



Crystal structure and activity of a de novo enzyme, ferric enterobactin esterase Syn-F4

Kodai Kurihara^a, Koji Umezawa^{b,c}, Ann E. Donnelly^{d,1}, Brendan Sperling^d, Guanyu Liao^d, Michael H. Hecht^{d,2}, and Ryoichi Arai^{a,c,2}

Edited by William DeGrado, University of California San Francisco, San Francisco, CA; received October 27, 2022; accepted August 7, 2023

Producing novel enzymes that are catalytically active *in vitro* and biologically functional *in vivo* is a key goal of synthetic biology. Previously, we reported Syn-F4, the first de novo protein that meets both criteria. Syn-F4 hydrolyzed the siderophore ferric enterobactin, and expression of Syn-F4 allowed an inviable strain of *Escherichia coli* (Δ fes) to grow in iron-limited medium. Here, we describe the crystal structure of Syn-F4. Syn-F4 forms a dimeric 4-helix bundle. Each monomer comprises two long α -helices, and the loops of the Syn-F4 dimer are on the same end of the bundle (*syn* topology). Interestingly, there is a penetrated hole in the central region of the Syn-F4 structure. Extensive mutagenesis experiments in a previous study showed that five residues (Glu26, His74, Arg77, Lys78, and Arg85) were essential for enzymatic activity *in vivo*. All these residues are located around the hole in the central region of the Syn-F4 structure, suggesting a putative active site with a catalytic dyad (Glu26–His74). The complete inactivity of purified proteins with mutations at the five residues supports the putative active site and reaction mechanism. Molecular dynamics and docking simulations of the ferric enterobactin siderophore binding to the Syn-F4 structure demonstrate the dynamic property of the putative active site. The structure and active site of Syn-F4 are completely different from native enterobactin esterase enzymes, thereby demonstrating that proteins designed de novo can provide life-sustaining catalytic activities using structures and mechanisms dramatically different from those that arose in nature.

binary patterned library | de novo enzyme | de novo protein | dimeric 4-helix bundle | ferric enterobactin esterase

Combinatorial libraries of de novo sequences can provide rich sources of diversity for the discovery of new proteins with chemically and biologically important functions (1, 2). Toward this goal, we have used binary patterning of polar and nonpolar amino acids to produce libraries of sequences that fold into stable α -helical structures (3–5). To isolate functional proteins from these libraries, collections of synthetic genes encoding the novel sequences can be transformed into auxotrophic strains of *Escherichia coli* missing a conditionally essential gene required for growth on minimal medium (6). The successful rescue of several such deletion strains previously demonstrated that new proteins unrelated to natural sequences can provide life-sustaining functions.

We previously reported that novel proteins isolated from a binary patterned library can rescue four different auxotrophic strains of *E. coli* (6). For two of these deletions—phosphoserine phosphatase and citrate synthase (encoded by *serB* and *gltA*, respectively)—the novel protein enables growth by altering gene regulation and/or metabolism in the mutant strain (7, 8). In related studies, we isolated a de novo protein capable of rescuing the deletion of threonine deaminase (Δ ilvA). This novel protein was promiscuous, as it also rescued Δ fes, which encodes ferric enterobactin (FeEnt) esterase. This bifunctional de novo protein, called Syn-IF, was then subjected to random mutagenesis followed by selections for faster rescue of Δ fes cells in iron-limited minimal medium. This led to the isolation of the evolved variant, Syn-F4 (Synthetic protein rescues Fes, 4th generation), which showed significant growth on minimal medium in 18 h, much faster than Syn-IF (9). Unlike its Syn-IF ancestor, Syn-F4 no longer rescues Δ ilvA. Thus, the bifunctional Syn-IF generalist had evolved into the more active Syn-F4 specialist (9).

The Syn-F4 protein hydrolyzes FeEnt, thereby enabling Δ fes cells to grow in minimal medium (10). Thus, this entirely new sequence, which did not evolve in living systems, provides a life-sustaining enzymatic function. Syn-F4 is the first example of an enantioselective enzyme based on a nonnatural sequence that provides an essential biological activity *in vivo*. Although the catalytic turnover (~ 0.6 per hour) of Syn-F4 (10) is $\sim 1,000$ -fold slower than that of a natural Fes enzyme (11), Δ fes cells do not require high levels of FeEnt esterase activity. Comprehensive mutagenesis of Syn-F4 followed by high-throughput sequencing revealed residues that must be conserved to maintain activity (10). There were 27 highly conserved positions. Of these, 15 are glycine, alanine, or leucine, and the remaining 12 are

Significance

Life on earth arose from common ancestry. Living systems share similar genes, proteins, and biochemistries. Can life-sustaining functions also be provided by nonnatural proteins?

Combinatorial libraries of de novo sequences present a rich source for exploring such possibilities. The de novo protein, Syn-F4, was previously isolated from such a library and catalyzed the hydrolysis of iron-enterobactin, a reaction essential for the survival of *Escherichia coli* in iron-limited environments. Here, we report the 3-dimensional structure of Syn-F4 and show that its structure and biochemical mechanism differ substantially from those of natural enterobactin esterases, thereby demonstrating that proteins which did not arise in living systems can perform life-sustaining enzymatic functions in ways dramatically different from those that evolved from common ancestry.

Author contributions: A.E.D., M.H.H., and R.A. designed research; K.K., K.U., A.E.D., B.S., G.L., and R.A. performed research; K.K., K.U., B.S., G.L., M.H.H., and R.A. analyzed data; and K.K., K.U., M.H.H., and R.A. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹Present address: Teledyne FLIR, Pittsburgh, PA 15238.

²To whom correspondence may be addressed. Email: hecht@princeton.edu or rarai@shinshu-u.ac.jp.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2218281120/-DCSupplemental>.

Published September 11, 2023.

polar or charged. Saturation mutagenesis of each of these 12 polar or charged residues revealed 5 essential residues (Glu26, His74, Arg77, Lys78, and Arg85) whose replacement by any other amino acid impaired activity (10).

In this study, we report the crystal structure of this first life-sustaining de novo enzyme. To correlate this structure with the observed enzymatic function, we mutated residues in the putative active site, purified the altered proteins, and demonstrated that each of these mutations destroyed enzyme activity.

Results

Crystallization of Syn-F4. We initially tried to crystallize the de novo enterobactin esterase protein, Syn-F4, but obtained only microcrystals with poor diffraction (*SI Appendix, Fig. S1*). Therefore, we attempted to improve crystallization by the surface entropy reduction (SER) strategy (12, 13). This strategy targets surface residues with large flexible side chains, such as Lys and Glu, and replaces them with smaller less flexible amino acids (Ala, Ser, Thr, etc.) and has been shown to improve the preparation of X-ray quality crystals. Four target residues (K4, E17, K42, and K66) of Syn-F4 were selected for mutagenesis. These residues were chosen because comprehensive mutagenesis in a previous study showed they were not important for the enzymatic activity (10). Several mutants to Ala or Thr were constructed and their protein expression was tested (*SI Appendix, Fig. S2*). Some mutants, especially K4T, improved both expression and solubility. In addition, the K4T mutant of Syn-F4 sustained the growth of Δfcs cells (*SI Appendix, Fig. S3*), thereby demonstrating that it retained the original Syn-F4 activity. The quality of the crystals and their X-ray diffraction were improved significantly (*SI Appendix, Fig. S4*), and diffraction data of Syn-F4 (K4T) were obtained at 2.2 Å resolution (crystal 1).

