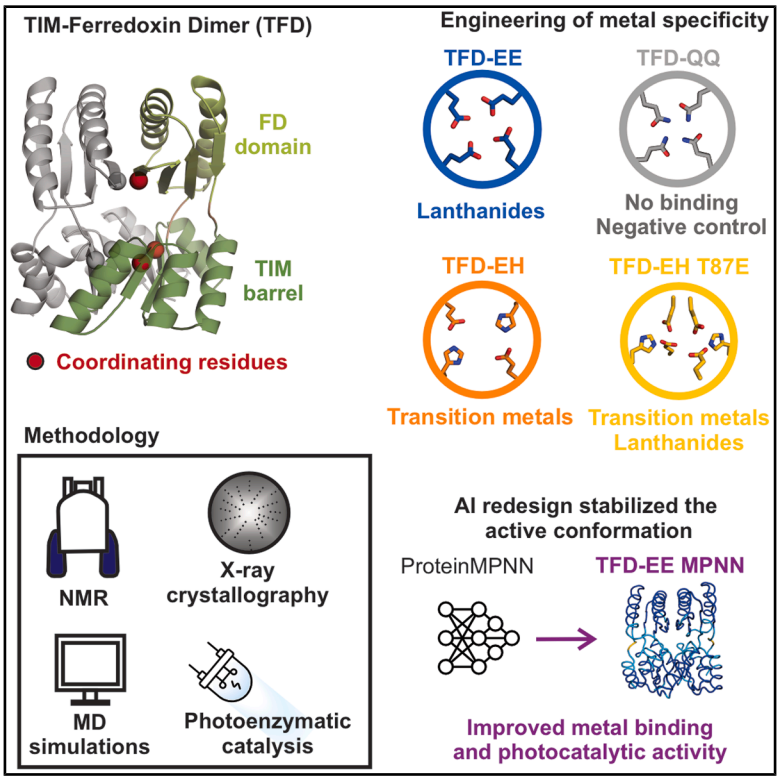
Structure

Article

Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering

Graphical abstract



Highlights

- Rational design modulates the metal binding specificity of a *de novo* protein scaffold
- Inter-domain linkers introduce conformational flexibility
- An undesired conformational state limits metal binding and photoenzymatic activity
- *ProteinMPNN*-based redesign shifts the equilibrium and stabilizes the active conformation

Authors

Paula Wagner Egea, Florent Delhommel, Ghulam Mustafa, ..., Michael Groll, Michael Sattler, Cathleen Zeymer

Correspondence

cathleen.zeymer@tum.de

In brief

Wagner Egea & Delhommel et al. performed an in-depth structural and biophysical study of a *de novo* protein scaffold with metal-dependent photoenzymatic activity. By combining X-ray crystallography, NMR spectroscopy, MD simulations, rational and AI-guided redesign, they analyzed and modulated scaffold dynamics and protein function, offering new strategies for metalloenzyme engineering.

Wagner Egea et al., 2026, Structure *34*, 1–13
January 8, 2026
https://doi.org/10.1016/j.str.2025.10.010

CellPress

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

Structure

CellPress
OPEN ACCESS

Article

Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering

Paula Wagner Egea,[1,5] Florent Delhommel,[2,3,5] Ghulam Mustafa,[1] Florian Leiss-Maier,[1] Lisa Klimper,[1] Thomas Badmann,[1] Anna Heider,[1] Idoia Wille,[1] Michael Groll,[1] Michael Sattler,[2,3] and Cathleen Zeymer[1,4,6,*]

[1]Center for Functional Protein Assemblies & Department of Bioscience, TUM School of Natural Sciences, Technical University of Munich (TUM), 85748 Garching, Germany

[2]Bavarian NMR Center & Department of Bioscience, TUM School of Natural Sciences, Technical University of Munich (TUM), 85748 Garching, Germany

[3]Institute of Structural Biology, Molecular Targets and Therapeutics Center, Helmholtz Munich, 85764 Neuherberg, Germany

[4]TUM Catalysis Research Center, Technical University of Munich (TUM), 85748 Garching, Germany

[5]These authors contributed equally

[6]Lead contact

*Correspondence: cathleen.zeymer@tum.de

https://doi.org/10.1016/j.str.2025.10.010

SUMMARY

Incorporating metal cofactors into computationally designed protein scaffolds provides a versatile route to novel protein functions, including the potential for new-to-nature enzyme catalysis. However, a major challenge in protein design is to understand how the scaffold architecture influences conformational dynamics. Here, we characterized structure and dynamics of a modular *de novo* scaffold with flexible inter-domain linkers. Three rationally engineered variants with different metal specificity were studied by combining X-ray crystallography, NMR spectroscopy, and molecular dynamics simulations. The lanthanide-binding variant was initially trapped in an inactive conformational state, which impaired efficient metal coordination and cerium-dependent photocatalytic activity. Stabilization of the active conformation by AI-guided sequence optimization using *ProteinMPNN* led to accelerated lanthanide binding and a 10-fold increase in k_{cat}/K_m for a photoenzymatic model reaction. Our results suggest that modular scaffold architectures provide an attractive starting point for *de novo* metalloenzyme engineering and that *ProteinMPNN*-based sequence redesign can stabilize desired conformational states.

INTRODUCTION

The field of protein design has witnessed major breakthroughs in recent years.[1,2] Building on the foundation of physics-based methods,[3] the transition to artificial intelligence (AI)-powered approaches has greatly increased the accessibility and accuracy of both protein structure prediction and *de novo* design.[4–11] Deep learning algorithms, trained on the large amount of publicly available protein sequences and structures, can generate robust and functional *de novo* proteins for a wide range of applications. However, while functions related to tight and specific binding have been implemented successfully,[12,13] the generation of highly active *de novo* enzymes remains challenging.[14–18]

Enzymes with new-to-nature activities can be obtained by integrating metal cofactors into *de novo* proteins,[19–21] with pioneering work in metalloprotein design dating back more than three decades.[22] Highlights of artificial metalloenzymes based on computationally designed protein scaffolds include hydrolases,[23,24] carbene transferases,[25,26] oxygen-activating enzymes,[27] and Diels-Alderases.[28] However, their catalytic efficiency usually had to be optimized by directed evolution, which utilizes iterative cycles of mutagenesis and screening or selection.[29] The current limitations can have different reasons, such as inefficient substrate recognition and product release, suboptimal active-site preorganization and catalytic residue pK_a values, but also the lack of conformational dynamics at relevant timescales, which drive catalysis but are rarely considered in the enzyme design process.[30–32] While the functional dynamics of many natural and re-engineered proteins have been studied in detail, often using NMR spectroscopy and molecular dynamics (MD) simulations,[33–41] our knowledge of *de novo* proteins remains limited in this context.

The TIM-ferredoxin dimer (TFD) scaffold is a *de novo* protein originally designed in a physics-based approach.[42,43] We previously reported a scaffold variant that binds lanthanide ions with femtomolar affinity (TFD-EE)[43,44] and further optimized it for cerium photoredox catalysis.[45] The protein scaffold comprises two homodimeric domains: a split *de novo* TIM barrel, which

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

carries the metal binding site in its dimer interface, and a *de novo* ferredoxin (FD) fold that serves as a lid covering the active site (Figures 1A and 1B). The domains are connected by two pairs of tri-glycine linkers. While most *de novo* protein scaffolds are hyperstable and rigid building blocks with limited conformational flexibility, TFD's modular architecture may enable inter-domain motions while maintaining crucial interactions between the domains. Here, we aimed to test this hypothesis, thereby evaluating the suitability as a starting scaffold for *de novo* metalloenzyme design. We started by rationally engineering TFD's metal binding preference, moving from lanthanides to first-row transition metal ions, and then studied the associated changes in protein structure and dynamics in an integrated approach using X-ray crystallography, NMR spectroscopy, MD simulations, and other biophysical methods. We identified an unfavorable conformational equilibrium as a functional bottleneck in this protein scaffold and applied AI-guided sequence redesign to overcome this limitation.

RESULTS

Point mutations modulate TFD's metal specificity and overall structure

Metal ions differ in their coordination preferences. In the dimeric TFD-EE scaffold, the metal binding site consists of two pairs of glutamate residues (E31 and E154). It is ideally suited to bind trivalent lanthanide ions, as they are oxophilic, hard Lewis acids that prefer high coordination numbers (Figures 1C and 1D).[43] To alter the metal specificity of our *de novo* scaffold, we rationally introduced the mutation E154H and named that variant TFD-EH. We hypothesized that the resulting binding site with two glutamates and two histidines inside the central cavity of the split TIM barrel would favor divalent first-row transition metal ions. A crystal structure of metal-free TFD-EH, determined at 1.6 Å resolution (PDB code: 9QUC), revealed that the single mutation substantially influenced the overall scaffold structure. Compared to TFD-EE (PDB code: 6ZV9),[43] the TIM barrel geometry is more oval-shaped and slightly distorted, the C-terminal helices point outward, and the FD domain is rotated (Figure 1C). In the absence of metal ions, the coordinating residues form stabilizing hydrogen bonds with each other (Figure 1E). A predicted pK_A = 9.4 for the histidine (calculated with *PROPKA3*),[46] even suggests ionic interactions at neutral pH. This is in agreement with a previously determined structure of a related variant, TFD-HE (PDB code: 6WXO),[43] in which E31 was mutated instead.

While TFD-HE did not exhibit specific metal binding, TFD-EH formed complexes with Cu^{II} and Ni^{II}, and crystal structures were obtained at 1.9 Å and 2.1 Å resolution, respectively (PDB codes: 9QUD and 9QUI). To our surprise, anomalous diffraction data clearly revealed a dinuclear metal binding site with two transition metal ions 5.5 Å apart (Figures 1F and S1). Each metal is coordinated by one glutamate and one histidine. Electron density at the free coordination sites was modeled with imidazole, a common ligand that was present in the crystallization conditions. Apart from local rearrangements in the coordinating residues, the overall structure closely resembles that of metal-free TFD-EH. However, compared to Tb^{III} in TFD-EE, the transition metal ions are located 4 Å higher above the TIM barrel core (Figure 1C). To further stabilize this dinuclear binding mode, we introduced an additional coordinating residue per monomer (T87E), projecting down from the FD domain. The crystal structures of TFD-EH T87E in complex with Zn^{II} and Co^{II} (PDB codes: 9QUL and 9QUO, both at 1.9 Å resolution) confirmed that the newly introduced glutamate indeed interacts with the bound metal ions (Figures 1G and S1). Notably, the TFD scaffold can support different coordination geometries: Zn^{II} shows tetrahedral geometry in TFD-EH T87E, while Cu^{II} coordination in the original TFD-EH is planar.

Next, we set out to compare the metal binding affinities of the different scaffold variants. A knock-out mutant with two pairs of glutamines (TFD-QQ) served as a negative control. The very tight lanthanide binding to TFD-EE was quantified previously in chelator-buffered titrations using tryptophan-enhanced Tb^{III} luminescence as a specific spectroscopic binding readout.[43,44] Here, a tryptophan (W6) in close proximity to the metal site serves as an intrinsic antenna (Figure 1H). Expectedly, the affinity for lanthanides dropped substantially due to the E154H mutation. While TFD-EE showed femtomolar affinity,[44] TFD-EH binds Tb^{III} in the micromolar range (K_D = 20 μM) at pH 7.5 (Figure 1I). For TFD-EH T87E, we observed a biphasic binding curve, indicating tight binding to one equivalent of Tb^{III} ($K_D \leq$ 500 nM) and weaker binding to a second one ($K_D \approx$ 10 μM).

Lanthanide luminescence also provides a readout for competitive titrations, displacing Tb^{III} with another metal ion of choice (Figure 1J). However, the ambiguity in lanthanide to protein stoichiometry, together with the assumption that the second equivalent of transition metal will not bind with the same affinity as the first one, impaired an exact quantitative analysis. Still, relative statements comparing the different metals and protein variants are possible. We titrated a selection of divalent first-row transition metal ions to Tb^{III}-bound protein and observed that the binding preference follows the Irving-Williams series,[47] with affinities decreasing from Cu^{II} to Zn^{II} to Ni^{II} to Co^{II}. While Tb^{III} was fully displaced in TFD-EH by sub-stoichiometric amounts of Zn^{II}, ca. 50% of the luminescence signal remained for TFD-EH T87E, suggesting that only the weakly bound equivalent can be displaced readily. Moreover, Cu^{II} displays an additional quenching of the lanthanide emission due to a photophysical process.[48]

Overall, our crystal structures and metal titrations demonstrate that this *de novo* scaffold has a high plasticity and can be rationally engineered to accept different metals in its central binding site. Single mutations led to pronounced effects on overall structure and metal coordination.

Flexible inter-domain linkers confer conformational dynamics

While natural enzymes typically exhibit conformational flexibility, *de novo* proteins are often considered hyperstable and rigid, with limited conformational dynamics. However, the TFD scaffold contains tri-glycine linkers that connect its two domains. We hypothesized that these linkers may provide flexibility and enable inter-domain motions. To study protein dynamics in our system, we combined NMR spectroscopy and MD simulations.

