

Laboratory evolution identifies elongated flavodoxins that support electron transfer to sulfite reductases

Albert Truong^{1,2}, Dru Myerscough², Ian Campbell², Josh 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

ABSTRACT

Flavodoxins (Flds) mediate the flux of electrons between oxidoreductases in diverse metabolic pathways. While dozens of Fld-partner oxidoreductases have been discovered, these only represent a subset of the oxidoreductases that couple with ferredoxin (Fd) protein electron carriers. To investigate whether Flds can support electron transfer to a sulfite reductase (SIR) that evolved to couple with a Fd, we evaluated the ability of Flds to transfer electrons from a Fd-NADP reductase (FNR) to a Fd-dependent SIR using growth complementation of a microbe with a sulfur metabolism defect. We show that Flds from cyanobacteria complement the growth of this microbe when coexpressed with an FNR and an SIR that evolved to couple with a plant Fd. To better understand the interaction of Fld with these partner oxidoreductases, we evaluated the effect of peptide insertion on Fld-mediated electron transfer. We observe a high insertion sensitivity within regions predicted to be proximal to the cofactor and partner binding sites and a high insertion tolerance within the loop that is used to differentiate short- and long-chain flavodoxins. These results represent the first evidence that Flds can support electron transfer to assimilatory SIRs, and they suggest that the pattern of peptide-insertion tolerance is influenced by interactions with oxidoreductase partners in electron transfer pathways.

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

INTRODUCTION

Flavodoxins are low potential electron carriers that use a flavin mononucleotide (FMN) cofactor to transfer electrons between partner oxidoreductases¹. Biochemical and genetic studies have shown that Flds couple with diverse oxidoreductases possessing cellular roles in glycolysis (pyruvate Fld oxidoreductase, pyruvate formate lyase activating enzyme)^{2,3}, photosynthesis (photosystem I)⁴, hydrogen metabolism (hydrogenase)⁵, amino acid synthesis (glutamate synthase, cobalamin-dependent methionine synthase)^{6,7}, nucleotide metabolism (ribonucleotide reductase)⁸, steroidogenesis (cytochrome P450)^{9,10}, isoprenoid biosynthesis (4-hydroxy-3-methylbut-2-enyl diphosphate synthase)¹¹, redox homeostasis (Fd-NADP reductase, Fld-quinone reductase)^{12,13}, lipid synthesis (acyl lipid desaturase)¹⁴, nitrogen assimilation (nitrogenase, nitrate reductase, nitrite reductase)^{6,15–17}, and sulfur metabolism (bisulfite reductase, biotin reductase)^{18,19}. In some organisms, Fld deletions result in growth defects, indicating the essential role of Flds in supporting metabolism²⁰. In addition, bioinformatic studies have revealed that some organisms have genomes with as many as ten Fld paralogs²¹. Currently, it is not clear why some organisms encode so many Fld paralogs. 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}. Currently, we do not understand how Flds paralogs evolve structures through mutation and elongation to support specialized coupling between different oxidoreductases and pathways.

Many genomes encoding Flds also have ferredoxin (Fd) protein electronic carriers²¹, which function as cellular electron transfer (ET) hubs. The relative

abundances of Fld and Fd protein electron carriers can vary widely. Gammaproteobacteria frequently have multiple Fld and Fd paralogs, while microbes from other taxonomic groups typically present higher abundances of Fd electron carriers²¹. Both Fds and Flds present low midpoint reduction potentials. Fld redox potentials range from -230 to -530 mV^{29,30}, 2Fe-2S Fds present potentials range from -150 to -500 mV, and 4Fe-4S Fds range from -200 to -650 mV³¹. In many organisms, Fld and Fd expression is controlled by environmental conditions^{32,33}, such as oxidative stress, salt stress, heavy metal toxicity, mineral availability, and light. Also, there is evidence that Fld and Fd expression levels are strongly coupled to iron availability^{1,34,35}, 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 Fds¹. While Fds and Flds support ET to an overlapping set of almost twenty partner proteins, Fds have been shown to couple with a much larger set (>80) of partner oxidoreductases^{1,31}. The extent to which Flds can support ET to many of these Fd-partner proteins in natural or synthetic cellular systems remains unclear.

Flds and Fds both support ET to nitrite reductases (NIR)³⁶, which share structural similarity with Fd-dependent SIRs³⁷. This observation led us to investigate whether Flds could support ET to Fd-type SIRs. Through genome mining, we previously identified microbes that encode Fd, Fd-dependent SIR, and Flds in their genomes³⁸. Herein, we show that cyanobacterial Flds support ET between a plant FNR and SIR using a cellular selection that requires ET from a FNR to a Fd-dependent SIR to complement a growth defect³⁹. Since some Fld paralogs evolved elongated structures to support recognition

of partner proteins⁴⁰, we use this selection to investigate where a cyanobacterial Fld tolerates insertions within its primary structure. We chose this class of mutational lesion because we hypothesized that it would only be tolerated within regions distal from residues that mediate molecular interactions with partner proteins. 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. Also, we find that the loop that differentiates short- and long-chain Flds presents a high tolerance to peptide insertion⁴⁰. These provide fundamental insight into the ways that Fld evolution could affect oxidoreductase interactions and implicate peptide-insertion profiling as a strategy to rapidly map oxidoreductase binding interfaces. They also identify new Fld-partner proteins that can be used as living electronic components for bioelectronics and synthetic biology^{41,42}.

