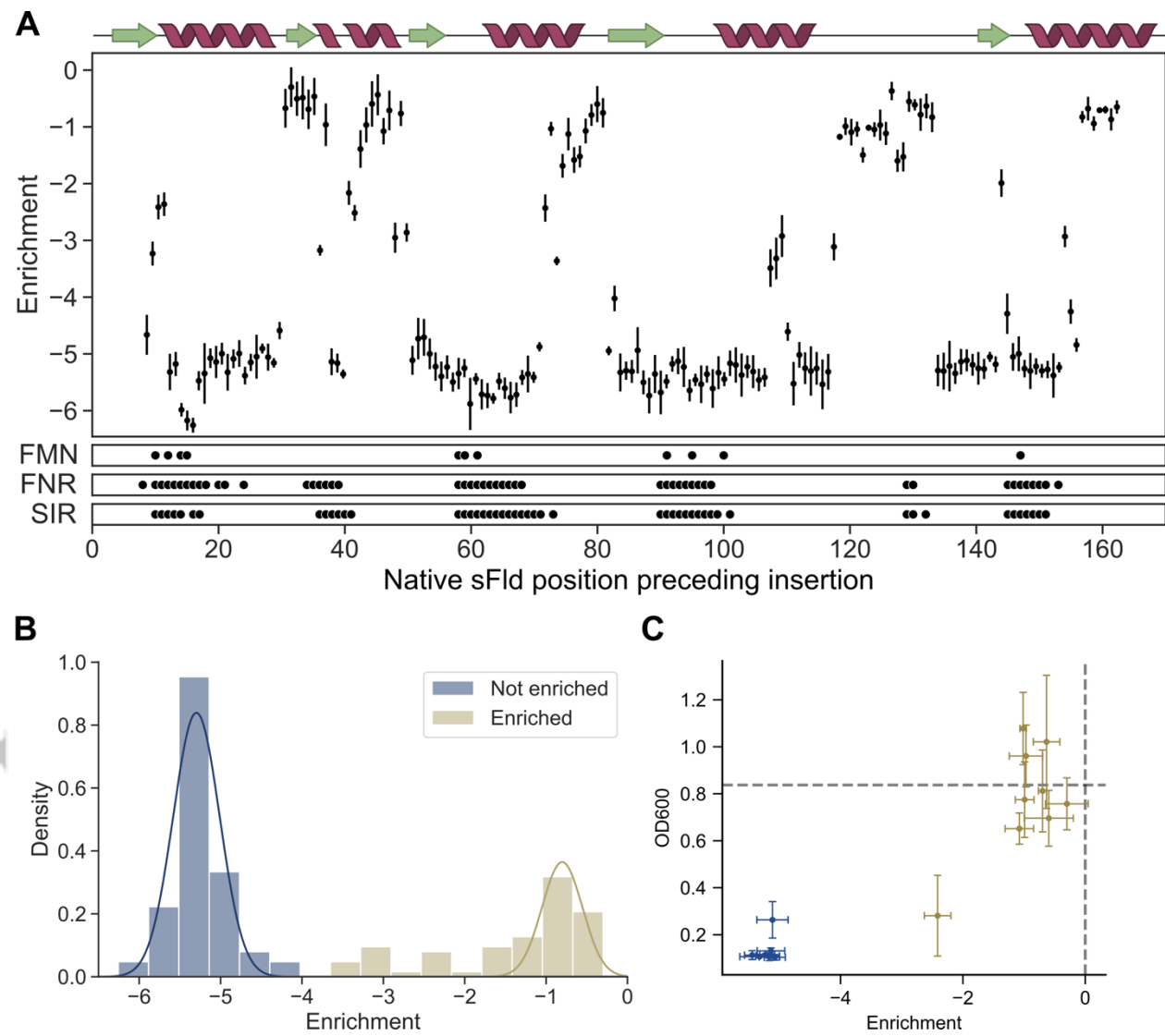


Figure 4