Due to the pronounced structural differences between TFD-EE and TFD-EH, their ^{1}H-^{15}N HSQC spectra differed so much that the backbone chemical shifts were assigned using triple-resonance NMR experiments (Figures S5 and S6). We next performed metal titrations and analyzed the chemical shift

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

CellPress
OPEN ACCESS

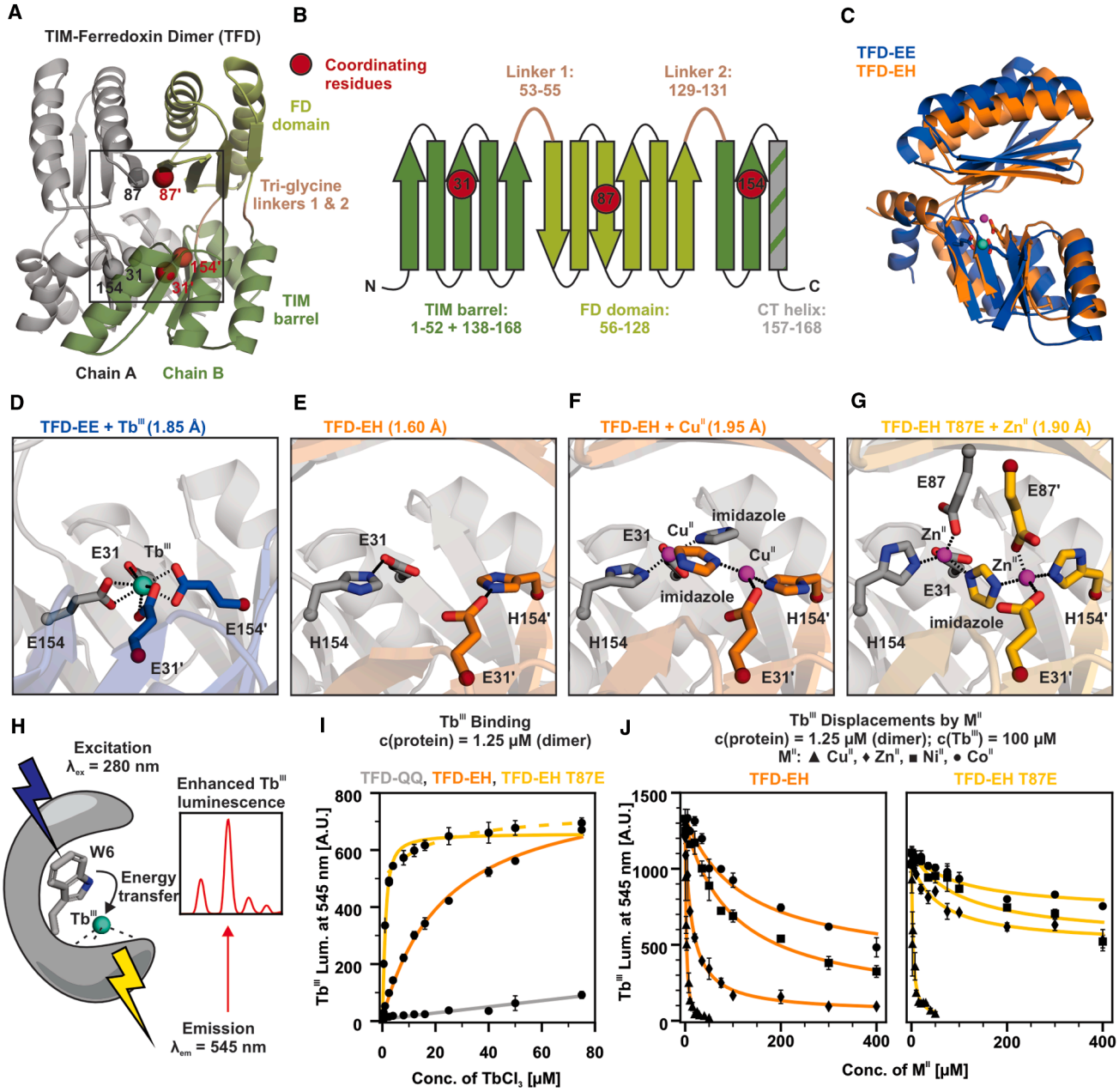


Figure 1. Structure and metal specificity of engineered TFD variants

(A) Overall structure of the homodimeric TFD scaffold with TIM barrel domain in dark green, ferredoxin (FD) domain in light green, tri-glycine linkers in brown, and positions of metal-coordinating residues as red spheres.

(B) Secondary structure diagram of a TFD monomer.

(C) Superposition of crystal structures obtained for TFD-EE + Tb^{III} and TFD-EH + Cu^{II}. For clarity, only one monomer with its two coordinating residues (31 + 154) and the respective metal ion position (Tb^{III} in green, Cu^{II} in pink) is shown. Note that these TFD variants differ by only one residue per monomer, namely E154H.

(D-G) Close-up views of the metal binding sites located in the dimer interface: TFD-EE + Tb^{III} (D), metal-free TFD-EH (E), TFD-EH + Cu^{II} (F), TFD-EH T87E + Zn^{II} (G). While TFD-EE binds one lanthanide ion per split TIM barrel dimer, each monomer of the TFD-EH variants coordinates one transition metal ion, generating a dinuclear site in the dimer interface.

(H) Tryptophan-enhanced Tb^{III} luminescence as a readout for metal binding titrations. Residue W6 is located in close proximity to the metal binding site and serves as an intrinsic antenna.

(I) Direct titration of $TbCl_3$ to metal-free TFD-EH (orange), TFD-EH T87E (yellow), and TFD-QQ (gray). The latter represents a negative control. The dashed yellow line is a separate fit for the binding of the second equivalent of metal. Measurements were done in triplicate, and their standard deviations are shown as error bars in the plot.

(J) Displacement titrations: a saturated complex formed from 1.25 µM dimeric TFD-EH (orange) or TFD-EH T87E (yellow) and 100 µM $TbCl_3$ was titrated with $CoSO_4$, $NiSO_4$, $ZnSO_4$, and $CuCl_2$. Measurements were done in duplicate, and their standard deviations are shown as error bars in the plot.

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

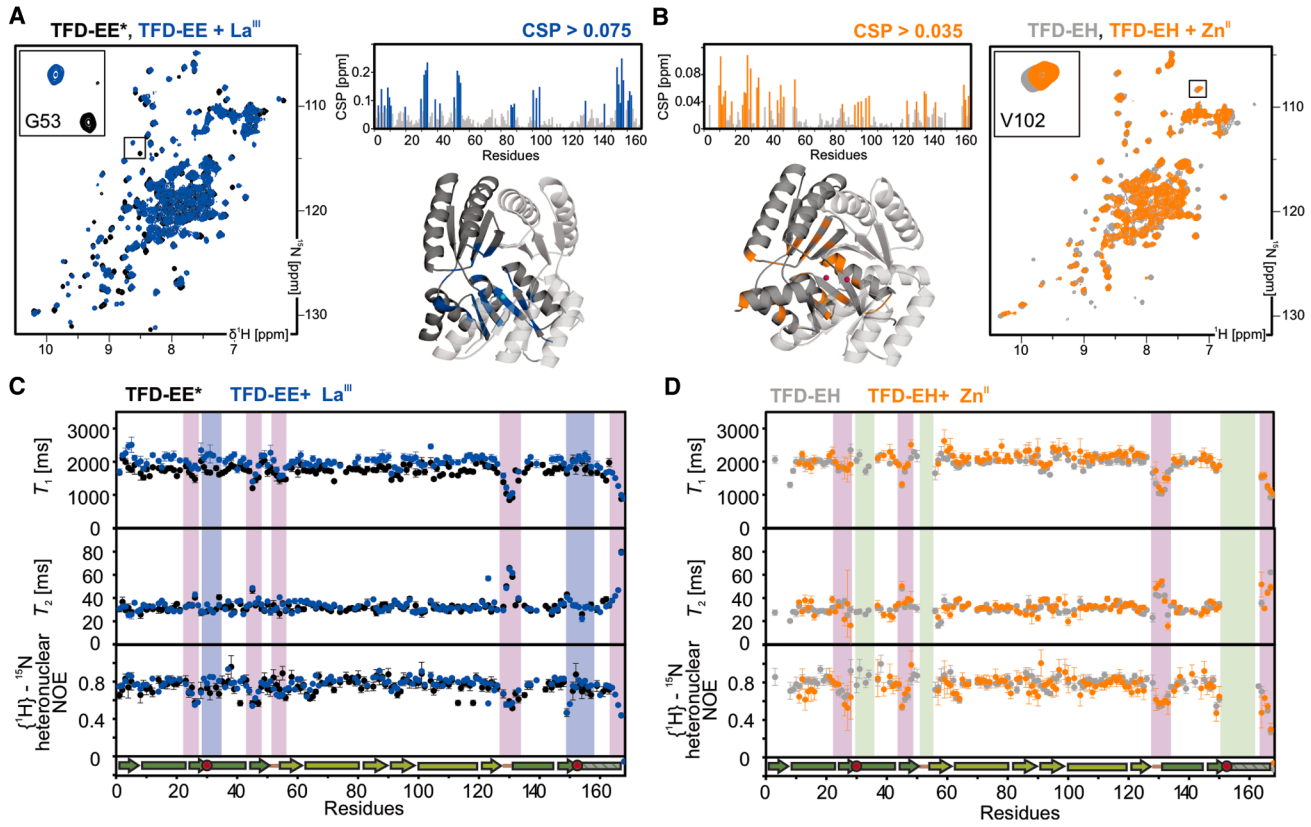


Figure 2. NMR chemical shift perturbation (CSP) and ^{15}N relaxation data

(A) ^{1}H-^{15}N HSQC spectra of metal-free TFD-EE after pre-incubation at 40°C (black) and TFD-EE + 1 eq $LaCl_3$ (blue). The inset shows residue G53. Residues with CSP>0.075 upon metal binding are highlighted and plotted on one monomer of the crystal structure.

(B) ^{1}H-^{15}N HSQC spectra of metal-free TFD-EH (gray) and TFD-EH + 2 eq $ZnSO_4$ (orange). The inset shows residue V102. Residues with CSP>0.035 upon metal binding are highlighted and plotted on one monomer of the crystal structure.

(C) ^{15}N relaxation data (T_1, T_2, and {^{1}H}-^{15}N heteronuclear NOE) for TFD-EE after pre-incubation at 40°C (black) and TFD-EE + 1 eq $LaCl_3$ (blue). Regions with increased flexibility are marked in pink. Regions with significantly different values between metal-free and metal-bound states are marked in blue. Measurements were done in triplicate, and their standard deviations are shown as error bars in the plot.

(D) ^{15}N relaxation data (T_1, T_2, and {^{1}H}-^{15}N heteronuclear NOE) for metal-free TFD-EH (gray) and TFD-EH + 2 eq $ZnSO_4$ (orange). Regions with increased flexibility are marked in pink. Residues experiencing significant line broadening are marked in mint. Measurements were done in triplicate, and their standard deviations are shown as error bars in the plot.

perturbations (CSPs). Given that most lanthanide ions and Cu^{II} are paramagnetic, we focused on the diamagnetic La^{III} and Zn^{II} for the NMR titrations (Figures 2A and 2B). When adding La^{III} to TFD-EE, we observed the most pronounced chemical shift changes for residues in the β-strands of the TIM barrel, where the metal binding site is located. Furthermore, the inner triglycine linker (Linker 1) as well as the C-terminal helix (CT helix) were strongly influenced. Similarly, Zn^{II} binding to TFD-EH induced distinct peak shifts, albeit with smaller amplitude. Imidazole, which was found as a ligand in the crystal structures, did not cause notable spectral changes. TFD-QQ served as a negative control to account for nonspecific metal interactions on the protein surface (Figure S7). TFD-EH T87E was also titrated with La^{III}, where concentration-dependent chemical shift changes support the proposed stoichiometry of two lanthanides per dimer (Figure S8).

We next assessed backbone flexibility from ^{15}N NMR relaxation data (T_1, T_2, {^{1}H}-^{15}N heteronuclear NOE) for TFD-EE and TFD-EH in the presence and absence of metal (Figures 2C and 2D). Both proteins showed remarkably similar relaxation profiles, with and without metal. The tri-glycine linkers and the CT helix showed high flexibility at fast, sub-nanosecond timescales, while both domains are otherwise rigid units. Furthermore, residues surrounding the metal-coordinating positions in TFD-EH exhibited line broadening, which suggests conformational exchange at μs-ms timescales. Unfortunately, TFD-EH T87E precipitated during extended NMR experiments, preventing us from obtaining high-quality relaxation data for this variant.