X-ray Crystallographic Analysis of Syn-F4 (K4T). We initially tried to solve the structure of Syn-F4 (K4T) by molecular replacement using the crystal structure of the de novo protein WA20, which was derived from the same binary patterned library (amino acid identity 44%, similarity 95%; *SI Appendix, Fig. S5*) as a search model (PDB ID: 3VJF) (14). However, this was not successful.

Also, the structure could not be solved by multiwavelength anomalous dispersion (MAD) using a SeMet-substituted protein. Finally, the structure was successfully solved by MAD using a platinum-derivative crystal (PDB ID: 8H7C). The crystallographic statistics are summarized in *SI Appendix, Table S1*. Two molecules of Syn-F4 (K4T) are located per asymmetric unit in the crystal. In the case of the crystal 1, the refined model includes 189 amino acid residues of two protein chains of Syn-F4 (K4T) in the asymmetric unit (PDB ID: 8H7D). The N-terminal and C-terminal residues and several loop residues (chain A, 1–2, 49–56, 101–102; chain B, 1–2, 49–55) are invisible due to disorder.

Overall Structure of Syn-F4 (K4T). The crystal structure of Syn-F4 (K4T) revealed a dimeric 4-helix bundle (Fig. 1A). Each monomer comprises two long α -helices, which span residues 5–47 (α 1), 59–98 (α 2) in chain A and residues 5–47 (α 1), 58–101 (α 2) in chain B. The loops of Syn-F4 (K4T) are on the same end of the bundle (*syn* topology). This contrasts with the previously solved cousins WA20 (PDB ID: 3VJF) (14) and SUWA (PDB ID: 6KOS) (15), which have bisecting U topology with loops at opposite ends of the bundles (*SI Appendix, Fig. S6*). The overall shape of Syn-F4 (K4T) is cylindrical with a length of ~8 nm and a diameter of ~3 nm. Interestingly, there is a penetrated hole in the central region of Syn-F4 structure, clearly seen in the surface representation in Fig. 1B.

In addition, to analyze the overall structure in solution, small-angle X-ray scattering (SAXS) analysis of Syn-F4 (K4T) was performed (*SI Appendix, Fig. S7 and Table S2*). The SAXS results indicate that Syn-F4 (K4T) in solution forms a dimer (*SI Appendix, Table S2*) and a cylindrical structure (*SI Appendix, Fig. S7C*) consistent with the crystal structure.

Orientation and Dynamics of the α -helices and Loops. In the crystal structure of Syn-F4 (K4T), the loop regions are disordered. Therefore, it is difficult to determine the precise connections between the four α -helices. To consider alternative connections, we constructed two model structures with different loop conformations, resulting in different connections between the helices (Fig. 2A). We named these “loopleft” and “loopright” to indicate that after ending the first helix, the loop can either

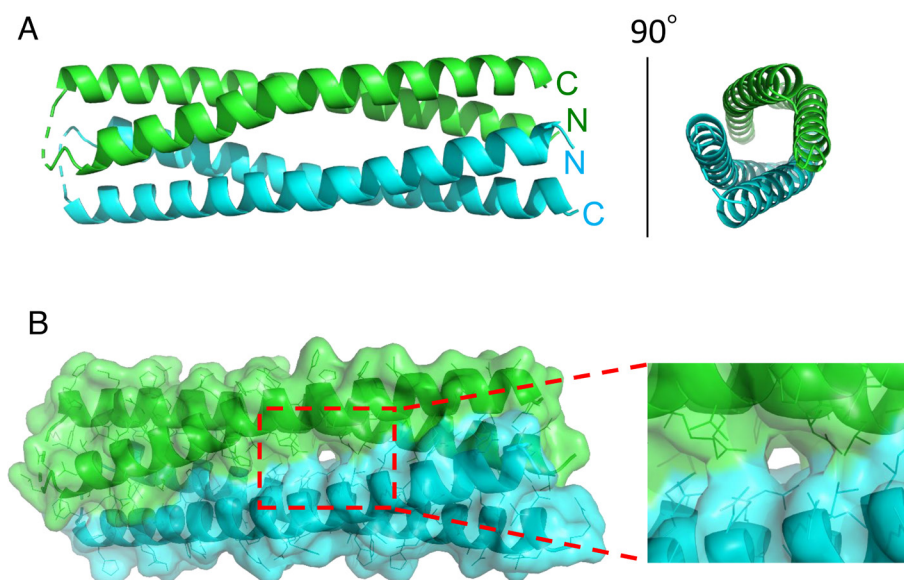


Fig. 1. Overall structure of Syn-F4 (K4T). (A) Ribbon representation of the crystal structure of Syn-F4 (K4T) (crystal 1). A side view (Left) and an axial view (Right). (B) Surface representation of the crystal structure of Syn-F4 (K4T) (crystal 1). Note the tunnel at the center of the structure. Chains A and B are shown in green and cyan, respectively.