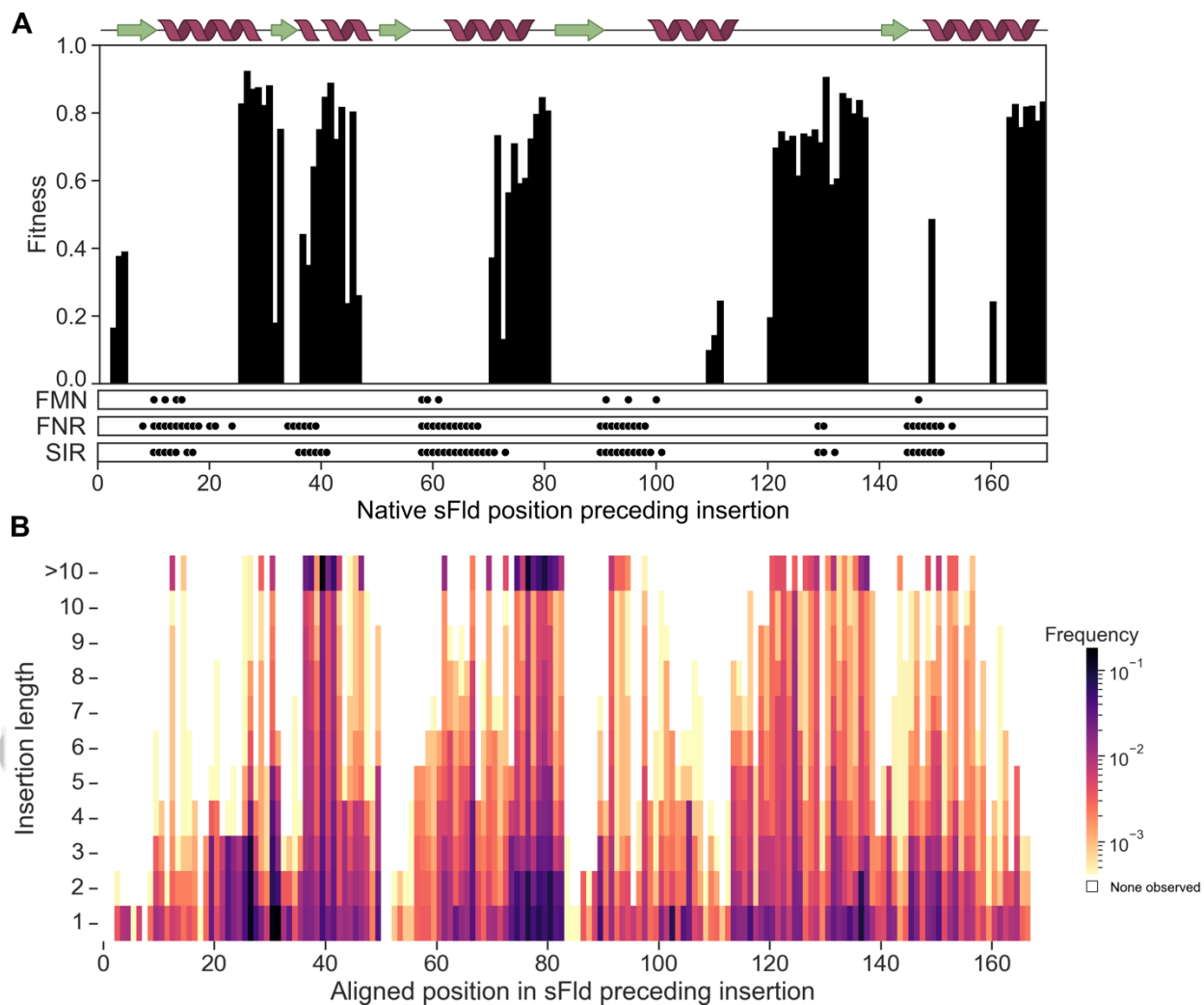


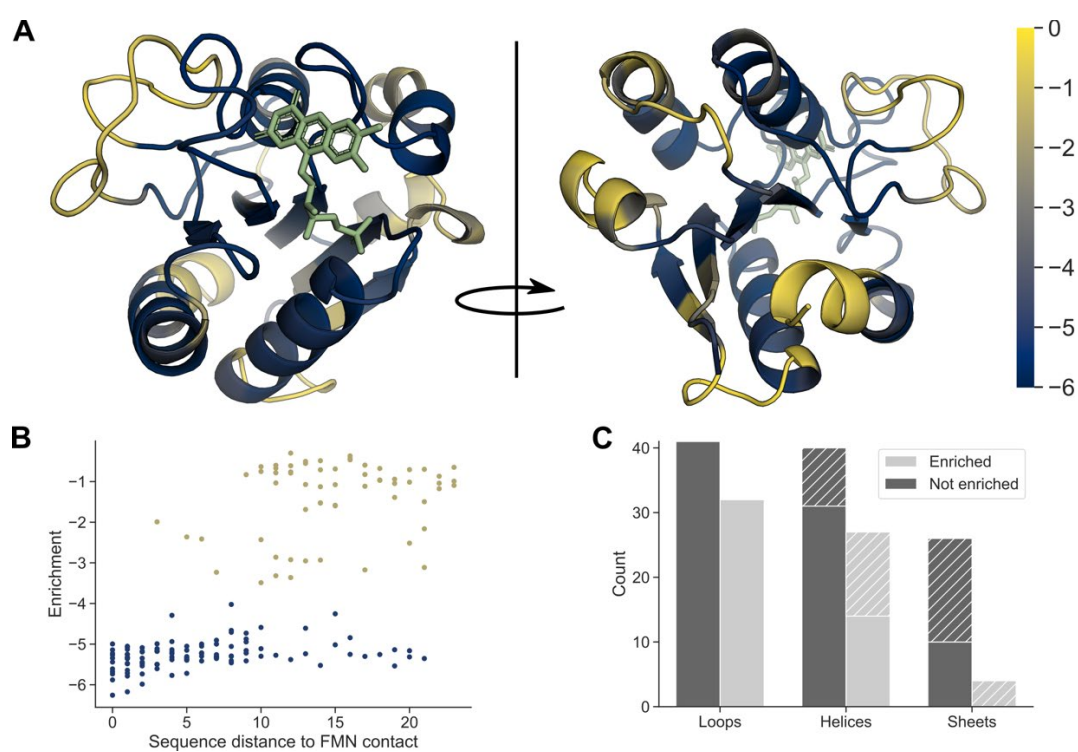