To explore conformational dynamics, we performed MD simulations. Replica simulations were combined to generate 3 μs trajectories for each scaffold variant, with and without bound metal. The use of optimized parameters for Zn^{II} and Cu^{II} was key to reliably model the metal coordination throughout the simulation.[49,50] We analyzed the RMSD values for individual regions of the protein and observed distinct differences (Figures 3A and S11). Representative MD trajectories are shown in Videos S1, S2, and S3, with per-residue RMSD values mapped on the structures and plotted against the sequence. This analysis

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

CellPress
OPEN ACCESS

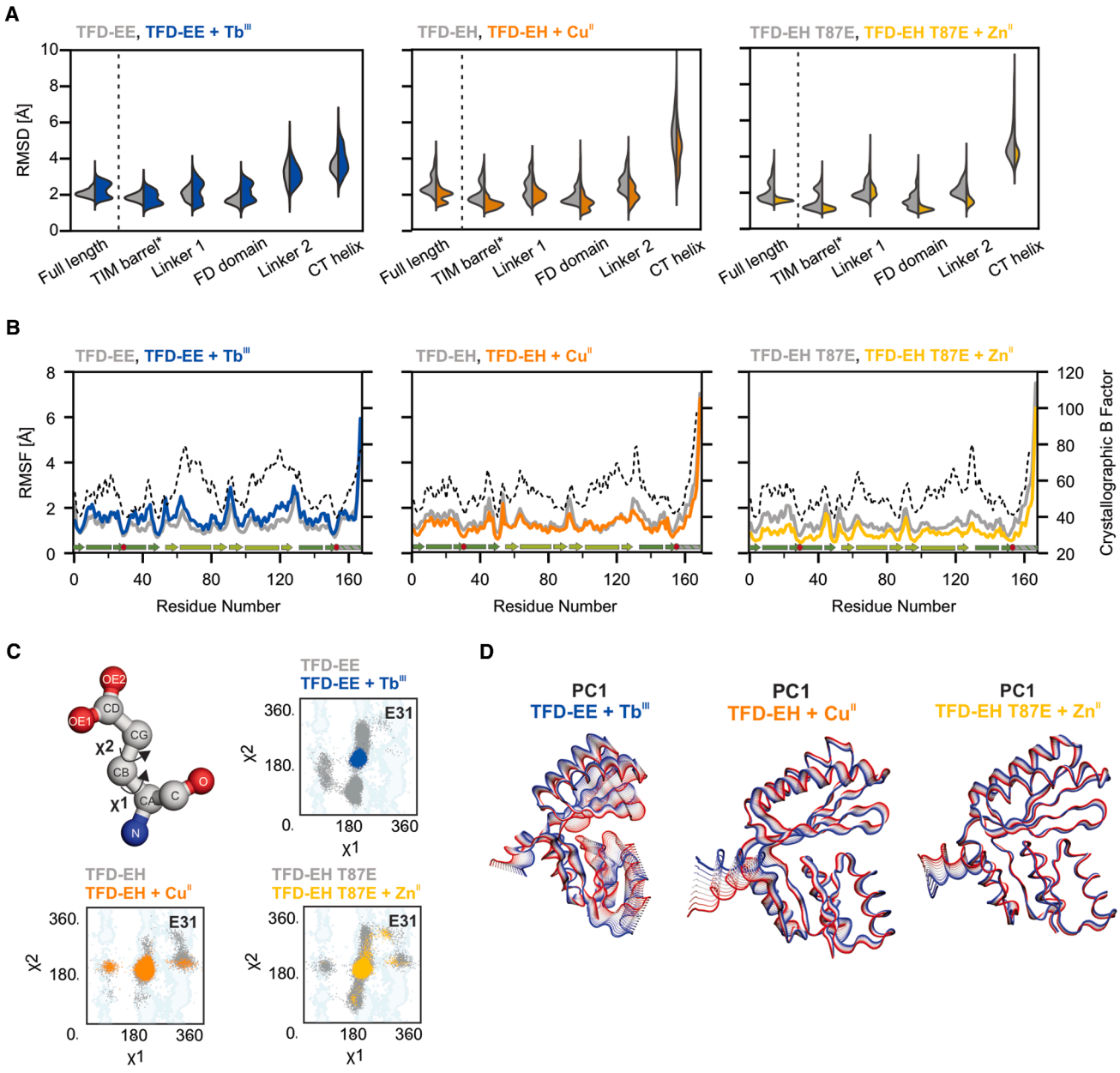


Figure 3. Molecular dynamics (MD) simulations

(A) Violin plots of RMSD values obtained from 3 μs simulations of the metal-free (gray) and metal-bound (blue, orange, yellow) TFD variants for the full-length proteins as well as separate regions as defined in Figure 1B. Here, TIM barrel* refers to the TIM barrel without the C-terminal helix, which was independently analyzed as CT helix.

(B) RMSF values (left axis) of metal-free (gray) and metal-bound (blue, orange, yellow) TFD variants, as well as experimental B-factors of the respective crystal structures (dashed black line, right axis), plotted against the amino acid sequence.

(C) Janin plots showing the distribution of side-chain dihedral angles χ1 and χ2 of the metal-coordinating glutamate E31 in the metal-free (gray) and metal-bound (blue, orange, yellow) state.

(D) Structural models from representative frames of pseudo-trajectories extracted from a principal component analysis (PCA) of the MD simulations.

(see also Figure S12) is in agreement with NMR relaxation data (Figures 2C and 2D), RMSF values and crystallographic B-factors (Figure 3B), consistently showing that the tri-glycine linkers and the C-terminal helix of the TIM barrel display high mobility. Global dynamics were strongly affected in the TFD-EH variants: metal binding decreased RMSD and RMSF values across nearly all regions of the protein, with TFD-EH T87E exhibiting the lowest overall flexibility. We concluded that the additional glutamate in TFD-EH T87E serves as an anchor that stabilizes the relative arrangement of the TIM barrel and FD domains in the presence of metal. Examination of side-chain dihedral angle distributions for the metal-coordinating residues showed

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

a localized reduction in conformational dynamics upon metal binding, most evident in the coordinating glutamates of TFD-EE (Figures 3C and S14). The MD simulations thus indicate that metal binding can modulate local and scaffold-wide protein dynamics.

We then performed a principal component analysis (PCA) to extract the key motions of the scaffold from the MD trajectories (Figures 3D, S15, and S16). We identified inter-domain motions mediated by the tri-glycine linkers and the flexible C-terminus, which seem to act as hinge, suggesting that the modular architecture can indeed enable conformational flexibility in an otherwise rigid *de novo* protein (Video S4). Furthermore, we measured NMR pseudo-contact shifts (PCS) for TFD-EE to provide experimental support for these conformational variations. Here, we exploited the paramagnetic properties of the lanthanides, which are associated with long-range distance- and orientation-dependent changes on the NMR chemical shifts.[51,52] As the metal binding site is located in the TIM barrel, relative domain motions will cause averaging of the PCS in the FD domain. We explored a minimal ensemble of conformations extracted from the MD simulations by calculating ensemble-averaged PCS depending on the ensemble size and compared these with the experimental PCS (Figure S10). This analysis showed that an ensemble of three structures of representative MD frames from the PCA analysis can reconcile the experimental PCS data. The main structural differences between selected frames reside in the relative position of the FD domain to the metal (Figure S10). As the two domains are connected by four symmetrically spaced linkers, the FD domain cannot move completely independently and only modest differences in the respective PCS were thus expected. The combined analysis of NMR data and MD simulations indicates that TFD is not a rigid protein scaffold. Its two-domain architecture with flexible linkers enables dynamic motions, which may be further tuned by mutagenesis to support catalysis.

TFD-EE is trapped in an undesired conformational state

We previously engineered the lanthanide-binding TFD-EE scaffold for cerium photoredox catalysis and noticed that efficient metal binding and photoenzymatic catalysis required a pre-treatment of the protein at elevated temperature (T > 40°C).[45] As this phenomenon may also be linked to conformational dynamics, we set out to investigate it further. While TFD-QQ, TFD-EH, and TFD-EH T87E showed well-resolved ^{1}H-^{15}N HSQC spectra in the presence and absence of metal, strong line broadening was observed for metal-free TFD-EE at 25°C (Figure 4A). Single-peak analysis using the tryptophan side chain resonances indicated a conformational equilibrium between two states shifting slowly in response to heating or metal binding. The metal-free protein had to be pre-incubated at 40°C to obtain well-resolved resonances at 25°C (Figure 4B). Upon addition of La^{III} at 25°C, the peak intensities increased up to 5-fold, albeit with very slow kinetics (Figure 4C).

To examine whether this phenomenon resulted from global unfolding, we recorded circular dichroism (CD) spectra at 25°C and 40°C, with and without metal, for each TFD variant. The spectra were very similar in all cases (Figure S2), indicating only minimal changes in the secondary structure. Still, local unfolding or destabilization of metal-free TFD-EE was plausible, as four negatively charged glutamates in the metal-binding site could repel each other in the absence of lanthanide. We thus calculated the contributions of all interface residues to the free energy of dimer formation based on the available MD simulations (Figures 4D and S17). The two pairs of glutamates in the binding site of TFD-EE add a strongly positive ΔG value in the absence of metal. The E154H mutation in TFD-EH, however, inverts this destabilizing effect. Analytical size-exclusion chromatography (SEC) coupled with static light scattering (SLS) detection confirmed these observations (Figures 4E and S4). Under all tested conditions (25°C and 40°C, with and without metal), TFD-QQ, TFD-EH, and TFD-EH T87E eluted primarily as a single peak of the expected dimer size (~38 kDa). For TFD-EE, however, we observed a more complex elution profile. In addition to a small fraction of oligomers, TFD-EE is present in an equilibrium of two dimeric species with distinct elution volumes. The earlier eluting dimer, suggesting a larger hydrodynamic radius, is present at low temperatures and in the absence of metal, while the latter dominates at high temperatures and in the presence of metal.

To study the TIM barrel in isolation, we removed the FD domain from the TFD-EE construct and directly closed the connecting loop from residues 54 to 130. This much smaller dimeric protein, named TIM-EE, showed a very similar behavior. Only after a pre-heating step and the addition of La^{III}, a well-resolved HSQC was obtained (Figure S9). SEC-SLS data confirmed the presence of two distinct dimeric species with variable distribution depending on temperature and lanthanide binding (Figure 4F).

We concluded that TFD-EE is trapped in an undesired conformational state (Figure 4G), which is still dimeric but associated with a locally destabilized dimer interface that impairs metal binding at low temperatures. Although heat treatment partially resolved this problem and accelerated metal binding, we wanted to overcome this limitation more generally by optimizing the scaffold's sequence while maintaining its modular architecture.

AI-guided sequence redesign accelerates metal binding and catalysis

The deep learning-based protein design tool *ProteinMPNN* has previously been applied to rescue failed designs that would not express or fold into the desired structure.[7] Here, we tested whether sequence redesign by *ProteinMPNN* could also shift a conformational equilibrium by stabilizing the desired metal binding-competent state of the TFD scaffold. We used the TFD-EE homodimer backbone as the input, but left functionally critical elements unchanged: the tryptophan antenna (W6), the metal-coordinating glutamates (E31, E154), and the tri-glycine linkers (G53-55 and G129-131) (Figure 5A). The redesigns were filtered for substrate pocket accessibility, low overall negative surface charge, and high-confidence structure prediction by AlphaFold. We also rationally modified selected surface residues to minimize nonspecific lanthanide binding. The most promising redesigned variant, TFD-EE MPNN, was tested experimentally.

TFD-EE MPNN has 48.8% sequence identity to the original TFD-EE (Figures 5A and S19). It expressed solubly, formed stable dimers, and had highly similar CD spectra compared to the parent scaffold (Figure S3). We obtained well-diffracting crystals