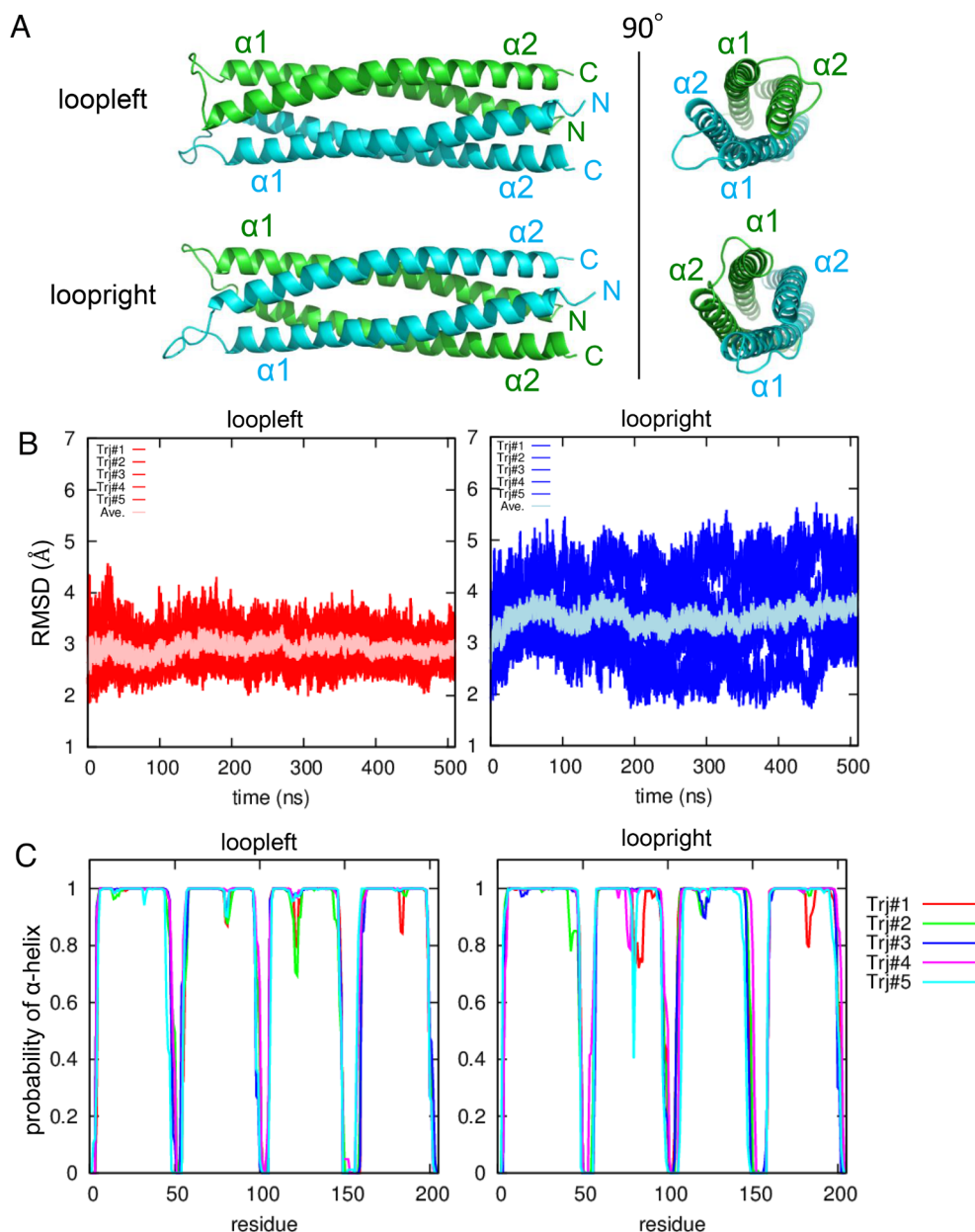


Fig. 2. MD simulations of Syn-F4. (A) Two model structures of Syn-F4 (K4T) with different loop conformations, resulting in different connections between the helices. These loopleft and loopright are indicated that after ending the first helix, the loop can either make a left turn or a right turn toward the second helix (viewed from the outside). Chains A and B are shown in green and cyan, respectively. (B) RMSD of Syn-F4 (K4T) structures from an initial model in MD simulations. (C) Probability of forming α -helical structure for each residue in Syn-F4 (K4T) dimer. Residue numbers 1–102 and 103–204 correspond to amino acid residues in chains A and B, respectively. A probability of 1 indicates that the α -helical structure is preserved over the simulation.

make a left turn (viewed from the outside) or a right turn toward the second helix. Both structures were subjected to molecular dynamics (MD) simulations. These simulations revealed that the structure with loopleft connections was more stable than the structure with the loopright connections (Fig. 2 B and C). These distinctions are nontrivial because, in addition to different conformations in the loops themselves, the two structures have slightly different packing interfaces between the helices. In addition to these distinctions, the simulations also suggest dynamic residues in the middle of the α -helices around residues 19–20 and 81–84.

Putative Active Site of Syn-F4. To find the active site of Syn-F4, we attempted to solve the structure of a crystal with the addition of the substrate, ferric enterobactin (FeEnt) after crystallization;

however, electron density for FeEnt or products was not found (*SI Appendix, Fig. S8*) at 2.0 Å resolution (crystal 2) (PDB ID: 8H7E). In a previous study (10), however, comprehensive mutagenesis of Syn-F4 revealed 27 residues that must be conserved to maintain activity. Of these, 12 residues (Y2, D18, E26, H51, E54, Q63, D65, D70, H74, R77, K78, and R85) are polar or charged. Saturation mutagenesis of each of these 12 positions revealed five residues (E26, H74, R77, K78, and R85) whose replacement by any other amino acid abrogated function (10). Fig. 3A shows that all five of these residues are located around the hole in the central region of the Syn-F4 structure, suggesting that the central hole is the active site. Notably, Glu26, His74, Lys78, and some water molecules formed salt bridges and hydrogen bonds, which may preorganize the structure of this site (Fig. 3B).

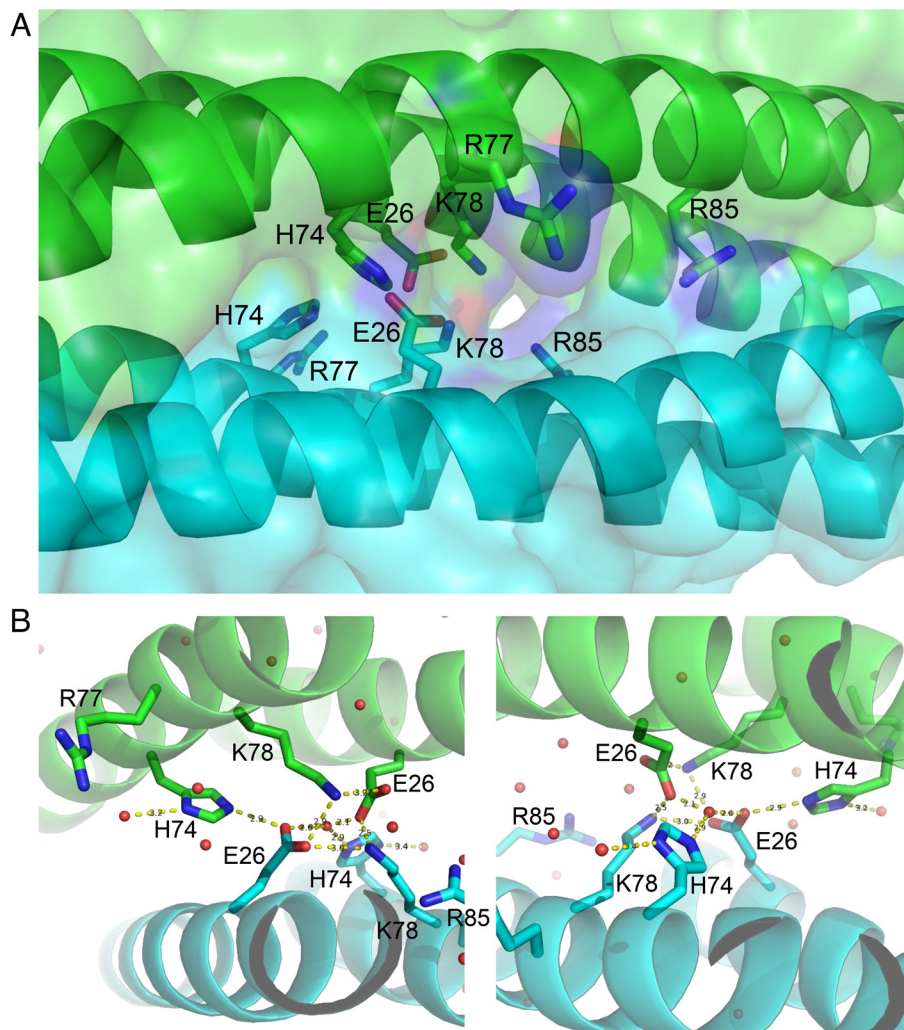


Fig. 3. Putative active site of Syn-F4. (A) Close-up view of the central region of Syn-F4 (K4T) (crystal 2). Five residues (Glu26, His74, Arg77, Lys78, and Arg85) essential for activity are located around the central hole. (B) Close-up views of the essential residues in the putative active site (crystal 2). Glu26, His74, Lys78, and several water molecules (red spheres) form a network of hydrogen bonds and salt bridges (dashed yellow lines).