Figure 5

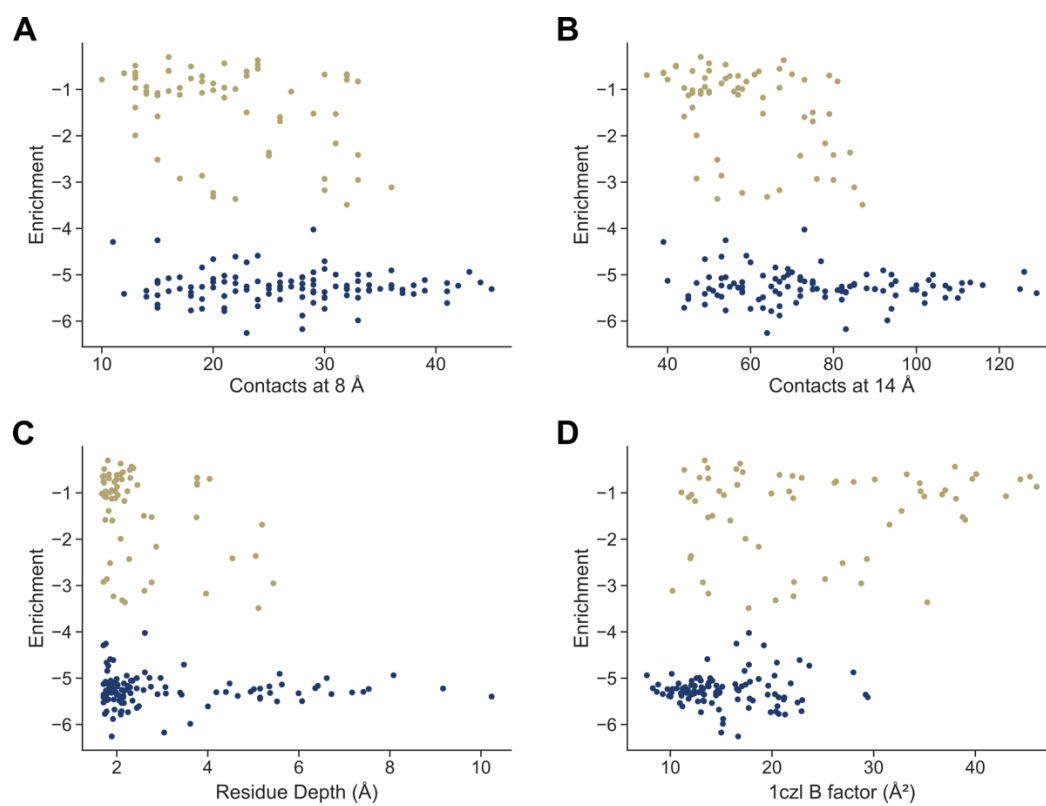
Figure 6

Figure 7