RESULTS

Flds support ET to Fd-dependent 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 small protein electron carrier (Fd) that transfers electrons, and a Fd-dependent sulfite reductase that catalyzes the six-electron reduction of sulfite to sulfide^{43,44}. To investigate whether Flds might have evolved to substitute for Fds in this pathway, a prior study used bioinformatics to identify organisms encoding Flds as well as this three-component pathway³⁸. Photosynthetic cyanobacteria and algae from over ten genera were identified that contain all of these proteins (Fld, Fd, FNR, and SIR), including *Acaryochloris*, *Anabaena*, *Crocospaera*, *Gloeotheca*, *Nostoc*, *Ostreococcus*, *Prochlorococcus*, *Synechococcus*, *Thalassiosira*, and *Trichodesmium*. All of these organisms contained a single Fd-dependent SIR, although many contained more than one Fld and Fd. These results suggest that Flds may be capable of supporting electron transfer from FNR to Fd-dependent SIRs.

In prior studies, an *Escherichia coli* strain (EW11) was engineered so that growth is dependent upon Fd-mediated ET from FNR to SIR³⁹. This strain cannot grow when sulfate is the sulfur source due to a sulfur metabolism defect, and growth complementation of this defect has been used to report on electron transfer mediated by diverse Fds, including natural pathways made up of corn FNR, Fd, and SIR³⁹ as well as synthetic pathways containing non-cognate protein electron carriers^{41,45–48}. As such, this strain represents a simple approach to assess whether Flds can support electron transfer between FNR to SIR that evolved to couple with Fds. We first investigated

whether *Synechocystis* sp. PCC6803 (sFld1) and *Nostoc* sp. PCC7120 (nFld) Flds, which exhibit 67% identity, could support ET from FNR to SIR (Figure 1a). Each Fld was expressed using an anhydrotetracycline (aTc) inducible promoter (Figure S1a) within a strain also expressing *Zea mays* FNR and SIR using constitutive promoters (Figure S1b)⁴⁵. To identify optimal assay conditions, we evaluated the growth of cells expressing *Mastigocladus laminosus* Fd and corn FNR and SIR, which complements the growth defect of *E. coli* EW11⁴⁵. 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 Fd-dependent SIR in a synthetic pathway. They also identify growth conditions where Fld ET can be analyzed using a selection.

Evaluating Fld mutation tolerance. Structural studies have shown that some Flds have evolved elongated structures to support high affinity interactions with partner oxidoreductases²¹. While rational design studies have shown that loop removal can be used to disrupt oxidoreductase partner binding⁴⁰, there have been no studies examining how peptide insertion affects Fld electron transfer. 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 as insertion libraries with this sequence are easy to generate using

an existing method⁴⁹. To allow comparison of Fld1 insertion variants with native, 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 Fld 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). 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 S5). 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.

To establish which Fld variants support cellular ET, we divided the mean frequency of each variant after selection by the value observed in the naive library (Figure 3a). This analysis yielded *enrichment values* for each variant, defined as the

$\log_2(\text{selected:naive ratio})$. Analysis of the distribution of enrichment values revealed two major clusters of phenotypes, with some variants of intermediate enrichment (Figure 3b). After fitting a three-component gaussian mixture model to the enrichment data, the mean values for the largest modes were -3.00 and 1.45, respectively. The latter cluster overlaps with the enrichment value observed for the sFld2, 1.90 ± 0.14 . 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 Fd-C42A mutant, which does not complement growth^{45,50}. After 48 hours (Figure 3c), cells expressing individual peptide-insertion variants presented growth with a linear correlation 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 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 4). 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 25 to 30, 39 to 45, 71 to 80, 121 to 137, and 163 to 169. The motif spanning residues 121 to 137 represents the loop that is used to differentiate short- and long-chain Flds⁴⁰. 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 4, 31, 36 to 37,

44, 46, 72, 109-111, 120, 149, and 160. These findings show that peptide-insertion sensitivity varies with location in Fld primary structure.

Mapping insertion sensitivity onto Fld structure. To better understand the Fld features that influence insertion sensitivity, we compared our results with a sFld1 structural model from AlphaFoldDB⁵¹. Structural alignment of this Fld with a homolog having 72% sequence identity, *Synechococcus elongatus* Fld⁵², yielded an RMSD of 0.27 Å and revealed similar backbone and side chain conformations of residues at the conserved FMN-binding site (Figure S6). This finding indicated that the FMN 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 FMN (Figure 5A). All peptide-insertion sites proximal to the FMN cofactor ($\leq 6\text{Å}$) were sensitive to peptide insertion (Figure 5B). These findings show that Fld is most sensitive to insertions near the FMN cofactor.

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 the highest proportion of functional variants resulting from peptide insertion in loops, and the lowest in beta sheets, which are largely found within the core of the protein. We found that many enriched variants with insertions in helices were near loops, so positions within two residues of a loop were considered separately. Of the functional variants within helical and sheet structures, almost half had insertions

proximal to a loop. 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 also compared fitness with crystallographic B factors (Figure 6D) from the structure of a homologous Fld to assess whether insertion tolerance may be related to local flexibility. These findings show that native positions that tolerate peptide insertion tend to be less buried and have fewer residue-residue contacts, as well as 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^{53,54}. We observed two binding modes for the Fld-FNR complex (Figure S7) and aligned Fld (PDB ID 1czi) and *Z. mays* FNR (PDB ID 1jb9) crystal structures to each prediction to obtain the orientations of the FMN and FAD cofactors. To determine if the predicted conformations are compatible with fast intermolecular ET, we compared the cofactor orientations to those of cytochrome P450 reductases (CPRs) (Figure S8), a protein family comprising fused Fld and FNR domains⁵⁵. One of the predicted Fld-FNR binding modes recapitulated the side-to-side orientation of the FMN and FAD cofactors observed in CPRs with an interatomic distance of 3.2 Å, which is conducive to intermolecular ET. When we analyzed how peptide-insertion sensitivity correlates with Fld-FNR intermolecular contact densities in the modeled structural complex (Figure 7A), we found that sensitivity was related to intermolecular contact density (Figure 7B). All Fld variants with insertions at positions making ≥ 5 contacts with FNR were non-functional. In regions making 0 to 4, the insertion tolerance was bimodal. This finding