Simulation of Substrate Docking. Since the crystal structure of Syn-F4 complexed with FeEnt could not be obtained, we analyzed binding by simulation. The docking simulation was performed targeting the vicinity of the putative active site residues (Glu26, His74, Arg77, Lys78, and Arg85). The simulation showed that the ligand binding sites are mainly on the molecular surface and cleft of the central hole (Fig. 4A). In particular, Arg77 and Arg85 were observed at the binding site of FeEnt, suggesting that these residues play important roles in substrate binding (Fig. 4B). Interestingly, some FeEnt molecules were also seen inside the central hole of Syn-F4 (Fig. 4C).

Asymmetric and Variable Local Structures in the Putative Active Site. As noted above, we solved two crystal structures: the initial crystal (crystal 1: PDB 8H7D) and the second crystal (crystal 2: PDB 8H7E), which had been exposed to FeEnt. Although bound substrate was not observed, the availability of two crystal structures provided additional insights into the putative active site. While the overall structures of crystal 1 and crystal 2 are nearly identical (RMSD: 0.52 Å), there are distinct differences in the putative active site (SI Appendix, Fig. S9). In both crystals, both Arg85 side chains in the dimer point toward the interior of the 4-helix bundle, thereby leading to electrostatic repulsion. This repulsion is relieved by the side chain of Arg85 from chain A

bending upward in crystal 1 (SI Appendix, Fig. S9B) and leftward in crystal 2 (SI Appendix, Fig. S9C). The alternative locations of this arginine cause Trp23 from chain A to occupy different rotamers in crystal 1 versus crystal 2 (SI Appendix, Fig. S9D). These subtle differences between the two crystals demonstrate that the local structure in the putative active site is asymmetric, variable, and dynamic. These observations are consistent with the MD simulations, which suggest that repulsion between the two Arg85 residues causes the putative active site to become dynamic and flexible, thereby facilitating binding and entry of FeEnt into the central hole of the 4-helix bundle dimer.

Mutating Residues in the Putative Active Site Destroys Enzymatic Activity. As noted above, extensive mutagenesis experiments described in a previous study (10) showed that Glu26, His74, Arg77, Lys78, and Arg85 are essential for enterobactin esterase activity in vivo. Thus, Δ_{fes} cells expressing the original Syn-F4 sequence formed large white colonies, whereas cells expressing any of these mutants formed tiny colonies (on rich medium), which were red due to the buildup of uncleaved FeEnt. Those genetic results, coupled with the current crystal structure, suggested that the active site occurs in the cavity shown in Figs. 3 and 4. To confirm the importance of these residues for activity, we mutated each of them individually and measured the activities of the

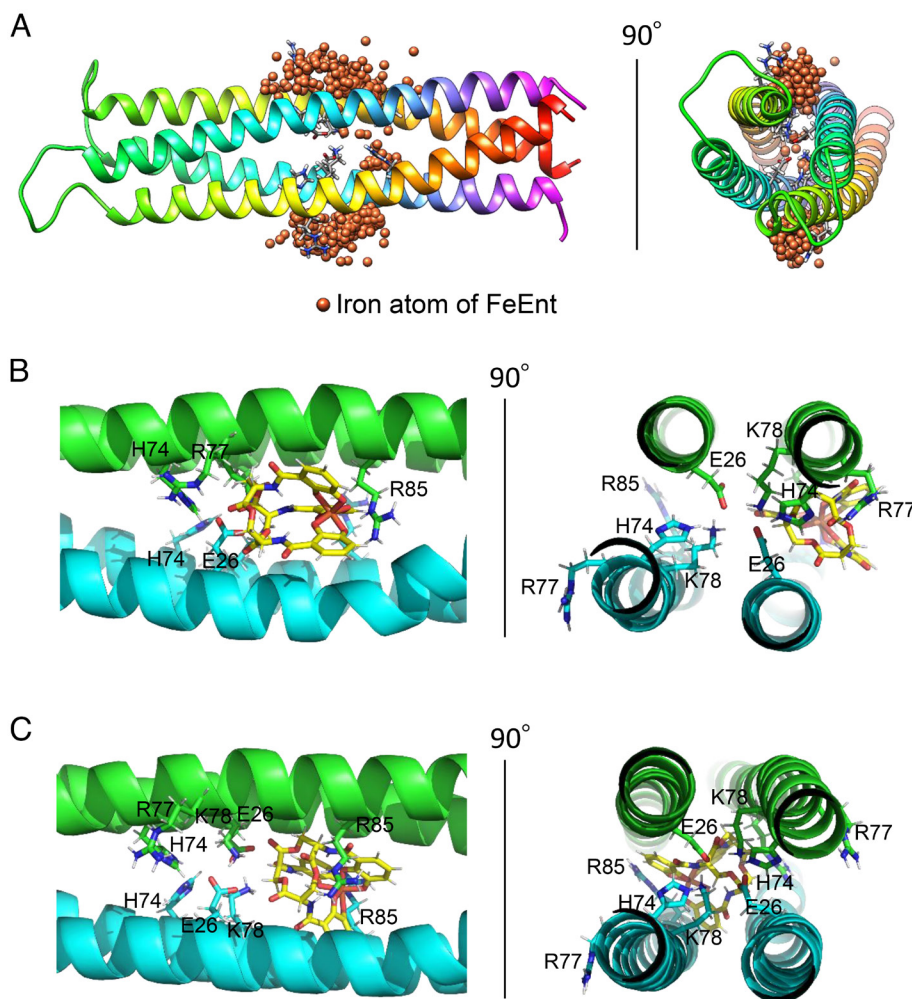


Fig. 4. Docking simulations of FeEnt to Syn-F4. (A) Bound positions of FeEnt on the Syn-F4 structure. Only iron atoms of FeEnt are shown. Five residues (Glu26, His74, Arg77, Lys78, and Arg85) essential for activity are shown as stick models. (B) A docking pose of FeEnt (yellow stick) on the surface cleft in the central region of Syn-F4. (C) A docking pose of FeEnt (yellow stick) inside the central hole in the 4-helix bundle of Syn-F4.

purified proteins. All five positions were mutated individually to alanine. In addition, Glu26 was mutated to Asp, His74 to Asn, and Lys78 to Tyr. The H74A protein was insoluble, and surprisingly, the E26A sequence was cytotoxic. All the other mutant proteins were purified and showed no detectable activity. A Michaelis–Menten plot showing the activity of the original Syn-F4 protein is shown in *SI Appendix, Fig. S10* and the k_{cat} and K_m values of Syn-F4 are 0.60 h^{-1} and $20 \mu\text{M}$, which are ~ 900 -fold lower and ~ 200 -fold higher than those of *E. coli* Fes enzyme (11), respectively (*SI Appendix, Table S3*). The experimental results of the mutant proteins are summarized in *SI Appendix, Table S3*. The total lack of activity for all of these singly substituted proteins highlights the crucial roles played by these residues in the active site of Syn-F4.