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

CellPress
OPEN ACCESS

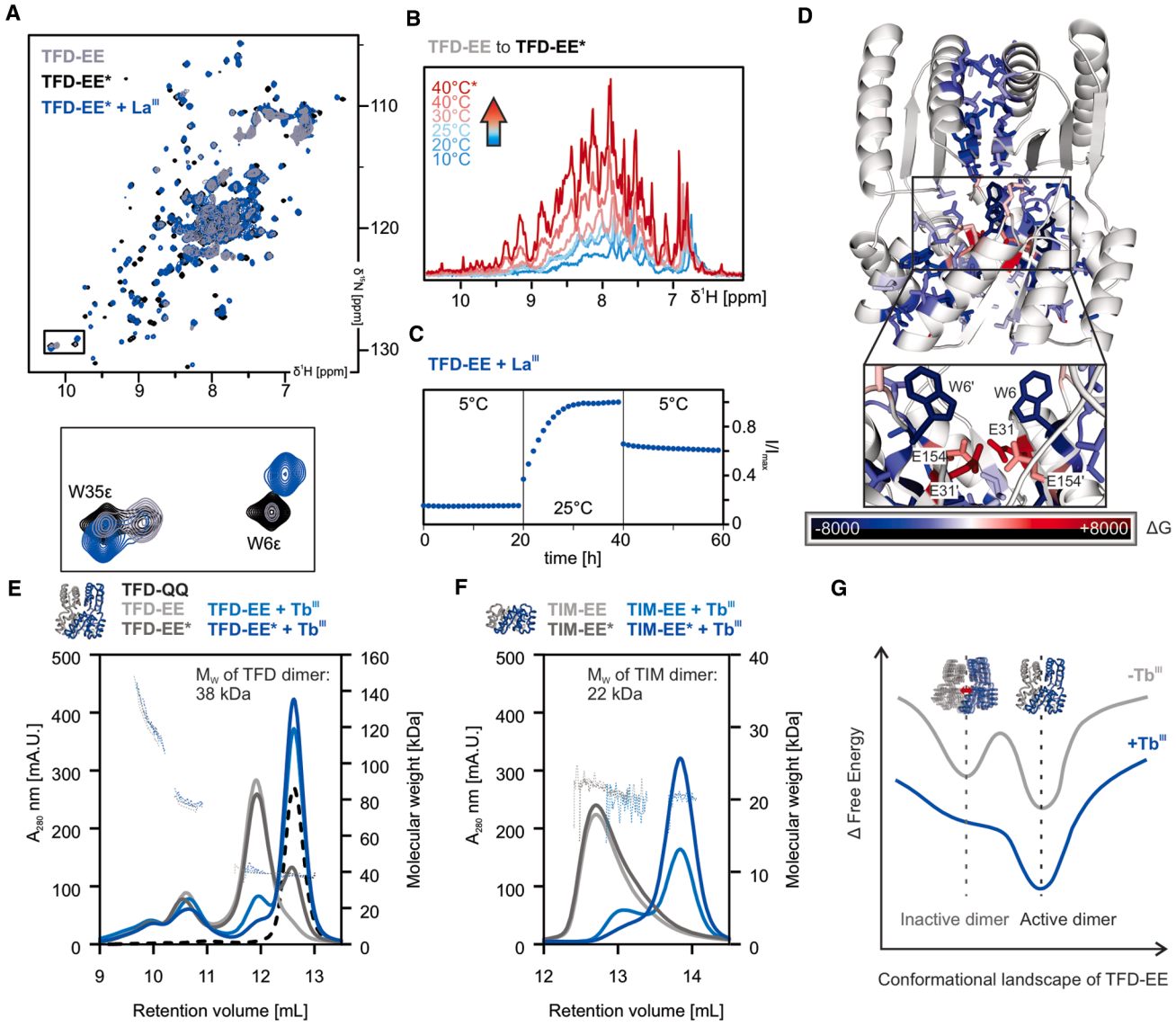


Figure 4. Conformational equilibria of TFD-EE

(A) ^{1}H-^{15}N HSQC spectra of TFD-EE, without (gray) and with (black) sample incubation at 40°C prior to measurement, and after the addition of 1 eq $LaCl_3$ (blue). A close-up view of the tryptophan side chain signals is shown below.

(B) The temperature dependence of the metal-free TFD-EE signal is shown in a ^{1}H chemical shift projection extracted from the 1st FID of ^{1}H-^{15}N HSQCs, highlighting the initially very pronounced line broadening in the metal-free sample.

(C) Time course of the signal intensity of a reference peak, normalized to its highest value; from the ^{1}H-^{15}N HSQC spectra of TFD-EE after addition of $LaCl_3$ and with varying temperatures.

(D) Contributions of the interface residues to the free energy of dimer formation were calculated from the MD simulations of TFD-EE and plotted on the structure, with ΔG>0 in red and ΔG<0 in blue.

(E) Size-exclusion chromatography with static light scattering detection (SEC-SLS) for TFD-EE in the absence (gray) and presence (blue) of $TbCl_3$ at 4°C and 25°C. The TFD-QQ control shows the same elution profile under all conditions and is shown for comparison (black line). The molecular weight of the eluted species is plotted on the right axis.

(F) SEC-SLS data of TIM-EE.

(G) Schematic energy landscape illustrating conformational equilibria of TFD-EE.

in complex with Tb^{III}, and determined its structure at 1.95 Å resolution (PDB code: 9QUP). With an all-atom RMSD of 0.5 Å, the experimental structure is in excellent agreement with its respective AlphaFold3 prediction (Figure 5A). The main deviations relate to the conformation of the outer tri-glycine linkers. MD simulations of the MPNN redesign showed stable structures that maintained flexibility through their linker regions (Figures 5B, S11, S12, and S13, Video S5). Lanthanide binding was confirmed experimentally by tryptophan-enhanced Tb^{III} luminescence. Titrations revealed a low nanomolar K_D value for Tb^{III} binding to

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

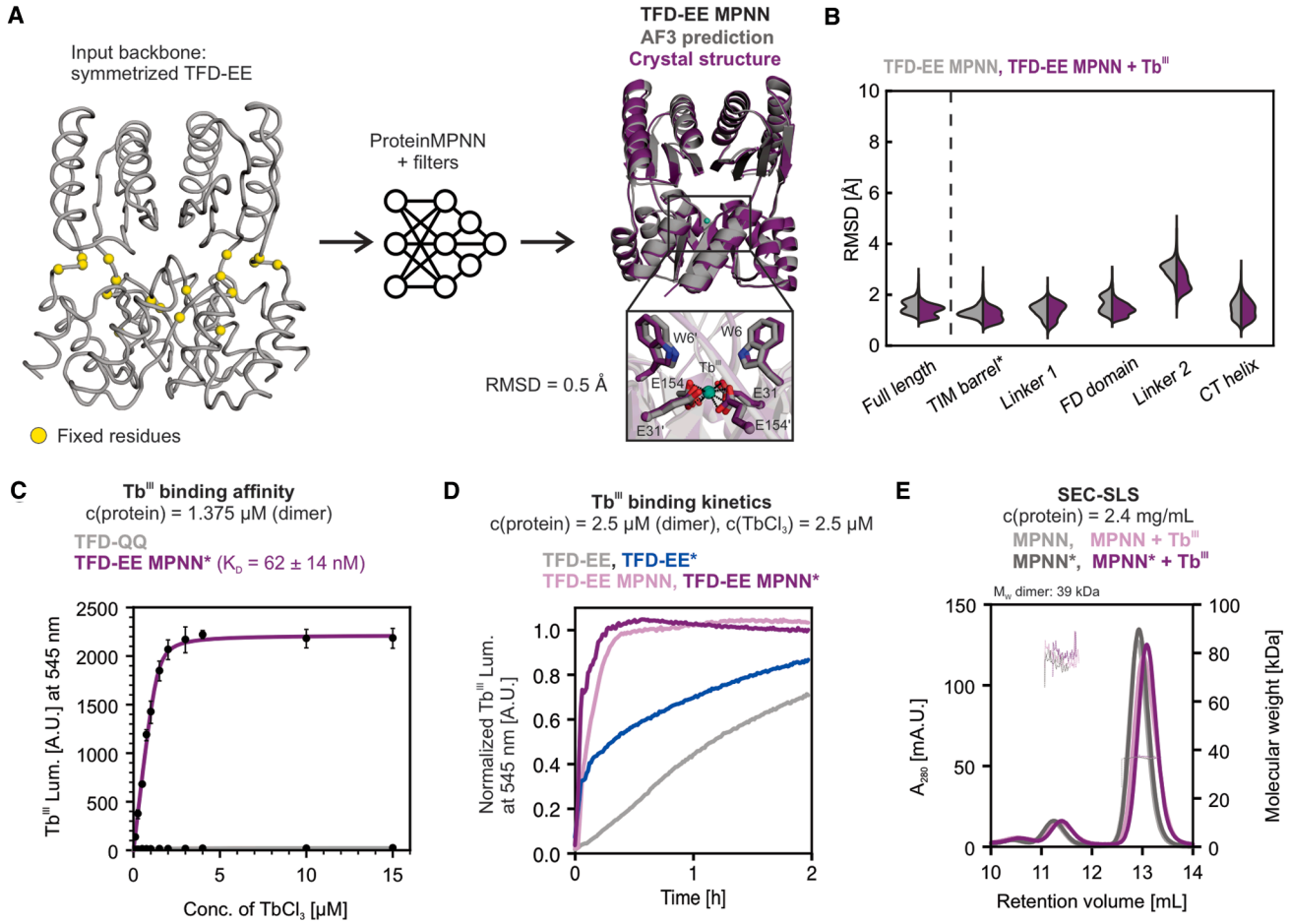


Figure 5. Characterization of the TFD-EE redesign

(A) Sequence redesign using *ProteinMPNN* and superposition of the crystal structure and AF3 model of Tb^{III}-bound TFD-EE MPNN.

(B) Comparative violin plots showing the overall RMSD distributions from MD simulations of metal-free (gray) and metal-bound (purple) TFD-EE MPNN.

(C) Tb^{III} binding to TFD-EE MPNN (purple), with TFD-QQ (gray) as control. Titration curves obtained for 1.375 µM protein dimer were fitted with a quadratic equation. Measurements were done in triplicate, and their standard deviations are shown as error bars in the plot.

(D) Kinetics of Tb^{III} binding to TFD-EE MPNN (purple) in comparison to the original TFD-EE (gray and blue) at room temperature and 40°C (indicated by *), respectively.

(E) SEC-SLS data of TFD-EE MPNN with TbCl$_3$ (light purple: 25°C, dark purple: 40°C) and without TbCl$_3$ (light gray: 25°C, dark gray: 40°C). The molecular weight of the eluted species is plotted on the right axis.

TFD-EE MPNN (Figure 5C), and the lanthanide binding kinetics were significantly accelerated (Figure 5D). Importantly, only a single dimeric species was observed in SEC-SLS runs under all conditions (Figure 5E). These findings confirmed that the *ProteinMPNN* redesign indeed stabilized the desired scaffold conformation.

To evaluate if the redesign's improved conformational properties also resulted in more efficient catalysis, we measured its photoenzymatic activity in comparison to the original TFD-EE. To that end, we chose the cerium(III/IV)-dependent radical C-C bond cleavage of hydrobenzoin to benzaldehyde, a photocatalytic reaction that we recently reported for this lanthanide-binding protein.[45] The diol cleavage proceeds upon visible-light irradiation, which triggers ligand-to-metal charge transfer followed by alkoxy radical formation and β-scission (Figure S18). Michaelis Menten kinetics revealed a 10-fold improved catalytic efficiency of TFD-EE MPNN, originating from both higher k_{cat} and lower K_M values compared to the original scaffold (Figure 6A).

To rationalize how the AI-guided sequence redesign enabled such improvements, we further characterized and analyzed the MPNN variant. Well-resolved ^{1}H-^{15}N HSQC spectra at 25°C were obtained in both metal-free and metal-bound states (Figure 6B). The absence of any temperature-dependent line broadening confirmed that the unfavorable conformational state was no longer present. A detailed structural analysis of the dimer interfaces showed additional stabilizing interactions that likely compensate for the charge frustration caused by the four glutamates in TFD-EE's metal-binding site (Figure 6C). In TFD-EE MPNN, enhanced hydrophobic packing in the TIM barrel core is facilitated by inter-domain π–π stacking of W152. In the solvent-exposed FD domain interface, residues R65, R69, K73, D77, and E81 form a zigzag of alternating charges across

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

CellPress
OPEN ACCESS

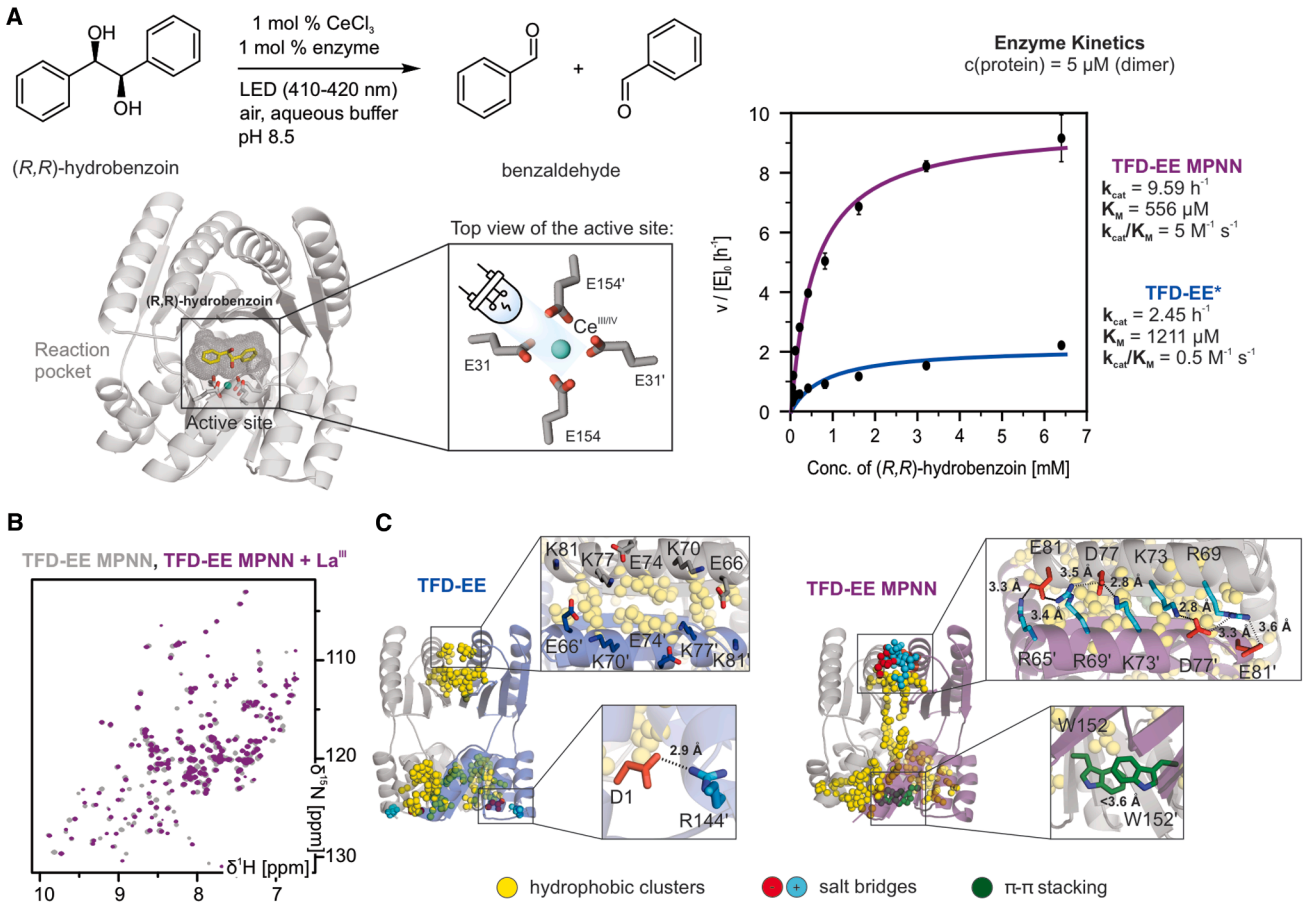


Figure 6. Lanthanide-dependent photoenzymatic activity and dimer interface stabilization

(A) Michaelis-Menten kinetics of the radical diol cleavage of (*R,R*)-hydrobenzoin to benzaldehyde catalyzed by TFD-EE (blue) and TFD-EE MPNN (purple) in complex with $CeCl_3$ upon visible-light irradiation. Measurements were done in triplicate, and their standard deviations are shown as error bars in the plot.