suggests that the pattern of peptide-insertion sensitivity contains 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 distances (~12 Å) observed in crystal structures of the SIR with its native Fd partner⁴³. A comparison of the intercofactor distances and orientations in the Fld-SIR and Fd-SIR complexes is shown in Figure S10. We next analyzed how peptide-insertion sensitivity correlates with Fld-SIR intermolecular contact densities in the predicted complex (Figure 7C). As observed with the Fld-FNR complex, peptide-insertion sensitivity correlated with intermolecular contact density in the Fld-SIR complex (Figure 7D). All variants having insertions at Fld locations making ≥5 contacts with FNR were non-functional. In regions making 0 to 4 contacts, the insertion tolerance was bimodal. This finding suggests that the pattern of insertion sensitivity contains information about the Fld regions that mediate binding to SIR.

DISCUSSION

Our results herein extend the list of Fld-partner oxidoreductases to include Fd-dependent SIRs. They also show the utility of synthetic electron transfer pathways for rapidly assessing coupling of protein electron carriers with different partner oxidoreductases³¹. The finding that cyanobacterial Flds can mediate ET from plant FNR to SIR implicate a role for this three-component ET pathway in microbes whose genomes encode all three of these oxidoreductases. Prior bioinformatic analysis revealed that a majority of the microbes containing FNR, Flds and Fd-dependent SIR are marine cyanobacteria, such as *Prochlorococcus* and *Synechocystis*³⁸. Biogeochemical studies have shown that iron is a limiting resource for primary productivity in marine settings⁵⁶, and they have shown that cyanobacteria compete for iron with heterotrophic bacteria and phytoplankton⁵⁷. Taken together with the finding that sulfate is abundant in marine settings, 28 mM⁵⁸, our results suggest that Fld-mediated electron transfer from FNR to SIR may occur under conditions when iron is sufficiently limiting so that it cannot support iron-sulfur cluster biogenesis critical to Fd electron transfer⁵⁹. To directly demonstrate a physiological role for the Fld-SIR interactions, studies will be needed to directly show coupling with cognate FNR and SIR under physiological conditions.

The finding that insertion sensitivity depends on proximity to cofactor and partner binding interfaces illustrates how peptide-insertion tolerance could be used to map residues that mediate molecular interactions in the absence of structural information. Peptide-insertion sensitivity correlates strongly with three structural parameters, including the density of intramolecular contacts, proximity to cofactor, and the density of

intermolecular contacts. Prior studies have probed protein structure by inserting peptides with a range of lengths^{60–64}. These studies found that peptide insertion can provide insight into molecular interactions⁶⁰, conditional phenotypes for genetic studies⁶¹, and sites compatible with the creation of non-disruptive affinity tags⁶². 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 libraries⁶⁵. Our study leveraged a recently described algorithm for domain insertion to create more uniform libraries⁶⁶, 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 length⁶⁷. Similar to a prior study that examined protein tolerance to single amino acid insertions⁶⁸, we found that the insertional tolerance correlates strongly with the identity of the secondary structure targeted by insertion, with loop tolerance being greater than that of helices and helices having a greater tolerance than beta strands, and we observed a correlation with insertion tolerance and proximity to the end of helix and strand secondary structures. Our study extends these findings by showing a correlation between insertion tolerance and proximity to cofactor and partner protein binding sites. In a recent study, peptide-insertion sensitivity also correlated with cofactor proximity within a membrane

protein wire⁶⁴. Taken together, these results support the idea that tolerance to insertion is constrained by the protein fold, requiring considerations of nonlocal interactions. Moreover, insertion tolerance at a given site varies based on the size and characteristics of the inserted peptide⁶⁹, which may display its own secondary and tertiary structure preferences. At times, these perturbations can lead to alternate frame folding⁷⁰. Such dramatic changes in protein structure create challenges for predicting the effects of insertions from native structures computationally except in specific cases⁷¹. Deep learning approaches offer promise for predicting fitness directly from sequence^{72–74}. However, they require larger insertion datasets for refinement and validation and to reveal principles governing insertion tolerance⁷⁵. This can be contrasted with deep mutational scanning experiments that study the effects of amino acid substitutions⁷⁶. The fitness effects of random mutations provide insight into how individual residues contribute to folding, stability, dynamics, and function. Amino acid preference at a given site in a protein is largely determined by the local molecular environment, as reflected by the recent success of deep learning approaches for variant prediction and sequence recovery from structure^{77–79}, rather than non-local interactions of importance for insertions.

Our results illustrate how insertional mutagenesis can be used to understand how proteins have evolved elongated structures, which has occurred through the recombination of protein domains⁸⁰. 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

mutations decreases as the targeted protein increases in thermostability^{81–83}. Peptide insertion profiles created with multiple homologs will be useful for establishing how topology, stability, and local energetics govern mutational tolerance⁸⁴. 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 insertion⁸⁵. Emerging machine learning approaches can be applied to these data sets in tandem with metagenomic data to predict design rules for elongated proteins^{69,74}. Such rules will be critical for rationally inserting larger polypeptides as a means to create protein switches for synthetic biology⁸⁶, 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^{41,45}.