Discussion

In this study, we revealed the crystal structure of the de novo ferric enterobactin esterase Syn-F4. Syn-F4 forms a dimeric 4-helix bundle with both interhelical turns on the same side end of the structure (*syn* topology) (*SI Appendix, Fig. S6C*). The structure of this de novo protein is completely different from the structures of the naturally evolved Fes or a related enzyme IroE (*SI Appendix, Fig. S11*). The structure of the native Fes from *Shigella flexneri* (PDB ID: 3C87) has an α/β -hydrolase fold with an N-terminal

lid domain (*SI Appendix, Fig. S11A*) and catalytic triad residues (Ser281, His375, and Glu345) typical for serine hydrolases (*SI Appendix, Fig. S12A*). The structure of the native enterobactin hydrolase IroE from uropathogenic *E. coli* CFT073 (PDB ID: 2GZR) has only an α/β -hydrolase fold (*SI Appendix, Fig. S11B*) and catalytic dyad residues (Ser189 and His287) (*SI Appendix, Fig. S12B*) (16). However, Syn-F4 has no serine residues in the entire sequence (*SI Appendix, Fig. S5*). Thus, Syn-F4 is not a serine hydrolase. Hence, both its 3-dimensional structure and its enzymatic mechanism are entirely different from those of naturally evolved enterobactin hydrolases.

The comprehensive mutagenesis experiments described in a previous study (10) showed that five polar residues (Glu26, His74, Arg77, Lys78, and Arg85) are essential for the enzymatic activity of Syn-F4. All these residues are located around the penetrated hole in the center of the Syn-F4 structure (Fig. 3A), suggesting that this central hole region is the active site. MD simulations demonstrated flexibility of the local structure around this hole (Fig. 3C). Docking simulation of FeEnt onto the Syn-F4 structure suggests binding of FeEnt not only on the surface of Syn-F4 but also inside the structure of the 4-helix bundle. These simulations also support the putative active site in the central hole region.

As described in a previous study (9), the bifunctional Syn-IF generalist was evolved into the more active Syn-F4 specialist with seven mutations (S27R, N29T, E34A, F44V, K62N, L85R, and

E94A). These mutations can be partly explained by the crystal structure of Syn-F4 (*SI Appendix, Fig. S13*). L85R is one of the essential residues for the activity of Syn-F4, and N29T is located close to His74, another essential residue for activity. S27R forms a salt bridge with D79. E34A and E94A eliminate electrostatic repulsions to D72 and D97, respectively. These mutations contribute to stabilizing the structure of Syn-F4. The effects of F44V and K62N are unclear but may be helpful for tuning the structure of Syn-F4. However, we do not fully understand how Syn-IF rescues $\Delta ilvA$ and why Syn-F4 does not. Syn-IF may rescue using a non-catalytic mechanism, as was seen for SynSerB3 (7) and SynGltA (8).

The crystal structure indicates that Glu26, His74, Lys78, and several water molecules form electrostatic interactions and hydrogen bonds (Fig. 3 *B* and *C*). There are ordered water molecules located in hydrogen bond distances from N δ or N ϵ of His74 in the crystal structure (crystal 2). MD simulations of Syn-F4 (K4T) also support that there are significantly high densities of water molecules in appropriate distances from N δ or N ϵ of His74 (*SI Appendix, Fig. S14*). Notably, the structure of the hydrogen bond network of Glu26–His74–water (Fig. 3*B*) suggests that Glu26 and His74 can work as a potential catalytic dyad which activates a water (Fig. 5), similar to the His–Asp catalytic dyads seen in some natural hydrolases, such as ribonuclease A (17) and N-acetylglucosaminidase (18). The complete inactivity of purified proteins with mutations at Glu26, His74, Arg77, Lys78, or Arg85 supports this mechanism.

Moreover, in the simulation docking FeEnt to Syn-F4, the distance distribution of Fe atom or ester carbon atom of FeEnt from His74 of Syn-F4 was analyzed (*SI Appendix, Fig. S15*). The distance distribution of ester carbon from His74 is slightly shorter than that of Fe from His74, indicating that there are many docking poses whose ester bond of FeEnt is oriented toward the catalytic residues of Syn-F4. Several binding poses which have short distance (<6 Å) from His74 and good binding scores were selected

(*SI Appendix, Fig. S16*). The selected poses showed the similar binding mode where FeEnt is partly inserted to the cleft on the molecular surface in the vicinity of the central hole (Fig. 4*B* and *SI Appendix, Fig. S16*). The ester bond of FeEnt is reasonably oriented to the catalytic dyad (Glu26–His74) (Fig. 5*A*), suggesting the putative mechanism of Syn-F4 for hydrolysis reaction of ester bond by the catalytic dyad and an activated water (Fig. 5*B*).

Four-helix bundles are typical of simple and versatile scaffolds in protein structures, and many natural and artificial 4-helix bundle proteins have been reported (19–21). Dali server search (22) revealed many structures partly similar to the backbone structure of Syn-F4 in natural proteins. *SI Appendix, Fig. S17* shows the superimposition of backbone structures of Syn-F4 and several highly ranked proteins with structural similarity: tripartite cytolytic toxins (PDB ID: 6T8D, 6GRK) (23, 24), peripherin-2 (PRPH2) (PDB ID: 7ZW1) (25), and Tweety homologs (TTYHs) (PDB ID: 7P54) (26). The findings in this study would be helpful for grafting or designing an active site on versatile scaffolds of 4-helix bundle proteins.

Taken together, the crystal structure of Syn-F4 revealed a structure dramatically different from native enterobactin esterases, suggesting that de novo designed proteins can create a new lineage of enzymes having structures and mechanisms different from native enzymes. These results, taken together with previous work on other binary patterned proteins demonstrate that the diversity of functional proteins can be expanded well beyond the set of proteins sampled by terrestrial evolution.