(B) ^{1}H-^{15}N HSQC spectra of TFD-EE MPNN before (gray) and after the addition of 1 eq. of $LaCl_3$ (purple).

(C) Structural analysis highlighting intermolecular protein-protein interactions in TFD-EE (blue) and TFD-EE MPNN (purple). Insets show close-up views of the homodimer interface. Color-coding: hydrophobic clusters (yellow), salt bridges (red and cyan), and π–π stacking (green). Protein interactions were calculated with *ProtInter*[62] and *ProteinTools*.[63]

subunits, creating additional stabilizing salt bridges. These interface interactions were consistent across the design model, crystal structure, and MD simulations. Overall, our results show that AI-guided sequence redesign can alleviate conformational constraints in this *de novo* enzyme by providing additional stabilizing interactions.

DISCUSSION

There is a high demand for robust and efficient enzymes to catalyze synthetically challenging reactions and fulfill sustainability goals in industrial processes.[53] Artificial metalloenzymes based on *de novo* protein scaffolds have great potential for addressing these needs, but further development is crucial.[21] Here, we used the TFD scaffold as a model system to show that both rational and AI-guided redesign can tune metal binding properties, conformational landscapes, and catalytic parameters in modular *de novo* proteins. We propose that such multi-domain scaffold architectures with flexible linker regions enable conformational dynamics in otherwise hyperstable *de novo* proteins. While these motions have not been explicitly targeted in the original design, they can likely be tailored by mutagenesis. One of the primary aims for future *de novo* enzyme development should be to incorporate protein dynamics early in the design process.[54,55] Deep learning tools currently being extended to predict[56,57] and control[58–60] different conformational states in proteins represent exciting milestones toward this goal.

ProteinMPNN, a deep learning-based tool for sequence design,[7] has already been shown to enhance expression, stability, and function of natural proteins, including a TEV protease variant with higher activity.[61] Here, we demonstrated that *ProteinMPNN* can also optimize an unfavorable conformational equilibrium in the well-expressed, folded, and highly stable TFD scaffold. The redesigned sequence encodes additional stabilizing interactions that counterbalance the charge frustration caused by the designed metal binding site in the homodimer interface. This stabilization of a desired, functional conformation

had a positive impact on metal binding kinetics and ultimately catalysis. It can be speculated that similar mutations could also arise through laboratory evolution, albeit at much higher experimental costs. Our results thus illustrate how AI-guided redesign could serve as a corrective strategy for conformational limitations in *de novo* designed enzymes with multi-domain architecture.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Prof. Dr. Cathleen Zeymer (cathleen.zeymer@tum.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Crystal structures have been deposited in the Protein Data Bank (PDB: 9QUC, 9QUD, 9QUI, 9QUL, 9QUO, 9QUP) and are publicly available as of the date of publication.
- NMR backbone assignments have been deposited in the Biological Magnetic Resonance Data Bank (BMRB: 53123, 53124) and are publicly available as of the date of publication.
- All other raw data associated with this work have been deposited at university's research data repository "mediaTUM" and are publicly available as of the date of publication at https://doi.org/10.14459/2025mp1798728.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was funded by the German Research Foundation (DFG) in the framework of SFB1035 (201302640, project A14). We further acknowledge support by an individual DFG grant (453748800) and the ERC Starting Grant "PhotoLanZyme" (101039592). A.H. received a PhD fellowship from "Stiftung der Deutschen Wirtschaft" (sdw). We are grateful to Astrid König, Annika Elimelech, Sam Asami and Gerd Gemmecker for technical support with protein crystallization, SEC-SLS measurements, and NMR experiments, respectively. Generous NMR measurement time at the Bavarian NMR Center is gratefully acknowledged. We thank Martin Zacharias and his group for helpful discussions and access to the in-house cluster to run MD simulations. The cluster infrastructure was funded by the German Research Foundation (DFG) via a Major Research Instrumentation Grant (439143547). Furthermore, we thank Zhen Li and Majid Jafari from the Merz group at Michigan State University for providing optimized LJ1264 parameters for zinc and copper ions prior to their publication. We acknowledge the staff members of the synchrotron beamlines X06SA at SLS (Villigen, Switzerland) as well as P13 of PETRA III at DESY (Hamburg, Germany, proposal number MX-970).

AUTHOR CONTRIBUTIONS

Conceptualization, P.W.E., F.D., M.G., M.S., and C.Z; Methodology, P.W.E., F.D., G.M., F.L.M., L.K., T.B., A.H., and I.W; Investigation, P.W.E., F.D., G.M., F.L.M., L.K., T.B., A.H., and I.W; Writing—original draft, C.Z., P.W.E., and F.D.; Writing—review & editing, C.Z., P.W.E., F.D., M.S., and M.G.; Funding acquisition, C.Z., M.S., and M.G.; Resources, C.Z., M.S., and M.G.; Supervision, M.G., M.S., and C.Z.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the author PWE used ChatGPT Plus (models 4o and o4-mini-high) in order to develop *Python* scripts for data analysis. After using this tool, PWE reviewed and edited the content as needed and takes full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Molecular cloning, recombinant expression, and protein purification
  - Metal binding studies based on tryptophan-enhanced Tb^{III} luminescence
  - Analytical size exclusion chromatography with static light scattering detection (SEC-SLS)
  - Circular dichroism spectroscopy
  - X-ray crystallography
  - NMR spectroscopy
  - Molecular dynamics (MD) simulations
  - Sequence redesign using *ProteinMPNN*
  - Protein intermolecular interaction analysis
  - Photoenzymatic activity
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2025.10.010.

Received: June 11, 2025
Revised: September 11, 2025
Accepted: October 10, 2025

REFERENCES

1. Kortemme, T. (2024). De novo protein design—From new structures to programmable functions. Cell *187*, 526–544.
2. Chu, A.E., Lu, T., and Huang, P.-S. (2024). Sparks of function by de novo protein design. Nat. Biotechnol. *42*, 203–215.
3. Korendovych, I.V., and DeGrado, W.F. (2020). De novo protein design, a retrospective. Q. Rev. Biophys. *53*, e3.
4. 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.
5. Wohlwend, J., Corso, G., Passaro, S., Getz, N., Reveiz, M., Leidal, K., Swiderski, W., Atkinson, L., Portnoi, T., Chinn, I., et al. (2025). Boltz-1 Democratizing Biomolecular Interaction Modeling. Preprint at bioRxiv. https://doi.org/10.1101/2024.11.19.624167.
6. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500.
7. Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky, B.I.M., Courbet, A., de Haas, R.J., Bethel, N., et al. (2022). Robust deep learning–based protein sequence design using ProteinMPNN. Science *378*, 49–56.
8. Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R.J., Milles, L.F., et al. (2023). De

novo design of protein structure and function with RFdiffusion. Nature *620*, 1089–1100.