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 complementation³⁹. 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 pSAC01⁴⁵. The control vectors expressing *Mastigocladus laminosus* Fd (pFd007) and an inactive C42A mutant of this Fd^{45,50}. 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 pFd007⁴⁵, 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 Gate⁸⁷. 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 described³⁹. To perform growth complementation, cells were transformed with pSAC01⁴⁵ 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 previously⁴⁹. In brief, the sFld1 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 step⁸⁸. 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, 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 containing m9sa medium (30 mL) containing with 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 add 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.

Peptide-insertion profile. The paired reads representing the forward and reverse sequencing data were merged using the BBmerge.sh script⁸⁹. 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 analysis⁹⁰. 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>). The frequency (F) of each insertion variant in each sequencing run was calculated by dividing the number of identical insertion variant reads by the total number of merged sequencing reads. The enrichment of mutants following selection was quantified by calculating the $\log_2(\text{fold change})$, where the fold change represents the ratio of the selected to the naive frequencies. When mutants had an enrichment score below -1.3, 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 = -3.0) 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 sFld1⁵¹. 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, PDB ID 1czl⁵², enabling identification of residues at

the conserved FMN binding site. AlphaFold-multimer was used to model the Fld-FNR or Fld-SIR complexes^{53,54}. 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 Ser12³⁹. Also, the N-terminal chloroplast localization sequence was omitted when modeling SIR. The primary structures used for all modeling are provided as supporting information. To obtain models for the Fld-FNR and Fld-SIR complexes, we used the ColabFold⁹¹ implementation of AlphaFold-multimer version 2.2.0 with multiple sequence alignments (MSAs) obtained from ColabFoldDB⁹¹ using the mmSEQs2 webserver^{92,93}. According to the default behavior of AlphaFold/ColabFold, sequences from the same organism were paired in the MSA, and unpaired sequences were also included. Templates were included from the PDB70 version 13Mar22^{94,95}. Cofactors were placed in each predicted structure by backbone alignment with structures of Fld PDB ID 1czi⁵², SIR PDB ID 5h92⁴³, and FNR PDB ID 1jb9⁹⁶.

The confidence metrics for each AlphaFold protein complex prediction are provided in Table S4. For the Fld-FNR complex, the first binding mode had a predicted template modeling (pTM) score of 0.91 and interface pTM (pTM) of 0.86, while the second binding mode had a pTM score of 0.89 and ipTM of 0.82. Predicted aligned error (PAE) and predicted local-distance difference test (pLDDT) plots are included in Figure S10. We chose the second binding mode for our analysis, despite the lower ipTM, since the orientations of the FNR and Fld as well as their cofactors closely recapitulated the productive binding mode of the Fld and FNR domains of cytochrome

reductases. This correspondence did not arise from the use of templates, as AlphaFold-multimer cannot use template information for predicting relative orientations of separate proteins⁵⁴. 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 approaches⁹⁷.

For the Fld-SIR complex, a single binding mode was predicted with a pTM score of 0.93 and interface pTM (ipTM) of 0.86. The PAE values are higher than the Fld-FNR complexes, which did not improve with additional recycling. Given the high MSA depth, this uncertainty is thought to arise from the shallow, smooth, and highly-charged interface of the corn SIR, which is compatible with multiple binding conformations. Ensemble binding has been observed with the corn SIR and its native ferredoxin partner, which binds in at least three conformations that primarily vary by rigid-body rotation along the inter-cofactor axis⁴³.

Structural calculations. To quantify intramolecular contact density, residue-residue interactions were identified using distance cutoffs of 8 Å and 14 Å, which are commonly used to consider medium-range contacts alone and medium-range and long-range contacts together^{98,99}. 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 residue¹⁰⁰. Local flexibility was inferred from the Cα B factors from a related Fld crystal structure⁵².

Statistics. Library selection data is presented as the mean and standard deviation of three biological (selective library) or technical (naive and non-selected library) replicates. P-values were obtained using a two-tailed Welch's t-test. A three-component

gaussian mixture model was fit to the enrichment data using the Scikit-learn library for Python¹⁰¹

SUPPLEMENTAL MATERIAL

Supplemental Figures S1-S10 and Tables S1-S4 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), and NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2010604 (to JTA).

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. Vigara 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, Dai 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. 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.

30. Kimmich N, Das A, Sevrioukova I, Meharena Y, Sligar SG, Poulos TL (2007) Electron transfer between cytochrome P450cin and its FMN-containing redox partner, cindoxin. *J. Biol. Chem.* 282:27006–27011.

31. 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.

32. 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.

33. Cassier-Chauvat C, Chauvat F (2014) Function and Regulation of Ferredoxins in the Cyanobacterium, *Synechocystis* PCC6803: Recent Advances. *Life Basel Switz.* 4:666–680.

34. 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.
35. 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.
36. Arizmendi JM, Serra JL (1990) Purification and some properties of the nitrite reductase from the cyanobacterium *Phormidium laminosum*. *Biochim. Biophys. Acta* 1040:237–244.
37. 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.
38. Guseva A (2020) Flavodoxin protein electron carriers: bioinformatic analysis and interactions with sulfite reductases. Available from: <https://scholarship.rice.edu/handle/1911/109196>
39. 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.
40. 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.
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. Atkinson JT, Chavez MS, Niman CM, El-Naggar MY (2022) Living electronics: A catalogue of engineered living electronic components. *Microb. Biotechnol.*
43. 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.
44. 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.
45. 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.

