bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Engineering a Covalent Linkage into a Dimeric *De Novo* Enzyme Reveals a Novel Life-Sustaining Mechanism

Guanyu Liao[1], Sha Tao[1], Mina Nagahara[2], Kodai Kurihara[2], Koji Umezawa[3], Ryoichi Arai[2], and Michael H. Hecht[1]*

[1]Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA

[2]Department of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, Ueda, Nagano 386-8567, Japan

[3]Department of Agricultural and Life Sciences, Faculty of Agriculture, Shinshu University, Minami-minowa, Nagano 399-4598, Japan

Present address:
G.L.: Department of Biochemistry and Molecular Biology, University of Chicago
S.T.: Department of Molecular Biology, Princeton University

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ABSTRACT

Designing novel proteins that share no homology with natural sequences, but which nonetheless provide life sustaining functions, is an important goal for synthetic biology. Towards this goal, we previously reported Syn-F4, the first *de novo* enzyme capable of catalyzing a life-sustaining reaction, both *in vitro* and *in vivo*. Syn-F4 catalyzes hydrolysis of the siderophore, ferric enterobactin, thereby releasing iron and enabling growth in iron-limited media of an otherwise inviable Δ*fes* strain of *Escherichia coli*. Although Syn-F4 provides a direct enzymatic replacement of the natural ferric enterobactin esterase encoded by Fes, it has a dramatically different structure and enzymatic mechanism than the natural Fes enzyme. The novel Syn-F4 enzyme forms a 4-helix bundle, comprising a homodimer of two α-helical hairpins. Here we describe the engineering of a covalent peptide linkage into the homodimer to generate a single chain 4-helix bundle. As expected, the resulting linked protein (Syn-F4-Link) also rescued Δ*fes* cells in iron-limited media. Moreover, X-ray crystallography revealed a 3D structure similar to the parental homodimer. Surprisingly, however, the linked protein was *not* enzymatically active. Instead, Syn-F4-Link rescues Δ*fes* cells by upregulating biosynthesis of the enterobactin siderophore thereby enabling assimilation of sufficient iron to sustain cell growth. These findings demonstrate that two very similar *de novo* proteins can sustain cell growth using dramatically different biological mechanisms.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

INTRODUCTION

Natural proteins perform a wide range of functions necessary to sustain the growth of living organisms. Some proteins are structural, others are catalytic, and yet others perform regulatory functions necessary to modulate cell growth. However, despite the enormous diversity of functions, structures, and amino acid sequences observed in biology, the total collection of proteins in all biological systems on earth (past and present) represents a miniscule fraction of what might be possible.[1-3] Indeed, the burgeoning field of *de novo* protein design – recognized recently by a Nobel Prize – demonstrates that it is now possible to devise entirely new sequences that fold into novel 3-dimensional structures not seen in nature.[4-7] Recent advances in protein design have also demonstrated the possibility of creating proteins with a range of novel functions.[8-14]

Our previous work in protein design focused on the synthesis of large combinatorial libraries of novel sequences, followed by screens and selections aimed at discovering non-natural proteins with new functions.[1,15,16] We focused not only on proteins that function in a test tube, but more importantly, on novel proteins that perform essential functions in living cells. For example, we discovered novel proteins that enable the growth of *E. coli* in concentrations of copper that would otherwise be toxic.[17] We also discovered novel proteins that rescue nutritional auxotrophs in which an essential gene is deleted.[18] For example, the deletion of *serB* (which encodes phosphoserine phosphatase) or *gltA* (which encodes citrate synthase) prevent the growth of *E. coli* in media lacking serine or glutamic acid, respectively. In both these cases, we isolated *de novo* proteins that rescue these deletions by altering gene regulation to increase the expression of alternative endogenous enzymes that have promiscuous activities, which if expressed at high levels, can rescue the auxotrophic strain.[19,20]

Novel proteins might also function *in vivo* by directly providing an essential enzymatic activity. For example, the growth of *E. coli* in iron-limited media depends on the endogenous enzyme Ferric Enterobactin Esterase (Fes), which degrades the siderophore Ferric Enterobactin (Fe(III)-Ent) to release iron and make it available for cell growth.[21-23] Mutant cells in which this enzyme is deleted (Δ*fes)* fail to grow in iron poor media.[24] Several years ago, we described the isolation of a binary patterned 4-helix bundle *de novo* protein, called Syn-F4, which rescues Δ*fes* cells.[25] Purification of Syn-F4 confirmed that the novel protein is catalytically active as an enterobactin esterase, and moreover, that it is stereospecific.[25] Recently, we reported the crystal structure of Syn-F4, demonstrating that it is a homodimeric 4-helix bundle with the C-terminus of one chain in close proximity to the N-terminus of the second chain.[26] Model building and extensive mutagenesis studies suggested the catalytic site is at the center of the 4-helix bundle comprising residues from both polypeptide chains.[26]