9. Anishchenko, I., Pellock, S.J., Chidyausiku, T.M., Ramelot, T.A., Ovchinnikov, S., Hao, J., Bafna, K., Norn, C., Kang, A., Bera, A.K., et al. (2021). De novo protein design by deep network hallucination. Nature *600*, 547–552.
10. Ferruz, N., Schmidt, S., and Höcker, B. (2022). ProtGPT2 is a deep unsupervised language model for protein design. Nat. Commun. *13*, 4348.
11. Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G.R., Morey-Burrows, F.S., Anishchenko, I., Humphreys, I.R., et al. (2024). Generalized biomolecular modeling and design with RoseTTAFold All-Atom. Science *384*, eadl2528.
12. Silva, D.A., Yu, S., Ulge, U.Y., Spangler, J.B., Jude, K.M., Labão-Almeida, C., Ali, L.R., Quijano-Rubio, A., Ruterbusch, M., Leung, I., et al. (2019). De novo design of potent and selective mimics of IL-2 and IL-15. Nature *565*, 186–191.
13. Bennett, N.R., Coventry, B., Goreshnik, I., Huang, B., Allen, A., Vafeados, D., Peng, Y.P., Dauparas, J., Baek, M., Stewart, L., et al. (2023). Improving de novo protein binder design with deep learning. Nat. Commun. *14*, 2625.
14. Buller, R., Damborsky, J., Hilvert, D., and Bornscheuer, U.T. (2025). Structure Prediction and Computational Protein Design for Efficient Biocatalysts and Bioactive Proteins. Angew. Chem. Int. Ed. *64*, e202421686.
15. Burton, A.J., Thomson, A.R., Dawson, W.M., Brady, R.L., and Woolfson, D.N. (2016). Installing hydrolytic activity into a completely de novo protein framework. Nat. Chem. *8*, 837–844.
16. Yeh, A.H.-W., Norn, C., Kipnis, Y., Tischer, D., Pellock, S.J., Evans, D., Ma, P., Lee, G.R., Zhang, J.Z., Anishchenko, I., et al. (2023). De novo design of luciferases using deep learning. Nature *614*, 774–780.
17. Romero-Romero, S., Braun, A.E., Kossendey, T., Ferruz, N., Schmidt, S., and Höcker, B. (2024). De novo design of triosephosphate isomerases using generative language models preprint at bioRxiv. https://doi.org/10.1101/2024.11.10.622869.
18. Lauko, A., Pellock, S.J., Sumida, K.H., Anishchenko, I., Juergens, D., Ahern, W., Jeung, J., Shida, A.F., Hunt, A., Kalvet, I., et al. (2025). Computational design of serine hydrolases. Science *388*, eadu2454.
19. Klein, A.S., and Zeymer, C. (2021). Design and engineering of artificial metalloproteins: from de novo metal coordination to catalysis. Protein Eng. Des. Sel. *34*, gzab003.
20. Chalkley, M.J., Mann, S.I., and DeGrado, W.F. (2022). De novo metalloprotein design. Nat. Rev. Chem *6*, 31–50.
21. Vornholt, T., Leiss-Maier, F., Jeong, W.J., Zeymer, C., Song, W.J., Roelfes, G., and Ward, T.R. (2024). Artificial metalloenzymes. Nat. Rev. Methods Primers *4*, 78.
22. Handel, T.M., Williams, S.A., and DeGrado, W.F. (1993). Metal Ion-Dependent Modulation of the Dynamics of a Designed Protein. Science *261*, 879–885.
23. Studer, S., Hansen, D.A., Pianowski, Z.L., Mittl, P.R.E., Debon, A., Guffy, S.L., Der, B.S., Kuhlman, B., and Hilvert, D. (2018). Evolution of a highly active and enantiospecific metalloenzyme from short peptides. Science *362*, 1285–1288.
24. Kim, D., Woodbury, S.M., Ahern, W., Tischer, D., Hanikel, N., Salike, S., Yim, J., Pellock, S.J., Lauko, A., Kalvet, I., et al. (2024). Computational Design of Metallohydrolases preprint at bioRxiv. https://doi.org/10.1101/2024.11.13.623507.
25. Stenner, R., Steventon, J.W., Seddon, A., and Anderson, J.L.R. (2020). A de novo peroxidase is also a promiscuous yet stereoselective carbene transferase. Proc. Natl. Acad. Sci. USA *117*, 1419–1428.
26. Kalvet, I., Ortmayer, M., Zhao, J., Crawshaw, R., Ennist, N.M., Levy, C., Roy, A., Green, A.P., and Baker, D. (2023). Design of Heme Enzymes with a Tunable Substrate Binding Pocket Adjacent to an Open Metal Coordination Site. J. Am. Chem. Soc. *145*, 14307–14315.
27. Lombardi, A., Pirro, F., Maglio, O., Chino, M., and DeGrado, W.F. (2019). De Novo Design of Four-Helix Bundle Metalloproteins: One Scaffold, Diverse Reactivities. Acc. Chem. Res. *52*, 1148–1159.
28. Basler, S., Studer, S., Zou, Y., Mori, T., Ota, Y., Camus, A., Bunzel, H.A., Helgeson, R.C., Houk, K.N., Jiménez-Osés, G., and Hilvert, D. (2021). Efficient Lewis acid catalysis of an abiological reaction in a de novo protein scaffold. Nat. Chem. *13*, 231–235.
29. Wang, Y., Xue, P., Cao, M., Yu, T., Lane, S.T., and Zhao, H. (2021). Directed Evolution: Methodologies and Applications. Chem. Rev. *121*, 12384–12444.
30. Childers, M.C., and Daggett, V. (2017). Insights from molecular dynamics simulations for computational protein design. Mol. Syst. Des. Eng. *2*, 9–33.
31. Davey, J.A., Damry, A.M., Goto, N.K., and Chica, R.A. (2017). Rational design of proteins that exchange on functional timescales. Nat. Chem. Biol. *13*, 1280–1285.
32. Gonzalez, N.A., Li, B.A., and McCully, M.E. (2022). The stability and dynamics of computationally designed proteins. Protein Eng. Des. Sel. *35*, gzac001.
33. Eisenmesser, E.Z., Millet, O., Labeikovsky, W., Korzhnev, D.M., Wolf-Watz, M., Bosco, D.A., Skalicky, J.J., Kay, L.E., and Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. Nature *438*, 117–121.
34. Henzler-Wildman, K.A., Lei, M., Thai, V., Kerns, S.J., Karplus, M., and Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature *450*, 913–916.
35. Lindorff-Larsen, K., Best, R.B., DePristo, M.A., Dobson, C.M., and Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. Nature *433*, 128–132.
36. Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., Eastwood, M.P., Bank, J.A., Jumper, J.M., Salmon, J.K., Shan, Y., and Wriggers, W. (2010). Atomic-Level Characterization of the Structural Dynamics of Proteins. Science *330*, 341–346.
37. Tokuriki, N., and Tawfik, D.S. (2009). Protein Dynamism and Evolvability. Science *324*, 203–207.
38. Hanoian, P., Liu, C.T., Hammes-Schiffer, S., and Benkovic, S. (2015). Perspectives on Electrostatics and Conformational Motions in Enzyme Catalysis. Acc. Chem. Res. *48*, 482–489.
39. Maria-Solano, M.A., Serrano-Hervás, E., Romero-Rivera, A., Iglesias-Fernández, J., and Osuna, S. (2018). Role of conformational dynamics in the evolution of novel enzyme function. Chem. Commun. *54*, 6622–6634.
40. Otten, R., Pádua, R.A.P., Bunzel, H.A., Nguyen, V., Pitsawong, W., Patterson, M., Sui, S., Perry, S.L., Cohen, A.E., Hilvert, D., and Kern, D. (2020). How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. Science *370*, 1442–1446.
41. Corbella, M., Pinto, G.P., and Kamerlin, S.C.L. (2023). Loop dynamics and the evolution of enzyme activity. Nat. Rev. Chem *7*, 536–547.
42. Huang, P.S., Feldmeier, K., Parmeggiani, F., Velasco, D.A.F., Höcker, B., and Baker, D. (2016). De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. Nat. Chem. Biol. *12*, 29–34.
43. Caldwell, S.J., Haydon, I.C., Piperidou, N., Huang, P.S., Bick, M.J., Sjöström, H.S., Hilvert, D., Baker, D., and Zeymer, C. (2020). Tight and specific lanthanide binding in a de novo TIM barrel with a large internal cavity designed by symmetric domain fusion. Proc. Natl. Acad. Sci. USA *117*, 30362–30369.
44. Mattocks, J.A., Tirsch, J.L., and Cotruvo, J.A., Jr. (2021). Determination of affinities of lanthanide-binding proteins using chelator-buffered titrations. Methods Enzymol. *651*, 23–61.
45. Klein, A.S., Leiss-Maier, F., Mühlhofer, R., Boesen, B., Mustafa, G., Kugler, H., and Zeymer, C. (2024). A De Novo Metalloenzyme for Cerium Photoredox Catalysis. J. Am. Chem. Soc. *146*, 25976–25985.
46. Olsson, M.H.M., Søndergaard, C.R., Rostkowski, M., and Jensen, J.H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. J. Chem. Theory Comput. *7*, 525–537.

47. Irving, H., and Williams, R.J.P. (1948). Order of Stability of Metal Complexes. Nature *162*, 746–747.

48. Trupp, L., Bruttomesso, A.C., Eliseeva, S.V., Petoud, S., Ramírez, J.A., and Barja, B.C. (2020). A Six-Armed Phenhomazine Ligand with a Potential "Turn-Off" Copper(II) Sensing Capability through Terbium(III) Luminescence Quenching. Chem. Eur J. *26*, 12645–12653.

49. Jafari, M., Li, Z., Song, L.F., Sagresti, L., Brancato, G., and Merz, K.M., Jr. (2024). Thermodynamics of Metal–Acetate Interactions. J. Phys. Chem. B *128*, 684–697.

50. Li, Z., Bhowmik, S., Sagresti, L., Brancato, G., Smith, M., Benson, D.E., Li, P., and Merz, K.M., Jr. (2024). Simulating Metal-Imidazole Complexes. J. Chem. Theory Comput. *20*, 6706–6716.

51. Müntener, T., Joss, D., Häussinger, D., and Hiller, S. (2022). Pseudocontact Shifts in Biomolecular NMR Spectroscopy. Chem. Rev. *122*, 9422–9467.

52. Otting, G. (2008). Prospects for lanthanides in structural biology by NMR. J. Biomol. NMR *42*, 1–9.

53. Buller, R., Lutz, S., Kazlauskas, R.J., Snajdrova, R., Moore, J.C., and Bornscheuer, U.T. (2023). From nature to industry: Harnessing enzymes for biocatalysis. Science *382*, eadh8615.

54. Casadevall, G., Duran, C., and Osuna, S. (2023). AlphaFold2 and Deep Learning for Elucidating Enzyme Conformational Flexibility and Its Application for Design. JACS Au *3*, 1554–1562.

55. Kouba, P., Planas-Iglesias, J., Damborský, J., Sedlar, J., Mazurenko, S., and Sivic, J. (2024). Learning to engineer protein flexibility. Preprint at arXiv. 2412.18275. https://doi.org/10.48550/arXiv.2412.18275.

56. Wayment-Steele, H.K., Ojoawo, A., Otten, R., Apitz, J.M., Pitsawong, W., Hömberger, M., Ovchinnikov, S., Colwell, L., and Kern, D. (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. Nature *625*, 832–839.

57. Sala, D., Engelberger, F., McHaourab, H.S., and Meiler, J. (2023). Modeling conformational states of proteins with AlphaFold. Curr. Opin. Struct. Biol. *81*, 102645.

58. Praetorius, F., Leung, P.J.Y., Tessmer, M.H., Broerman, A., Demakis, C., Dishman, A.F., Pillai, A., Idris, A., Juergens, D., Dauparas, J., et al. (2023). Design of stimulus-responsive two-state hinge proteins. Science *381*, 754–760.

59. Abrudan, A., Ojeda, S.P., Joshi, C.K., Greenig, M., Engelberger, F., Khmelinskaia, A., Meiler, J., Vendruscolo, M., and Knowles, T.P. (2025). Multi-state Protein Design with DynamicMPNN preprint at arXiv. 2507. 21938. https://doi.org/10.48550/arXiv.2507.21938.

60. Guo, A.B., Akpinaroglu, D., Stephens, C.A., Grabe, M., Smith, C.A., Kelly, M.J.S., and Kortemme, T. (2025). Deep learning–guided design of dynamic proteins. Science *388*, eadr7094.

61. Sumida, K.H., Núñez-Franco, R., Kalvet, I., Pellock, S.J., Wicky, B.I.M., Milles, L.F., Dauparas, J., Wang, J., Kipnis, Y., Jameson, N., et al. (2024). Improving Protein Expression, Stability, and Function with ProteinMPNN. J. Am. Chem. Soc. *146*, 2054–2061.

62. Borry, M., and Schmidt, A. (2025). Protinter 0.9.2 (0.9.2). Zenodo. https://doi.org/10.5281/zenodo.14851670.

63. Ferruz, N., Schmidt, S., and Höcker, B. (2021). ProteinTools: a toolkit to analyze protein structures. Nucleic Acids Res. *49*, W559–W566.

64. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

65. Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687–696.

66. Orton, H.W., Huber, T., and Otting, G. (2020). Paramagpy: software for fitting magnetic susceptibility tensors using paramagnetic effects measured in NMR spectra. Magn. Reson. *1*, 1–12.

67. Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. Curr. Protoc. Bioinform *54*, 5.6.1–5.6.37.

68. McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernández, C.X., Schwantes, C.R., Wang, L.-P., Lane, T.J., and Pande, V. S. (2015). MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys. J. *109*, 1528–1532.

69. Gowers, R.J., Linke, M., Barnoud, J., Reddy, T.J.E., Melo, M.N., Seyler, S. L., Domanski, J., Dotson, D.L., Buchoux, S., and Kenney, I.M. (2019). MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations (Los Alamos, NM (United States): Los Alamos National Laboratory (LANL)), pp. 2575–9752.

70. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., and Beckstein, O. (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem. *32*, 2319–2327.

71. Janin, J., Wodak, S., Levitt, M., and Maigret, B. (1978). Conformation of amino acid side-chains in proteins. J. Mol. Biol. *125*, 357–386.

72. Salomon-Ferrer, R., Case, D.A., and Walker, R.C. (2013). An overview of the Amber biomolecular simulation package. WIREs Comput. Mol. Sci. *3*, 198–210.

73. Case, D.A., Aktulga, H.M., Belfon, K., Cerutti, D.S., Cisneros, G.A., Cruzeiro, V.W.D., Forouzesh, N., Giese, T.J., Götz, A.W., Gohlke, H., et al. (2023). AmberTools. J. Chem. Inf. Model. *63*, 6183–6191.

74. Tian, C., Kasavajhala, K., Belfon, K.A.A., Raguette, L., Huang, H., Migues, A.N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., and Simmerling, C. (2020). ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. J. Chem. Theory Comput. *16*, 528–552.

75. Li, Z., Song, L.F., Li, P., and Merz, K.M., Jr. (2021). Parametrization of Trivalent and Tetravalent Metal Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. J. Chem. Theory Comput. *17*, 2342–2354.

76. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. J. Mol. Graph. *14*, 33.

77. Valdés-Tresanco, M.S., Valdés-Tresanco, M.E., Valiente, P.A., and Moreno, E. (2021). gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS. J. Chem. Theory Comput. *17*, 6281–6291.

78. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. *26*, 795–800.

79. Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

80. Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. Sect. D - Biol. Crystallogr. *67*, 355–367.

81. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

82. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. *27*, 293–315.

83. Sánchez-Aparicio, J.-E., Tiessler-Sala, L., Velasco-Carneros, L., Roldán-Martín, L., Sciortino, G., and Maréchal, J.-D. (2021). BioMetAll: Identifying Metal-Binding Sites in Proteins from Backbone Preorganization. J. Chem. Inf. Model. *61*, 311–323.

84. Schlee, S., and Reinstein, J. (2005). Characterization of ATPase Cycles of Molecular Chaperones by Fluorescence and Transient Kinetic Methods. In Protein Folding Handbook, J. Buchner and T. Kiefhaber, eds. (Wiley-VCH), pp. 105–161.

85. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

86. Turk, D. (2013). MAIN software for density averaging, model building, structure refinement and validation. Acta Crystallogr. D Biol. Crystallogr. *69*, 1342–1357.

87. Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. *3*, 1171–1179.

88. Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. Nucl. Magn. Reson. Spectrosc. *34*, 93–158.

89. Williamson, M.P. (2013). Using chemical shift perturbation to characterise ligand binding. Prog. Nucl. Magn. Reson. Spectrosc. *73*, 1–16.

90. Barbato, G., Ikura, M., Kay, L.E., Pastor, R.W., and Bax, A. (1992). Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. Biochemistry *31*, 5269–5278.

91. Dürr, S.L. (2023). ProteinMPNN Gradio Webapp (v0.3). Zenodo. https://doi.org/10.5281/zenodo.7630417.

92. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682.