46. Campbell IJ, Kahanda D, Atkinson JT, Sparks ON, Kim J, Tseng C-P, Verduzco R, Bennett GN, Silberg JJ (2020) Recombination of 2Fe-2S Ferredoxins Reveals Differences in the Inheritance of Thermostability and Midpoint Potential. *ACS Synth. Biol.* 9:3245–3253.
47. 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.
48. 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.
49. 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.
50. 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.
51. 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.
52. Hoover DM, Drennan CL, Metzger AL, Osborne C, Weber CH, Pattridge 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.
53. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589.
54. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek 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>
55. 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.
56. Martin JH Iron as a Limiting Factor in Oceanic Productivity. In: Falkowski PG, Woodhead AD, Vivirito K, editors. *Primary Productivity and Biogeochemical Cycles in the Sea. Environmental Science Research.* Boston, MA: Springer US; 1992. pp. 123–137. Available from: https://doi.org/10.1007/978-1-4899-0762-2_8

57. Moore CM, Mills MM, Arrigo KR, Berman-Frank I, Bopp L, Boyd PW, Galbraith ED, Geider RJ, Guieu C, Jaccard SL, et al. (2013) Processes and patterns of oceanic nutrient limitation. *Nat. Geosci.* 6:701–710.
58. Andreae MO (1990) Ocean-atmosphere interactions in the global biogeochemical sulfur cycle. *Mar. Chem.* 30:1–29.
59. Erdner DL, Anderson DM (1999) Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. *Limnol. Oceanogr.* 44:1609–1615.
60. 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.
61. 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.
62. 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.
63. 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.
64. 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.
65. 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.
66. Coyote-Maestas W, He Y, Myers CL, Schmidt D (2019) Domain insertion permissibility-guided engineering of allostery in ion channels. *Nat. Commun.* 10:290.
67. Miton CM, Tokuriki N (2023) Insertions and Deletions (Indels): A Missing Piece of the Protein Engineering Jigsaw. *Biochemistry* 62:148–157.
68. 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.
69. 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.

70. Mitrea DM, Parsons LS, Loh SN (2010) Engineering an artificial zymogen by alternate frame protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 107:2824–2829.
71. Plaks JG, Brewer JA, Jacobsen NK, McKenna M, Uzarski JR, Lawton TJ, Filocamo SF, Kaar JL (2020) Rosetta-Enabled Structural Prediction of Permissive Loop Insertion Sites in Proteins. *Biochemistry* 59:3993–4002.
72. Notin P, Niekerk LV, Kollasch AW, Ritter D, Gal Y, Marks DS (2022) TranceptEVE: Combining Family-specific and Family-agnostic Models of Protein Sequences for Improved Fitness Prediction. :2022.12.07.519495. Available from: <https://www.biorxiv.org/content/10.1101/2022.12.07.519495v1>
73. Nijkamp E, Ruffolo J, Weinstein EN, Naik N, Madani A (2022) ProGen2: Exploring the Boundaries of Protein Language Models. Available from: <http://arxiv.org/abs/2206.13517>
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. Notin P, Dias M, Frazer J, Marchena-Hurtado J, Gomez A, Marks DS, Gal Y (2022) Tranception: protein fitness prediction with autoregressive transformers and inference-time retrieval. Available from: <http://arxiv.org/abs/2205.13760>
76. Fowler DM, Fields S (2014) Deep mutational scanning: a new style of protein science. *Nat. Methods* 11:801–807.
77. Hsu C, Verkuil R, Liu J, Lin Z, Hie B, Sercu T, Lerer A, Rives A (2022) Learning inverse folding from millions of predicted structures. :2022.04.10.487779. Available from: <https://www.biorxiv.org/content/10.1101/2022.04.10.487779v2>
78. Anand N, Eguchi R, Mathews II, Perez CP, Derry A, Altman RB, Huang P-S (2022) Protein sequence design with a learned potential. *Nat. Commun.* 13:746.
79. Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF, Wicky BIM, Courbet A, de Haas RJ, Bethel N, et al. (2022) Robust deep learning-based protein sequence design using ProteinMPNN. *Science* 378:49–56.
80. Netzer WJ, Hartl FU (1997) Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388:343–349.
81. 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.
82. 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.

83. 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.
84. 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.
85. Ostermeier M (2005) Engineering allosteric protein switches by domain insertion. *Protein Eng. Des. Sel. PEDS* 18:359–364.
86. 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.
87. Engler C, Kandzia R, Marillonnet S (2008) A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE* 3:e3647.
88. Bosley AD, Ostermeier M (2005) Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomol. Eng.* 22:57–61.
89. Bushnell B, Rood J, Singer E (2017) BBMerge - Accurate paired shotgun read merging via overlap. *PloS One* 12:e0185056.
90. 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.
91. 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.
92. Steinegger M, Söding J (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35:1026–1028.
93. 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.
94. 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.
95. 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.
96. 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.

97. 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.
98. 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.
99. 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.
100. Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 38:305–320.
101. 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. Fld 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 complemented the growth of *E. coli* EW11 after 48 hours at 37°C when the inducer anhydrotetracycline (aTc) was added to growth medium having sulfate as the only sulfur source. As a positive control, we evaluated complementation when cells expressed *Mastigocladus laminosus* (Fd), which has been shown to couple with FNR and SIR. As a negative control, we evaluated complementation using a C42A mutant of this Fd, which cannot coordinate an iron-sulfur cluster. Cells expressing Fd and sFld1 presented significantly higher growth than the negative control ($p < 0.02$; two-tailed t test). While the average nFld OD is ~5-fold higher than C42A, the difference in the mean value of OD from the C42A or uninduced culture only approaches significance ($p = 0.0501$ and $p = 0.0503$, respectively; two-tailed t test). Cells expressing a Fld from *Synechocystis* PCC6803 containing silent mutations codons 30 and 31 (sFld2) also presented significantly higher growth than this control ($p < 0.02$, two-tailed t test) and do not grow to significantly higher ODs than cells expressing the native sFld1 ($p = 0.887$). Error bars represent the standard deviation from three biological replicates.