Materials and Methods

Protein Expression and Purification. The Syn-F4 gene was cloned in an IPTG-inducible expression vector, pCA24N, with chloramphenicol resistance. The Syn-F4 mutants were prepared by site-directed mutagenesis using the Transfer-PCR method (27) with degenerate primers (*SI Appendix, Table S4*). The Syn-F4 and Syn-F4 (K4T) proteins were expressed in *E. coli* BL21 Star (DE3) (Invitrogen, Carlsbad, CA) using 2 L of LB broth (Lennox) (Nacalai Tesque, Kyoto, Japan) with 34 μ g/mL chloramphenicol at 30 °C. Expression was induced with 0.1 mM IPTG (at OD₆₀₀ = ~0.5), and cells were further cultured for 16 h at 30 °C. The protein was extracted from harvested cells using the freeze-thaw method (28) in lysis buffer (50 mM sodium phosphate (pH 7.0), 300 mM NaCl, 10% glycerol). The protein was purified by immobilized metal ion affinity chromatography with COSMOSIL His-Accept (Nacalai Tesque) [equilibration/wash buffer: 50 mM sodium phosphate buffer (pH 8.0) containing 300 mM NaCl and 10% glycerol; elution buffer: 50 mM sodium phosphate buffer (pH 8.0) containing 300 mM NaCl, 10% glycerol and 250 mM imidazole]. Even without a His-tag, the Syn-F4 and Syn-F4 (K4T) proteins can bind to the nickel ion affinity resin, presumably because of the relatively high percentage of His residues (13.7%: 14/102 residues) in their sequences (*SI Appendix, Fig. S5*). The protein was further purified by cation exchange chromatography (25 mM MES buffer (pH 6.5) containing 10% glycerol with a linear gradient of NaCl from 0 to 1.5 M) with a RESOURCE S 6-mL column (Cytiva, Little Chalfont, Buckinghamshire, UK) and size-exclusion chromatography [25 mM MES buffer (pH 6.5) containing 100 mM NaCl, 10% glycerol, and 200 mM Arg-HCl with a Superdex 75 Increase 10/300 GL column (Cytiva)].

Crystallization. Several crystals of the Syn-F4 (K4T) protein were obtained in a drop composed of 0.5 μ L of the protein solution (~10 mg/mL) and 0.5 μ L of the reservoir solution at 20 °C by the sitting drop vapor diffusion method against 50 μ L of the reservoir solution in an Index crystallization screening kit (Hampton Research, Aliso Viejo, CA) with a 96-well Viologam protein crystallization plate VCP-1 (AS ONE, Osaka, Japan). Optimized crystals were obtained in a drop composed of 1 μ L of the protein solution and 1 μ L of the reservoir solution [100 mM Tris (pH 8.5 to 8.9), 200 mM ammonium acetate, 8 to 20% w/v polyethylene glycol 3,350] by the hanging drop vapor diffusion method against 500 μ L of the reservoir solution at 20 °C with a 24-well plate. Plate-like crystals were obtained in a few weeks (*SI Appendix, Fig. S4*). Platinum-derivative crystals were prepared by soaking crystals in the reservoir solution containing 10 mM K₂PtCl₄ for 1 d.

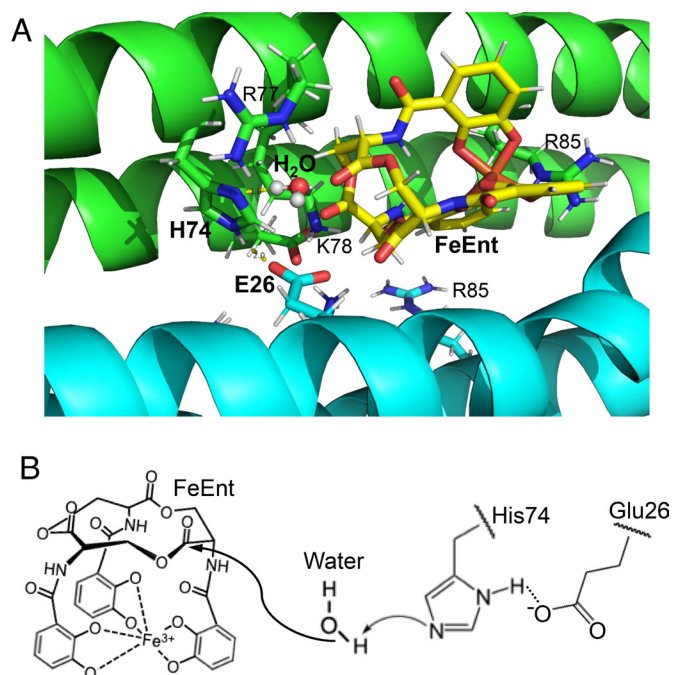


Fig. 5. Putative reaction mechanism of Syn-F4. (A) A binding pose of FeEnt on the Syn-F4 (K4T) structure in a short distance from His74 by docking simulation. A water molecule was added to explain a reaction mechanism. (B) Putative reaction scheme of the de novo enzyme Syn-F4 with the catalytic dyad residues (Glu26 and His74) and an activated water.

Data Collection, Structure Determination, and Refinement. X-ray diffraction data were collected at the Photon Factory, BL-5A and AR-NW12A (KEK, Tsukuba, Japan). The data collection was carried out at 95 K with 30% glycerol as a cryoprotectant. All diffraction data were processed with the program HKL2000 (29) (*SI Appendix, Table S1*). The program AutoSol in Phenix was used to locate the heavy atom sites and to calculate the phases by the MAD method, and the program was used for the density modification and partial model building (30, 31). The model was built and corrected with the program COOT (32) and was refined with the program REFMAC5 (33, 34) in the CCP4 suite (35). All refinement statistics are presented in *SI Appendix, Table S1*. The quality of the model was inspected by the programs PROCHECK (36) and MolProbity (37–39). The atomic coordinates and the structure factors have been deposited in the Protein Data Bank with the accession codes 8H7C, 8H7D, and 8H7E. The graphic figures were created using the programs PyMOL (Schrödinger, New York, NY) and UCSF Chimera (40).

SAXS. SAXS measurements were performed for samples (~5 mg/mL) of Syn-F4 (K4T) and hen egg lysozyme dissolved in 25 mM MES buffer (pH 6.5) containing 100 mM NaCl, 10% glycerol, and 200 mM Arg-HCl at 20 °C using synchrotron radiation ($\lambda = 1.2 \text{ \AA}$) at the Photon Factory BL-10C beamline (41) (KEK, Tsukuba, Japan) with a PILATUS3 2M detector (Dectris, Baden, Switzerland) at a sample-detector distance of 1 m. The two-dimensional scattering images were integrated into one-dimensional scattering intensities $I(q)$ as a function of the magnitude of the scattering vector $q = (4\pi/\lambda)\sin(\theta/2)$ using SAnGler (42), where θ is the total scattering angle. The forward scattering intensity, $I(q \rightarrow 0)$, and radius of gyration, R_g , were estimated by Guinier approximation (43) using AUTORG in ATSAS (44) with SAnGler (42). Forward scattering intensity normalized by the protein concentration (mg/mL), $I(q \rightarrow 0)/c$, is proportional to weight average molecular mass (M_w). Lysozyme ($M_w = 14.3 \text{ kDa}$) was used as a reference standard of the molecular mass. A low-resolution dummy atom model was constructed from the SAXS data using ab initio shape modeling programs in the ATSAS program suite (44). Calculations of rapid ab initio shape determination were performed ten times by DAMMIF (45) without a symmetry constraint, and the generated models were aligned and averaged by DAMAVER (46). The averaged model was modified with the fixed core by DAMSTART, and further refinement of the model was performed by DAMMIN (47). Superimposing the dummy atom model and crystal structures was performed by SUPCOMB (48). The SAXS data and dummy-atom model of Syn-F4 have been deposited in Small-Angle Scattering Biological Data Bank (49) (SASBDB accession code: SASDRJ8).