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
$LaCl_3$	Sigma-Aldrich	Cat. # 262072
$TbCl_3$	Sigma-Aldrich	Cat. # 212903
$CeCl_3$	Sigma-Aldrich	Cat. # 228931
$YbCl_3$	Sigma-Aldrich	Cat. # 337927
$CoSO_4$	ROTH	Cat. # T889.2
$NiSO_4$	Sigma-Aldrich	Cat. # 227676
$ZnSO_4$	Sigma-Aldrich	Cat. # 793523
$CuCl_2$	Sigma-Aldrich	Cat. # 467847
(*R,R*)-hydrobenzoin	Sigma-Aldrich	Cat. # 256277
Critical commercial assays		
PCR Purification Kit	Jena Bioscience GmbH	Cat. # PP-201L
PureYield™ Plasmid Miniprep System	Promega	Cat. # A1223
Monarch DNA Gel Extraction Kit	New England Biolabs GmbH	Cat. # T1020
Deposited data		
TFD-EE with terbium(III) structure	Caldwell et al.[43] 2020	PDB: 6ZV9
TFD-EH apo structure	This paper	PDB: 9QUC
TFD-EH with copper(II) structure	This paper	PDB: 9QUD
TFD-EH with nickel(II) structure	This paper	PDB: 9QUI
TFD-EH T87E with zinc(II) structure	This paper	PDB: 9QUL
TFD-EH T87E with cobalt(II) structure	This paper	PDB: 9QUO
TFD-EE MPNN with terbium(III) structure	This paper	PDB: 9QUP
Backbone resonance assignments of TFD-EE with lanthanum(III)	This paper	BMRB: 53123
Backbone resonance assignments of TFD-EH apo	This paper	BMRB: 53124
15N relaxation data for TFD-EE and TFD-EH with and without metal	This paper	https://doi.org/10.14459/2025mp1798728
Other raw and analyzed data	This paper	https://doi.org/10.14459/2025mp1798728
Oligonucleotides		
Primers for mutagenesis, see Table S3; Sigma-Aldrich	This paper	N/A
Recombinant DNA		
Synthetic plasmids, see Section Methods S1; TWIST Bioscience	This paper	N/A
Software and algorithms		
NMRpipe	Delaglio et al.[64] 1995	https://www.ibbr.umd.edu/nmrpipe/install.html
CCPNMR 2.4	Vranken et al.[65] 200	http://www.ccpn.ac.uk/v2-software
paramagpy	Orton et al.[66] 2020	https://github.com/henryorton/paramagpy
MODELLER software	Webb & Sali[67] 2016	https://salilab.org/modeller
Python version 2.7	Python Software Foundation	https://www.python.org
MDTraj	McGibbon et al.[68] 2015	https://www.mdtraj.org/1.9.8.dev0/installation.html
Janin plots, part of the MDAnalysis package	Gowers et al., 2019; Michaud-Agrawal et al., 2011; Janin et al., 1978[69–71]	https://docs.mdanalysis.org/2.0.0/documentation_pages/analysis/dihedrals.html

(*Continued on next page*)

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AMBER software package	Salomon-Ferrer et al., 2013; Case et al., 2023[72,73]	https://ambermd.org
AMBER force field parameters for protein (ff19SB)	Tian et al.[74] 202	https://ambermd.org
AMBER force field parameters for OPC water model and metal ion parameters LJ1264	Li et al.[75] 2021	https://ambermd.org
AMBER force field default LJ1264 parameters for terbium (III)	Li et al.[75] 2021	https://ambermd.org
AMBER force field modified LJ1264 parameters for copper (II) and zinc (II)	Jafari et al., 2024; Li et al., 2024[49,50]	https://ambermd.org
AmberTools	Case et al.[73] 2023	https://ambermd.org/AmberTools.php
CCPTRAJ	McGibbon et al.[68] 2015	https://github.com/Amber-MD/cpptraj
VMD	Humphrey et al.[76] 1996	https://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=VMD
PyMOL	Schrodinger, 2015.	https://www.pymol.org
MDAnalysis	Michaud-Agrawal et al.[70] 2011	https://www.mdanalysis.org
GROMACS MMPBSA	Valdés-Tresanco et al.[77] 2021	https://valdes-tresanco-ms.github.io/gmx_MMPBSA/dev/
ProtInter	Borry & Schmidt[62] 2025	https://github.com/maxibor/protinter/tree/master
ProteinTools	Ferruz et al.[63] 2021	https://proteintools.uni-bayreuth.de
XDS	Kabsch[78,79] 2010	http://xds.mpimf-heidelberg.mpg.de/
AlphaFold 3	Abramson et al.[6] 2024	https://deepmind.google/science/alphafold/
REFMAC5	Murshudov et al.[80] 2004	http://www.ccp4.ac.uk/
Coot	Emsley et al.[81] 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
MolProbity online tool	Williams et al.[82] 2010	http://molprobity.biochem.duke.edu/
ProteinMPNN	Dauparas et al.[7] 2022	https://github.com/dauparas/ProteinMPNN
BioMetAll	Sánchez-Aparicio et al.[83] 2021	https://github.com/insilichem/biometall
ChemDraw Professional 18.0	PerkinElmer	https://www.perkinelmer.com/category/chemdraw
Other		
NiNTA resin	Qiagen	Cat. # 30210
Superdex 75 Increase 10/300 GL gel filtration column; Cytiva 29-1487-21	Sigma-Aldrich	Cat. # GE29-1487-21
Vanquish System HPLC	Thermo Fisher Scientific	N/A
Achiral Hypersil Gold C18 column	Thermo Fisher Scientific	Cat. # 25003-102130
Varioskan LUX plate reader with time-resolved fluorometry module	Thermo Fisher Scientific	Cat. # 15360777
Viscotek TDA 305 triple array detector	Malvern Instruments	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

E. coli strains DH10B and BL21(DE3) were cultivated in LB broth supplemented with antibiotics while shaking at 37°C.

METHOD DETAILS

Molecular cloning, recombinant expression, and protein purification

Synthetic genes were ordered from Twist Bioscience and subcloned into pET-M11 or pET29b(+) expression vectors. The sequences of all constructs, primers used to introduce point mutations, and the associated cloning protocols are available in the SI. The proteins were produced recombinantly in *E. coli* BL21(DE3) grown either in LB medium or M9 minimal medium, which was prepared with H_2O or D_2O and supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose for isotope labeling. All constructs contained a cleavable N-terminal His_6-tag and were purified by Ni-NTA affinity chromatography (washing buffer: 25 mM HEPES pH 7.5, 25 mM imidazole, 300 mM

NaCl; elution buffer: 25 mM HEPES pH 7.5, 300 mM imidazole, 300 mM NaCl) followed by TEV protease cleavage for affinity tag removal (dialysis buffer: 25 mM HEPES pH 7.5, 0.5 mM EDTA, 0.5 mM DTT, 300 mM NaCl) and subsequent size-exclusion chromatography (SEC) with a Superdex 75 column (SEC buffer: 25 mM HEPES pH 7.5, 100 mM NaCl). Residual metal ions bound to the proteins were removed by incubation in EDTA-containing buffer or dialysis in the presence of Chelex100 resin. Purity and integrity of all proteins were evaluated by SDS-PAGE, protein mass spectrometry, and CD spectroscopy. Protein concentrations were determined by absorption at 280 nm. Concentrated protein aliquots were flash-frozen and stored at -70°C in SEC buffer (25 mM HEPES pH 7.5, 100 mM NaCl). Detailed protocols are available in the SI (section Methods S1).

Metal binding studies based on tryptophan-enhanced Tb^{III} luminescence

Measurements were performed in black 96-well microtiter plates using a Varioskan LUX plate reader (Thermo Fisher Scientific) in time-resolved fluorescence mode (TRF settings: 75 μs delay time, 1 ms integration time, 250 ms measurement time). Upon excitation at 280 nm, the tryptophan-enhanced Tb^{III} luminescence signal was recorded at 545 nm emission wavelength.

Tb^{III} binding to TFD and the MPNN variant was measured in direct titrations at 1.25 μM and 1.375 μM dimeric protein, respectively, and $TbCl_3$ concentrations ranging from 0 to 70 μM in buffer containing 25 mM HEPES pH 7.0, 25 mM NaCl. Data points were recorded in triplicate after 3 hours of incubation. Due to mostly low K_D values in the range of c(protein) or lower, a quadratic binding equation was used for data fitting.

$$F = F_0 + F_{Ampl} \cdot \frac{\frac{[A_0]+[B_0]+K_{D_{AB}}}{2} - \sqrt{\left(\frac{[A_0]+[B_0]+K_{d_{AB}}}{2}\right)^2 - [A_0]\cdot[B_0]}}{[A_0]}$$

Tb^{III} binding kinetics were recorded at 10 μM dimeric protein and 10 μM $TbCl_3$ over the course of 10 hours.

Tb^{III} displacement titrations were performed for TFD-EH and TFD-EH T87E to estimate the relative binding affinities of divalent transition metal ions (Cu^{II}, Zn^{II}, Ni^{II}, Co^{II}). To that end, 1.25 μM dimeric protein was incubated with an excess of 100 μM $TbCl_3$ and then titrated with $CoSO_4$, $NiSO_4$, $ZnSO_4$, or $CuCl_2$ in concentrations ranging from 0 to 400 μM. Data points were recorded in triplicate after 3 hours of incubation. A cubic binding equation[84] was used for data fitting. However, the complexity of this particular metal displacement with a change in metal-to-protein stoichiometry impaired a quantitative analysis based on the available data.

Analytical size exclusion chromatography with static light scattering detection (SEC-SLS)

Static light scattering measurements were performed using a Viscotek TDA 305 triple array detector (Malvern Instruments) attached to an analytical size exclusion chromatography column. Protein samples were incubated overnight in the presence or absence of metal at different temperatures (4, 25, and 40 °C) and then run over a Superdex 75 column at a flow rate of 0.5 mL/min. All experiments were performed in SEC buffer (25 mM HEPES pH 7.0, 25 mM NaCl). The molecular masses of the eluted species were calculated from the refractive index and right-angle light scattering signals using the Omnisec software (Malvern Instruments). The SLS detector was calibrated with a bovine serum albumin (BSA) solution at 4 mg/ml using 66.4 kDa for the BSA monomer and a dn/dc value of 0.185 ml/g for all protein samples.

Circular dichroism spectroscopy

CD spectra over a wavelength range of 190-260 nm were recorded on a Chirascan-plus instrument (Applied Photophysics) or a Jasco J-815 spectropolarimeter using a quartz cuvette with 1 mm path length, and the following settings: 1.0 nm bandwidth, 1.0 nm step size, 0.5-1.0 s response time, and at least 3 accumulations. Protein samples were diluted to a concentration of 0.1 mg/mL in ddH_2O and measured at different temperatures. Metal-containing samples were incubated for 2 hours prior to measurement. The CD signals were recorded in millidegrees (m°) and subsequently converted to mean residue molar ellipticity (MRE) values.

X-ray crystallography

Protein crystallization

Crystals for the TFD-EH scaffolds were obtained by hanging drop vapor diffusion. Samples were prepared by incubating the purified proteins with 2-fold molar excess of the respective metal ion at 37°C overnight in 25 mM HEPES pH 7.0 and then concentrating to a final protein concentration of ca. 9 mg/mL. Hanging drops were set up by mixing the protein solutions and reservoir solutions as specified in Table S5. Crystals grew after 48 h equilibration against 300 μμL of reservoir solution. In preparation for data collection, the crystal of metal-free TFD-EH was cryoprotected by adding 5 μL of an 8:2 mixture of mother liquor and 100% (v/v) ethylene glycol. All other crystals were cryoprotected by adding 5 μL of a 7:3 mixture of mother liquor and 100% (v/v) glycerol. Cryoprotected crystals were vitrified in liquid nitrogen. TFD-EE MPNN was crystallized in sitting drop vapor diffusion experiments. Samples were prepared by incubating the purified protein with equimolar amounts of $TbCl_3$ at room temperature for 2 h in 25 mM HEPES pH 7.5 and 25 mM NaCl. The protein sample was concentrated to a final protein concentration of ca. 9 mg/mL, from which a volume of 0.2 μL was mixed with 0.1 μL of reservoir solution, and the mixture was equilibrated against 50 μL of reservoir solution (0.2 M NaCl, 0.1 M Tris pH 8.5, 25% (w/v) PEG3350) at 20°C. Crystals were detected after 3 days and cryoprotected by adding 1 μL of 10:7:3 mixture of reservoir solution, 100% (v/v) glycerol, and ddH_2O.

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

Data collection and structure determination

The native datasets were recorded with synchrotron radiation (λ = 1.0 Å or 0.976 Å) at the beamlines X06SA (Swiss Light Source, Villigen, Switzerland) and P13 at the PETRA III storage ring light source (Deutsches Elektronen-Synchrotron, Hamburg, Germany) (Tables S6–S9). Anomalous datasets were recorded at 1.377 Å (Cu), 1.6021 Å (Co), 1.485 Å (Ni), and 1.282 Å (Zn) for crystals co-crystallized with the respective metal ions. Reflection intensities were evaluated using the XDS software package.[78,79] Data reductions were carried out using XSCALE.[78,79] Phasing of TFD-EH apo was achieved by molecular replacement. The deposited structure of TFD-EE (PDB: 6ZV9) was utilized as a search model to generate an electron density map with phase information through Patterson search calculations.[85] The remaining TFD-EH and TFD-EH T87E structures were determined using the TFD-EH apo structure's phase information. For TFD-EE MPNN, the AlphaFold structure prediction was used as the search model for molecular replacement. The models were built, refined, and completed in iterative rounds of restrained refinements using REFMAC5[80] and model building with the 3D graphics program Coot[81] and Main.[86] Water molecules were positioned using ARP/wARP solvent.[87] Restrained and TLS (Translation/Liberation/Screw) refinements with REFMAC5[80] provided satisfactory R_{work} and R_{free} values as well as RMSD, bond length and angle values (Tables S6–S9). The resulting structures were validated using the MolProbity online tool,[82] and deposited in the RCSB Protein Data Bank (PDB entries: 9QUC, 9QUD, 9QUI, 9QUL, 9QUO, 9QUP).