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 the cofactor (FMN), and intermolecular contacts made with the donor (FNR) and acceptor (SIR) partners. Pairs of residues were considered in contact if any constituent heavy atoms were within 8 Å, while enrichment is defined as an enrichment score greater than -1.3. The areas of dots corresponding to interactions between proteins are proportional to the number of residue-residue contacts. (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, 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. A linear fit yields an $r^2 = 0.865$. Error bars represent standard deviation from three biological replicates. Dashed lines indicate the mean calculated enrichment value and growth complementation of the sFld2-expressing positive control.

Figure 4. Insertion variant fitness comparison. The fitness of each insertion variant relative to wild type-like Fld is mapped onto primary structure, contacts with the cofactor (FMN), and intermolecular contacts made with the donor (FNR) and acceptor (SIR) oxidoreductase partners.

Figure 5. Insertion tolerance correlates with loop proximity and cofactor distance. (A) The AlphaFold-predicted structure of the *Synechocystis* Fld colored by enrichment

values from the selection. The FMN cofactor was placed by backbone alignment with the crystal structure 1czi. **(B)** Relative enrichments of all variants plotted against the minimum distance between any atom in the residue preceding the insertion site and any atom in the FMN cofactor. **(C)** Relative abundances of enriched and depleted variants by secondary structure. Diagonal lines denote 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. Relationship 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 *Synechococcus elongatus* Fld (PDB ID 1czi).

Figure 7. Positions at the predicted interfaces with partner oxidoreductases do not tolerate insertions. **(A)** One of two predicted binding modes predicted for the Fld-FNR complex. This binding mode recapitulates the interflavin geometry observed among cytochrome reductases. **(B)** Enrichments of Fld insertion variants plotted against numbers of intermolecular residue-residue contacts with FNR at 8 Å. **(C)** The predicted binding modes for the Fld-SIR complex. **(D)** Enrichments of 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 insertion.