MD Simulation and Docking Simulation. MD simulations for the two types of loop connections (loopleft and loopright) of Syn-F4 were performed under explicit solvent condition with periodic boundary of rectangular box system. The initial conformations were built by MODELLAR (50). Amber ff14SB force field (51) was used for protein. TIP3P model (52) was used for water molecules. The potassium and chloride ions were added to the system for neutralizing the total charge and setting adequate salt concentration of 100 mM. After energy minimization, the simulation system was heated from 1 K to 310 K. The initial velocities of atoms were randomly assigned according to the maxwell-boltzmann distribution. After equilibration for 10 ns under 1 bar at 310 K, production run was conducted for 500 ns. During the production run, the snapshots were saved every 100 ps. Then, 5,000 snapshots were sampled. The MD simulation was repeated five times with different initial velocities to produce five trajectories. Finally, 25,000 snapshots were saved for analysis of secondary structure and rmsd against the initial structure. The 250 snapshots, which were saved every 10 ns, were used for the following docking simulation. MD simulation and analysis were done by using Amber18 and AmberTools18 (53).

Docking simulation of FeEnt was conducted onto the Syn-F4 conformations (255 snapshots) sampled by the MD simulation. The molecular model of FeEnt was prepared by conformational optimization with the quantum mechanics simulation of B3LYP/6-311++G, which was calculated by Gaussian09 (54). The partial atomic charges of FeEnt were fitted by the RESP method by antechamber of AmberTools18. The FeEnt model was docked onto a targeted site of the Syn-F4 by the docking software package, SievGene (55). The targeted site was located around the atom coordinates of residues essential for enzymatic activity. The top-ranking score and its docking pose were saved for each of the Syn-F4 conformations.

Mutagenesis, Protein Purification, and Enzymatic Activity. Mutations of the putative active site residues (E26, H74, R77, K78, and R85) were generated by site-directed mutagenesis using PCR. The PCR products were treated with a kinase-ligase-DpnI (KLD) mix (New England Biolabs.) The products were cloned into DH5 α cells and plated on LB agar containing 30 $\mu\text{g/mL}$ chloramphenicol. Single colonies were picked and sequence-validated. Validated sequences were transformed into BL21 cells for protein expression. Expression was induced with 100 μM IPTG and cultures were grown at 18 °C for 18 h. Cells were harvested and lysed by freeze-thaw and sonication. Soluble material was loaded on a Ni-NTA column, and eluted fractions were run on a size-exclusion column using an AKTA Pure system. Fractions containing protein with the correct molecular weight were collected and stored at -80°C for future assays.

The substrate, ferric enterobactin, was prepared as described in Donnelly et al. (10). To assay Syn-F4 and mutants, 25 μM purified protein was incubated with 0, 10, 20, 40, 80, 120, and 200 μM substrate in 75 mM HEPES, 1 M NaCl, pH 8.0 at 37 °C. Aliquots of reactions were quenched with 2.5 M HCl in methanol at time points 0, 10, and 30 min, followed by extraction with ethyl acetate. The organic layer was analyzed by reverse-phase HPLC monitored at 316 nm. The final concentrations of reactant and products were calculated from areas under peaks in the chromatogram. Michaelis-Menten kinetics were determined by nonlinear regression.

Data, Materials, and Software Availability. Atomic coordinates and the structure factors data have been deposited in Protein Data Bank (8H7C, 8H7D, and 8H7E) (56–58). SAXS data and dummy-atom model have been deposited in Small-Angle Scattering Biological Data Bank (SASDRJ8) (59).

ACKNOWLEDGMENTS. We thank Ryuichi Yamanaka and Sha Tao for help in preliminary experiments and Jacob Davis for assistance with enzyme kinetics. We thank Betsy Smith for discovering Syn-F4. The synchrotron X-ray experiments were performed at Photon Factory (PF), KEK, under the approval of PF program advisory committee (Proposal No. 2016G606, 2016G617, 2018G634, 2018G636, and 2020G658). We thank the beamline scientists and staff at PF, KEK. We are indebted to Research Center for Advanced Science and Technology, Shinshu University for providing facilities. This work was supported by JSPS KAKENHI Grant numbers JP24780097, JP16K05841, JP16H00761, JP17KK0104, and JP19H02522 to R.A.; by JSPS Invitational Fellowship S18035 to R.A. and M.H.H.; and by NSF grant MCB-1947720 to M.H.H.

Author affiliations: ^aDepartment of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, Ueda, Nagano 386-8567, Japan; ^bDepartment of Agricultural and Life Sciences, Faculty of Agriculture, Shinshu University, Minami-minowa, Kami-ina, Nagano 399-4598, Japan; ^cDepartment of Biomolecular Innovation, Institute for Biomedical Sciences, Interdisciplinary Cluster for Cutting Edge Research, Shinshu University, Matsumoto, Nagano 390-8621, Japan; and ^dDepartment of Chemistry, Princeton University, Princeton, NJ 08544

1. A. D. Keefe, J. W. Szostak, Functional proteins from a random-sequence library. *Nature* **410**, 715–718 (2001).
2. B. Seelig, J. W. Szostak, Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **448**, 828–831 (2007).
3. S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, M. H. Hecht, Protein design by binary patterning of polar and nonpolar amino acids. *Science* **262**, 1680–1685 (1993).
4. M. H. Hecht, A. Das, A. Go, L. H. Bradley, Y. Wei, De novo proteins from designed combinatorial libraries. *Protein Sci.* **13**, 1711–1723 (2004).
5. L. H. Bradley, P. P. Thumfort, M. H. Hecht, De novo proteins from binary-patterned combinatorial libraries. *Methods Mol. Biol.* **340**, 53–69 (2006).
6. M. A. Fisher, K. L. McKinley, L. H. Bradley, S. R. Viola, M. H. Hecht, De novo designed proteins from a library of artificial sequences function in *Escherichia coli* and enable cell growth. *PLoS One* **6**, e15364 (2011).
7. K. M. Digianantonio, M. H. Hecht, A protein constructed de novo enables cell growth by altering gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2400–2405 (2016).
8. K. M. Digianantonio, M. Korolev, M. H. Hecht, A non-natural protein rescues cells deleted for a key enzyme in central metabolism. *ACS Synth. Biol.* **6**, 694–700 (2017).
9. B. A. Smith, A. E. Mularz, M. H. Hecht, Divergent evolution of a bifunctional de novo protein. *Protein Sci.* **24**, 246–252 (2015).
10. A. E. Donnelly, G. S. Murphy, K. M. Digianantonio, M. H. Hecht, A de novo enzyme catalyzes a life-sustaining reaction in *Escherichia coli*. *Nat. Chem. Biol.* **14**, 253–255 (2018).
11. H. Lin, M. A. Fischbach, D. R. Liu, C. T. Walsh, In vitro characterization of salmochelin and enterobactin trilactone hydrolases IroD, IroE, and Fes. *J. Am. Chem. Soc.* **127**, 11075–11084 (2005).
12. Z. S. Derewenda, Rational protein crystallization by mutational surface engineering. *Structure* **12**, 529–535 (2004).