NMR spectroscopy

Backbone resonance assignment

All NMR experiments were recorded on Bruker Avance III spectrometers equipped with cryogenically cooled TCI probe heads, operating at magnetic field strengths corresponding to ^{1}H Larmor frequencies of 600 to 1200 MHz. The sample temperature was set at 298 K for all experiments unless otherwise specified. All experiments were performed in 25 mM HEPES buffer, 25 mM NaCl, pH 7.0, and 8% D_2O. All data were processed using NMRpipe[64] and analyzed in CCPNMR 2.4.[65] The backbone assignment of all TFD-EE with La^{III} and TFD-EH-apo was performed on ^{15}N, ^{13}C, and random fractionally deuterated samples using standard 3D heteronuclear experiments HNCACB, HNCA, HNCO, and HN(CA)CO.[88] Backbone assignments were deposited to the BMRB (ID 53123 and 53124). The assignment of TFD-EE apo, TIM-EE, TFD-EH with metal, and TFD-EH T87E was done by titrations and spectra similarity.

^{1}H-^{15}N HSQC temperature scans and metal titrations

All titrations were done on ^{15}N-labeled samples, at 25°C, in 25 mM HEPES, 25 mM NaCl, pH 7.0, and 8% D_2O. The titration of TFD-EE was done on a sample pre-incubated at 40°C for 3h, concentrated to 700 μM dimer, and with the addition of 700 μM of $LaCl_3$. Titrations of TFD-EH were done at 345 μM dimer with seven consecutive additions of $ZnSO_4$, up to 1380 μM, and at 240 μM dimer with seven consecutive additions of $LaCl_3$ up to 625 μM. The titration on TFD-EH T87E was done at 288 μM dimer with five consecutive additions of $LaCl_3$ up to 860 μM. The TFD-QQ spectra and CSPs were acquired with 550 μM of dimeric protein, with the addition of 550 μM $LaCl_3$. The spectra for the other metals were acquired at 362 μM dimer with equimolar metal addition. The TIM-EE spectra were recorded at 300 μM dimer before and after the addition of 300 μM $LaCl_3$ and incubation at 40°C for 7 h. Finally, the spectra for TFD-EE MPNN were performed at 250 μM dimer with the addition of 250 μM $LaCl_3$. All metals were kept in a water stock solution with concentrations ranging from 100-200 mM. All CSPs were calculated as a weighted average following the equations $CSP = ((\Delta\delta^{1}H)^2 + (\Delta\delta^{15}N*0.15)^2)^{1/2}$ for ^{1}H, ^{15}N spectra.[89]

The temperature jump experiment was performed with ^{15}N-labeled TFD-EE at 700 μM dimer and 700 μM $LaCl_3$ added to a cooled sample. We recorded 20 x 1h ^{1}H-^{15}N HSQCs per temperature point. For the figure, the intensity of an unassigned Asn/Gln resonance was followed and normalized to the highest intensity peak (20 h at 25°C). The peak was selected to reduce noise, as it is one of the most intense at 5°C. The points were plotted according to the time at the start of the experiment.

^{15}N relaxation experiments

^{15}N relaxation (T_1 and T_2 and $\{^{1}H\}$–^{15}N heteronuclear NOE) was measured on samples at around 400 μM dimer. The metal-bound states were prepared at saturation with centrifugation to avoid any aggregation. Data were recorded as described[90] in an interleaved manner with a recycling time of 3.8 s and with eight relaxation delays for T_1 (between 200 and 3400 ms) and T_2 (between 17 and 119 ms). The heteronuclear NOE was recorded in the presence and absence of a 3.8 s ^{1}H saturation sequence (120° ^{1}H pulse train).

PCS measurements, analysis, and cross-validation with MD simulations

PCS were recorded at 1200 MHz and 25°C on TFD-EE concentrated to 700 μM dimer with 700 μM lanthanides. The assignment was performed by comparing the shifts between the metals of different tensor strengths and directions. The tensor was calculated with paramagpy[66] using a fixed metal position, SVD grid search, and NLR gradient descent. The tensor was calculated using PCS from the FL sequence (258 resonances) and quality factors (Q-factors) were calculated according to:

$$Q = \sqrt{\frac{\sum_i \left[\left(\sum_m \left[a_i^{exp} - a_{m,i}^{cal} \right] \right)^2 \right]}{\sum_i \left[\left(\sum_m \left[a_i^{exp} \right] \right)^2 \right]}}$$

where PCS values are denoted a^{exp} and a^{cal}, the index m corresponds to the ensemble averaging of between models, and the index i is the summation of all spins of the molecule. For cross-validation, we used 6 structures representative of the clusters extracted from the PCA analysis of the MD trajectory of TFD-EE with Tb^{III}. The metal was placed in the models by aligning their TIM barrel to the crystal structure of TFD-EE with Tb^{III} (PDB code: 6ZV9). Tensors for all possible ensembles from these 7 models were manually calculated using the paramagpy ensemble fitting function.

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

Molecular dynamics (MD) simulations

Structure modeling and preparation

Missing N- or C-terminal residues in the TFD-EH and TFD-EH T87E crystal structures were modeled using the MODELLER software.[67] The multi-chain model program in MODELLER was applied to the TFD-EE crystal structure to generate a symmetric dimer model by using symmetry constraints. The almost perfectly symmetric crystal structure of TFD-EE MPNN was used directly without symmetrization. The input structures were utilized for molecular dynamics (MD) simulations in apo (metal-free) form and holo (metal-bound) form. For the holo simulations, metal coordinates were taken from the crystal structures by superimposing the modeled structures on the experimental ones.

MD simulation procedure

The molecular dynamics simulations were performed using the AMBER software package.[72,73] The latest AMBER force field parameters for protein (ff19SB),[74] OPC water model, and metal ion parameters LJ1264 were used.[75] For Tb^{III}, the default LJ1264 parameters were used, whereas for Cu^{II} and Zn^{II}, modified LJ1264 parameters were applied.[49,50] Charges on proteins were neutralized, and the OPC water model was used to solvate the protein in a cubic box with a padding of 20 Å. Furthermore, additional Na^+ and Cl^- were added to reach a 150 mM salt concentration.

Following the system preparation, gradual energy minimization was performed in 14 steps, applying restraints on the protein's heavy atoms, while no restraints were used on the water, ions, and hydrogen atoms. The restraints were gradually reduced, starting from 1,000 kcal mol^{-1} A^{-2} to no restraints in the last step. The steepest descent followed by conjugate gradient algorithms was used for energy minimization in each step. Heating simulations using NVT ensembles –constant number of particles (N), volume (V), and temperature (T)– were performed by a gradual increase of temperature from 10 K to 300 K in the first 30 ps. The Langevin thermostat algorithm was used to control bath temperature, keeping the collision frequency at 2.0 ps. The integration time was set to 1.0 fs, and the SHAKE algorithm was used to constrain bonds involving hydrogen atoms. Periodic boundary conditions were applied. Heating simulations were performed using weak constraints on heavy atoms (50 kcal/mol), which were gradually decreased. After heating the equilibration simulation with constant pressure, the NPT ensemble –constant number of particles (P), pressure (P), and temperature (T) was performed. Equilibration simulations of 5 ns with constraints of 5.0 kcal/mol were performed before running a 20 ns simulation with no constraints. The pressure was maintained using the Berendsen barostat. The integration time was increased to 2.0 fs in the second NPT simulation, and trajectory frames and restart files were saved every 100 ps. The final coordinates from the equilibration simulation were taken to perform 3 replica simulations, each simulation was 1.0 μs, with new starting velocities.

MD simulation analysis

The MD trajectories were analyzed after combining the replica simulations in a long, concatenated trajectory. Different analysis methods, including RMSD, RMSF, dihedral angle distributions, and principal component analysis (PCA), were used to extract the dynamic motions of the TFD scaffolds and to understand differences in dynamics between different scaffolds, apo vs. holo, and single point mutations. MD trajectories were analyzed using the MDAnalysis[70] and CCPTRAJ[68] programs. The trajectories and structure files were further visualized using VMD[76] and PyMOL (Schrodinger 2015).

Sequence redesign using *ProteinMPNN*

Sequence redesign was performed using *ProteinMPNN*[7] via the Gradio Webapp (v0.3).[91] The symmetrized TFD-EE backbone (PDB ID: 6ZV9) served as the input structure. The protein was redesigned as a symmetric homodimer, with fixed positions being the tryptophan antenna (W6), the metal-coordinating glutamates (E31, E154), and the tri-glycine linkers (G53-55 and G129-131). The vanilla-v_48_020 model with a sampling temperature of 0.1 and a backbone noise of 0 was applied. A total of 85 new sequences were generated and filtered subsequently. The first selection criterion was the total amount of charged residues per dimer (set to max. 110), resulting in 15 candidate sequences. Next, the AlphaFold2[4] implementation of ColabFold (v.1.5.5)[92] and, later, a local installation of AlphaFold3[6] were used for protein structure prediction. Designs with an average pLDDT>80 were accepted. The final filtering steps were based on the scaffold's cavity size and accessibility as determined with the PyVOL plug-in for PyMOL. We then analyzed the protein surface of selected designs for potential metal binding sites using *BioMetAll*[83] and rationally introduced point mutations to break up clusters of negatively charged residues with the goal of suppressing unspecific lanthanide binding. Synthetic genes of the two redesigns were ordered and tested experimentally (sequence alignment available in the SI).

Protein intermolecular interaction analysis

Intermolecular interactions in the dimer interface of TFD-EE and TFD-EE MPNN were analyzed using *ProtInter*[62] and *ProteinTools.*[63] Since ProtInter calculates interactions for a single chain only, a multi-step protocol was developed to evaluate interactions within the homodimer's chains (chain A and chain B) as well as at their interface. First, the input PDB file was cleaned to remove crystallographic ligands and metals, the chains were merged and renumbered, and exported as a single-chain molecule. Next, the residues corresponding to chain A and chain B were exported separately. ProtInter was then used to calculate the interactions for chain A, chain B, and the combined chain A + chain B structure. To determine the interactions occurring at the interface, the interactions identified in the individual chains (chain A and chain B) were subtracted from those calculated for the merged structure (chain A + chain B).

The types of interactions computed include hydrophobic interactions, ionic interactions, cation-π interactions, main chain–side chain contacts, side chain–side chain contacts, and π-π stacking. The resulting contact data are provided in the SI (section Methods S7, Tables S10–S13). A custom script was developed to visualize these interactions in PyMOL.

Please cite this article in press as: Wagner Egea et al., Modular protein scaffold architecture and AI-guided sequence optimization facilitate de novo metalloenzyme engineering, Structure (2025), https://doi.org/10.1016/j.str.2025.10.010

Photoenzymatic activity

The photoenzymatic diol cleavage was performed and analyzed as described previously.[45] Samples were prepared by incubating the dimeric protein with 1 equivalent of $CeCl_3$ in 25 mM HEPES pH 8.5, 100 mM NaCl at 37°C for 2 hours. Michaelis-Menten kinetics were measured for TFD-EE and TFD-EE MPNN. The reactions were irradiated in a custom-made photoreactor at 410-420 nm for 1 hour at final concentrations of 5 μM dimeric protein, 5 μM $CeCl_3$, substrate concentrations ranging from 25 μM to 6.4 mM, and 10% acetonitrile as co-solvent. After the reaction, the protein was precipitated with acetonitrile. After centrifugation and addition of trimethoxybenzene as an internal standard, samples were subjected to HPLC analysis to quantify the reaction product, benzaldehyde. The initial velocities of product formation were calculated accordingly. Measurements were performed in triplicate. The kinetic parameters k_{cat} and K_M were determined by data fitting with the Michaelis-Menten equation:

$$k_{obs} = \frac{v}{[E]_0} = \frac{k_{cat}\,[S]_0}{K_M + [S]_0}$$

All HPLC analyses were performed on a Vanquish System (Thermo Fisher Scientific) equipped with an achiral Hypersil Gold C18 column (100 x 2.1 mm, 3 μm particle size, Thermo Fisher Scientific).

QUANTIFICATION AND STATISTICAL ANALYSIS

Metal binding studies and photoenzymatic activity measurements were performed in duplicate or triplicate. Given errors refer to the calculated standard deviation and were plotted accordingly in the figures. X-ray crystallography data collection and refinement statistics are summarized in Tables S6, S7, S8, and S9.