Figure 1

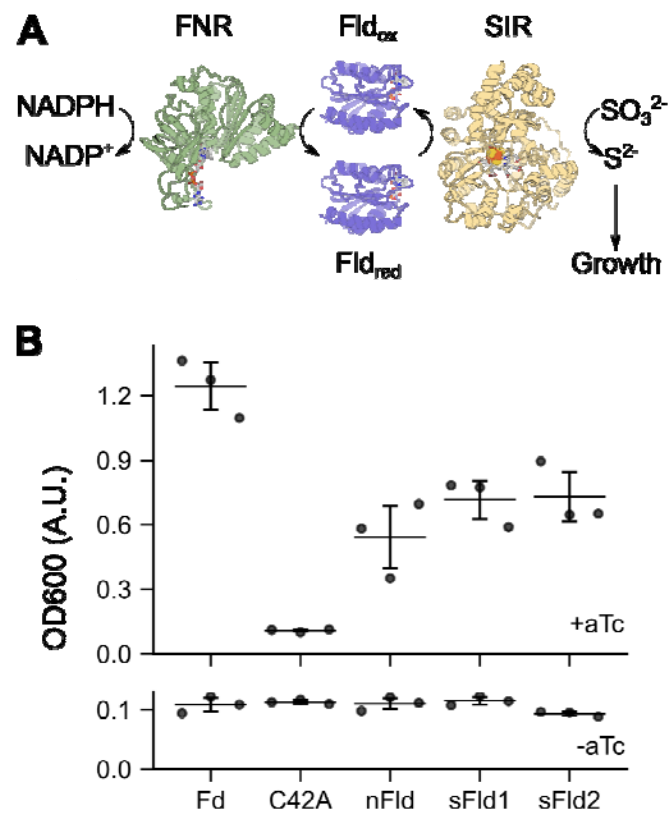


Figure 2

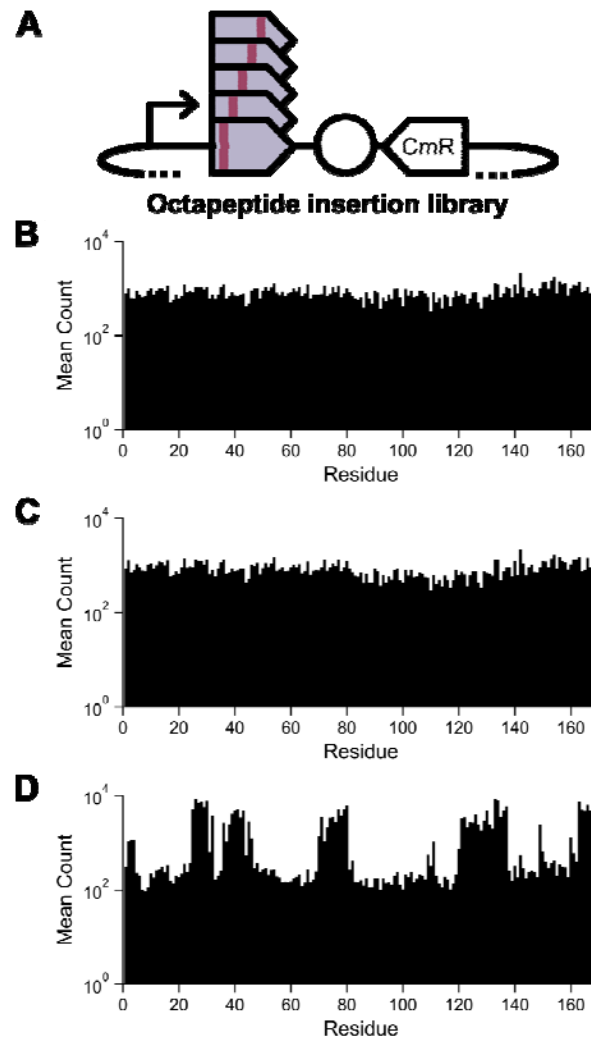


Figure 3

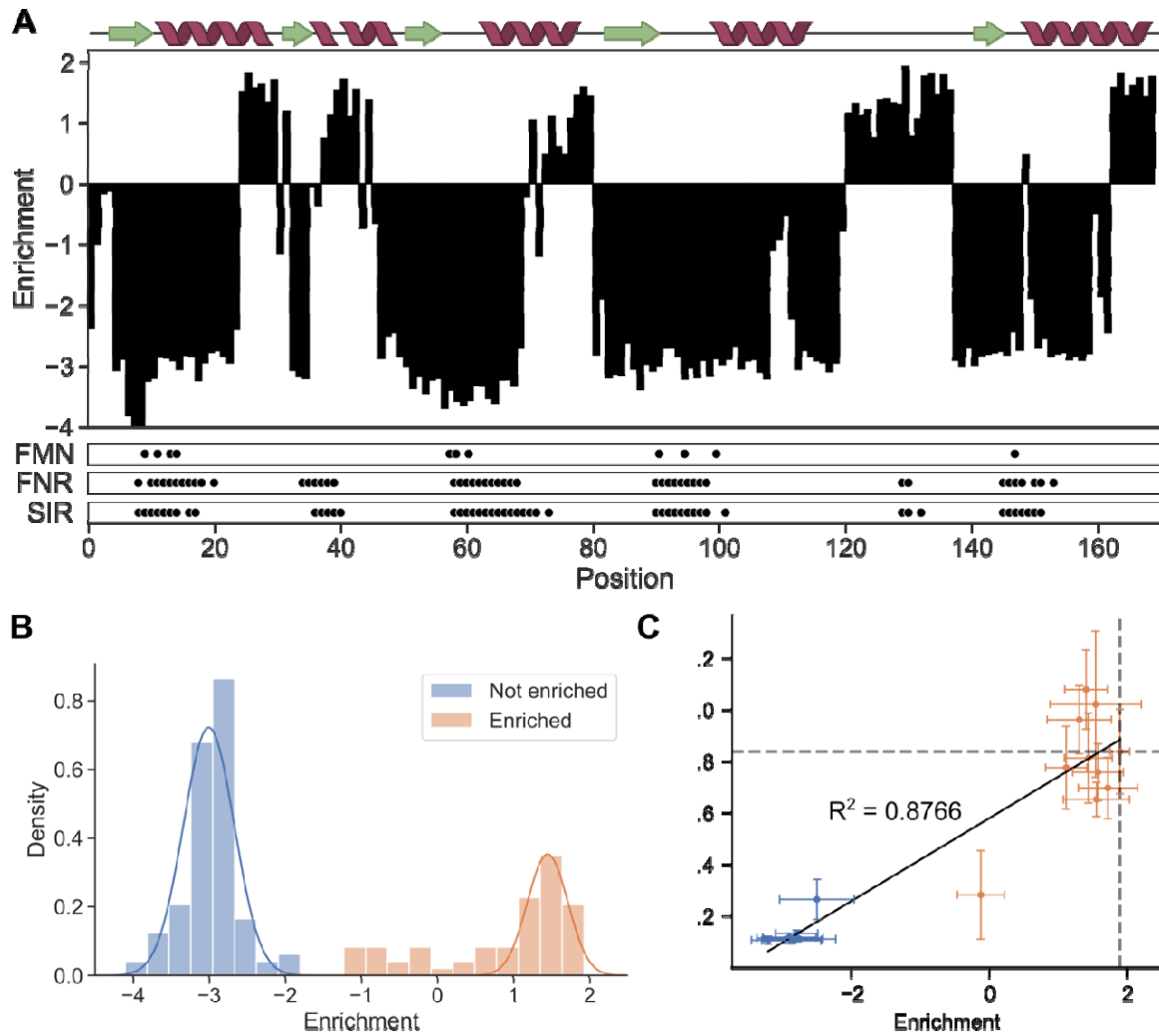