13. D. R. Cooper *et al.*, Protein crystallization by surface entropy reduction: Optimization of the SER strategy. *Acta Crystallogr. D* **63**, 636–645 (2007).
14. R. Arai *et al.*, Domain-swapped dimeric structure of a stable and functional *de novo* four-helix bundle protein, WA20. *J. Phys. Chem. B* **116**, 6789–6797 (2012).
15. N. Kimura, K. Mochizuki, K. Umezawa, M. H. Hecht, R. Arai, Hyperstable *de novo* protein with a dimeric bisecting topology. *ACS Synth. Biol.* **9**, 254–259 (2020).
16. N. A. Larsen, H. Lin, R. Wei, M. A. Fischbach, C. T. Walsh, Structural characterization of enterobactin hydrolase IroE. *Biochemistry* **45**, 10184–10190 (2006).
17. L. W. Schultz, D. J. Quirk, R. T. Raines, His-Asp catalytic dyad of ribonuclease A: Structure and function of the wild-type, D121N, and D121A enzymes. *Biochemistry* **37**, 8886–8898 (1998).
18. S. Litzyngier *et al.*, Structural and kinetic analysis of Bacillus subtilis N-acetylglucosaminidase reveals a unique Asp-His dyad mechanism. *J. Biol. Chem.* **285**, 35675–35684 (2010).
19. R. B. Hill, D. P. Raleigh, A. Lombardi, W. F. Degradó, De novo design of helical bundles as models for understanding protein folding and function. *Acc. Chem. Res.* **33**, 745–754 (2000).
20. G. G. Rhys *et al.*, Navigating the structural landscape of de novo alpha-helical bundles. *J. Am. Chem. Soc.* **141**, 8787–8797 (2019).
21. A. Lombardi, F. Pirro, O. Maglio, M. Chino, W. F. DeGrado, De novo design of four-helix bundle metalloproteins: One scaffold, diverse reactivities. *Acc. Chem. Res.* **52**, 1148–1159 (2019).
22. L. Holm, Dali server: Structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215 (2022).
23. A. Nadeem *et al.*, A tripartite cytolytic toxin formed by Vibrio cholerae proteins with flagellum-facilitated secretion. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2111418118 (2021).
24. J. S. Wilson *et al.*, Identification and structural analysis of the tripartite alpha-pore forming toxin of *Aeromonas hydrophila*. *Nat. Commun.* **10**, 2900 (2019).
25. D. El Mazouni, P. Gros, Cryo-EM structures of peripherin-2 and ROM1 suggest multiple roles in photoreceptor membrane morphogenesis. *Sci. Adv.* **8**, eadd3677 (2022).
26. A. Sukalskaia, M. S. Straub, D. Deneka, M. Sawicka, R. Dutzler, Cryo-EM structures of the TTYH family reveal a novel architecture for lipid interactions. *Nat. Commun.* **12**, 4893 (2021).
27. A. Erijman, A. Dantes, R. Bernheim, J. M. Shifman, Y. Peleg, Transfer-PCR (TPCR): A highway for DNA cloning and protein engineering. *J. Struct. Biol.* **175**, 171–177 (2011).
28. B. H. Johnson, M. H. Hecht, Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. *Biotechnology* **12**, 1357–1360 (1994).
29. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
30. P. D. Adams *et al.*, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
31. T. Terwilliger, SOLVE and RESOLVE: Automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52 (2004).
32. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
33. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
34. G. N. Murshudov *et al.*, REFMACS for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
35. Collaborative Computational Project, The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
36. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, Procheck-A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
37. V. B. Chen *et al.*, MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
38. I. W. Davis *et al.*, MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
39. S. C. Lovell *et al.*, Structure validation by Calpha geometry: Phi, psi and Cbeta deviation. *Proteins* **50**, 437–450 (2003).
40. E. F. Pettersen *et al.*, UCSF Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
41. N. Shimizu *et al.*, BL-10C, the small-angle x-ray scattering beamline at the photon factory. *AIP Conf. Proc.* **2054**, 060041 (2019).
42. N. Shimizu *et al.*, Software development for analysis of small-angle X-ray scattering data. *AIP Conf. Proc.* **1741**, 050017 (2016).
43. O. Glatter, O. Kratky, *Small-Angle X-Ray Scattering* (Academic Press, New York, 1982).
44. D. Franke *et al.*, ATLAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
45. D. Franke, D. I. Svergun, DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
46. V. V. Volkov, D. I. Svergun, Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
47. D. I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
48. M. B. Kozin, D. I. Svergun, Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41 (2000).
49. A. G. Kikhney, C. R. Borges, D. S. Molodenskiy, C. M. Jeffries, D. I. Svergun, SASBDB: Towards an automatically curated and validated repository for biological scattering data. *Protein Sci.* **29**, 66–75 (2020).
50. B. Webb, A. Sali, Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **47**, 5.6.1–5.6.32 (2014).
51. J. A. Maier *et al.*, ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
52. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
53. D. A. Case *et al.*, *AMBER 2018* (University of California, San Francisco, 2018).
54. M. J. Frisch *et al.*, *Gaussian 09* (Gaussian Inc., Wallingford, CT, 2016).
55. Y. Fukunishi, Y. Mikami, H. Nakamura, Similarities among receptor pockets and among compounds: Analysis and application to in silico ligand screening. *J. Mol. Graph. Model.* **24**, 34–45 (2005).
56. K. Kurihara, K. Umezawa, A. E. Donnelly, M. H. Hecht, R. Arai, Crystal structure of a de novo enzyme, ferric enterobactin esterase Syn-F4 (K4T) - Pt derivative. *Protein Data Bank*. <https://www.rcsb.org/structure/8H7C>. Deposited 19 October 2022.
57. K. Kurihara, K. Umezawa, A. E. Donnelly, M. H. Hecht, R. Arai, Crystal structure of a de novo enzyme, ferric enterobactin esterase Syn-F4 (K4T). *Protein Data Bank*. <https://www.rcsb.org/structure/8H7D>. Deposited 19 October 2022.
58. K. Kurihara, K. Umezawa, A. E. Donnelly, M. H. Hecht, R. Arai, Crystal structure of a de novo enzyme, ferric enterobactin esterase Syn-F4 (K4T) at 2.0 angstrom resolution. *Protein Data Bank*. <https://www.rcsb.org/structure/8H7E>. Deposited 19 October 2022.
59. K. Kurihara *et al.*, A de novo enzyme, ferric enterobactin esterase Syn-F4. *Small Angle Scattering Biological Data Bank*. <https://www.sasbdb.org/data/SASDRJ8>. Deposited 7 April 2023.