Figure 4

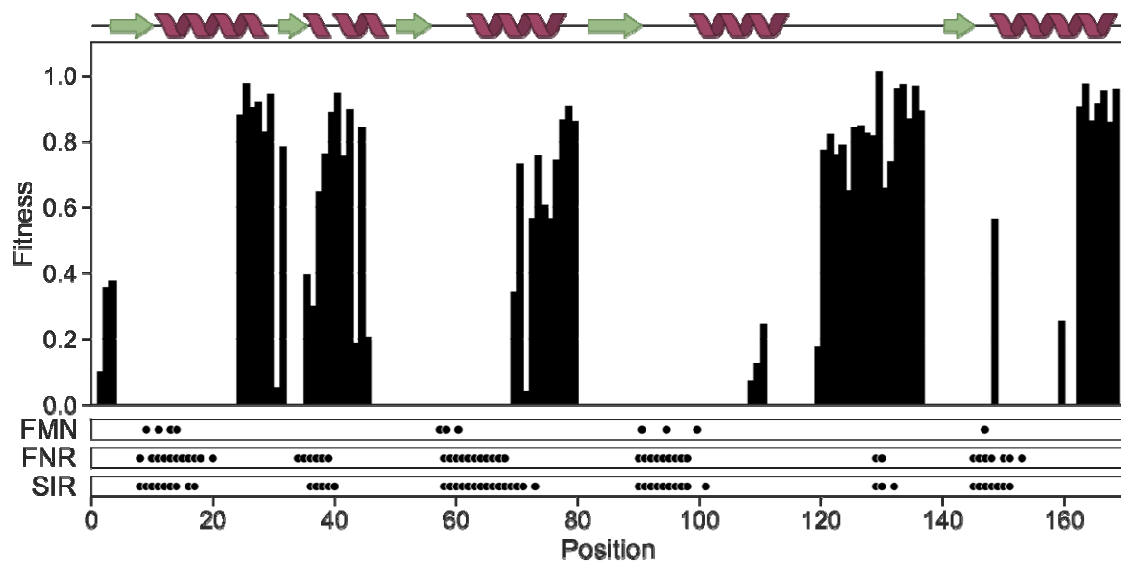


Figure 5

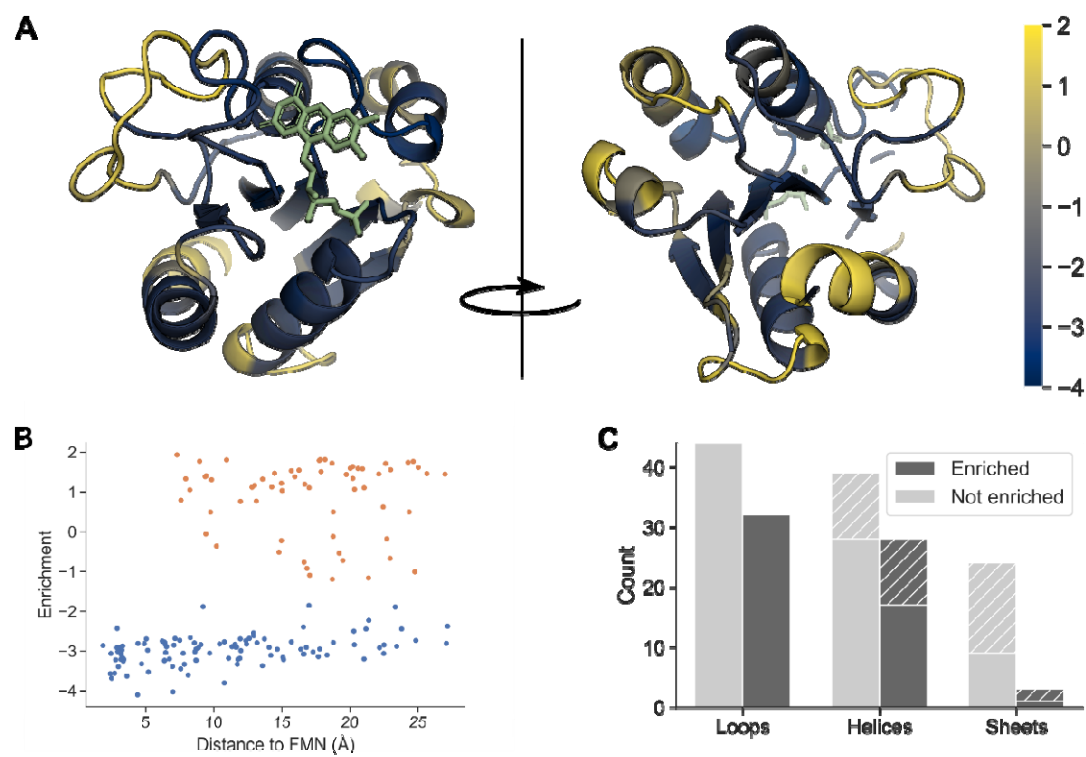


Figure 6

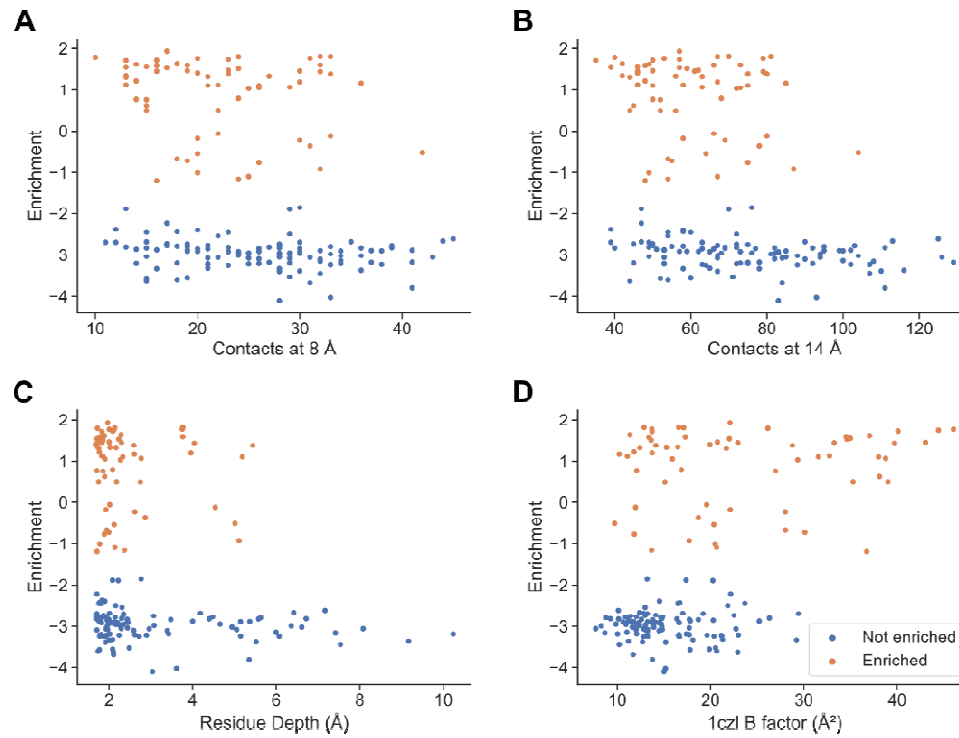


Figure 7