The homodimeric 4-helix bundle structure of Syn-F4 is entirely different from the larger and more complex structure of the natural Fes enzyme. Moreover, the *de novo* protein catalyzes hydrolysis of its Fe(III)-Ent substrate using a mechanism that differs from that of the naturally evolved enzyme: whereas the natural enzyme is a classic serine hydrolase with a catalytic triad comprising serine, histidine, and aspartic acid, Syn-F4 contains no serine in its entire sequence

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and uses water as a nucleophile to hydrolyze the enterobactin substrate.[26] The fact that the sequence, structure, and catalytic mechanism of Syn-F4 differ so dramatically from those of the natural Fes protein provide compelling evidence that life-essential biological challenges can be solved in ways that are very different from those selected by evolution.[26]

Because the termini of the two chains are close to one another in the crystal structure of the Syn-F4 homodimer, we surmised that linking these chains into a single double-length polypeptide would diminish the entropic cost of dimerization and thereby produce a more stable and/or more active enzyme. Here we describe construction of the covalent dimer and report that its crystal structure, as expected, is similar to that of the non-covalent dimer. Surprisingly, however, the enzymatic properties of these very similar proteins are dramatically different: the linked protein *cannot* catalyze the hydrolysis of Fe(III)-Ent. Consequently, although both the noncovalent dimer (Syn-F4) and the newly constructed covalent dimer (Syn-F4-Link) rescue Δ*fes* cells *in vivo,* they do so using distinctly different biological mechanisms.

RESULTS

Creation of a covalent dimer of Syn-F4 by inserting a flexible linker

Previously, we determined the crystal structure of Syn-F4 and found that it forms a homodimeric 4-helix bundle, with the N- and C-termini of both chains occurring at the same end of the bundle (*syn* topology, **SI Figure 1**).[26] Based on this structure, we assumed that linking the two chains with a short flexible peptide would produce a covalently linked homodimer with a structure and properties similar to the original non-covalent dimer of Syn-F4. Toward this goal, we constructed a gene encoding a protein in which the two chains are joined by the flexible decapeptide, Gly-Gly-Gly-Gly-Ser-Gly-Gly-Gly-Gly-Ser (**Figure 1A**). We also made two mutations on the linked version of Syn-F4 at K4T and K116T corresponding to the previous mutation K4T on both chains of Syn-F4, which had been incorporated to enhance crystallization. We named this covalent dimer Syn-F4-Link K4Tx2, or Syn-F4-Link in brief.

Syn-F4-Link and the non-covalent Syn-F4 homodimer have similar structures, yet slightly different active site conformations

We solved the crystal structure of Syn-F4-Link and found it to be nearly identical to that of the original Syn-F4, with C-alpha RMSD = 0.45 Å (**Figure 2A, B**). The catalytic residues for Fe(III)-Ent hydrolysis have an all-atom RMSD of 0.78 Å, with Glu26 accounting for the slightly larger RMSD among the active site residues (RMSD = 1.0 Å) (**Figure 2C**). This shift in the orientation of Glu26 disables the H-bonding between Glu26, His74, and water, which was seen in the original Syn-F4 structure (**Figure 2D**). This alteration may prevent formation of the catalytic dyad, and therefore hamper the activity for Fe(III)-Ent hydrolysis.

In addition to the crystallographic studies, we performed biophysical characterizations in solution. Analytical ultracentrifugation and size exclusion chromatography both demonstrated that Syn-F4-Link and the noncovalent dimer of Syn-F4 have similar apparent molecular weights

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in solution (**SI Figure 2A**). Moreover, small-angle X-ray scattering (SAXS) and 1D NMR suggest the solution structures of the covalent and non-covalent dimers are similar (**SI Figure 2B, 3, SI Table 2**). Therefore, we concluded that while biophysical studies suggest Syn-F4 and Syn-F4-Link have similar overall structures, the crystal structure of Syn-F4-Link indicates a change of conformation of the catalytic residues, which may disrupt enzymatic activity.

Molecular dynamics (MD) simulations of Syn-F4-Link

The crystallographic data obtained for Syn-F4-Link does not reveal the structures of the loops connecting the α-helices. There is insufficient electron density, both for the turns connecting successive helices and for the residues engineered to link two monomers into a dimer. This ambiguity concerning inter-helical regions was also true for the parental protein (Syn-F4). In the case of the parental protein, we performed MD simulations, which ultimately showed that the structure with loopleft connections was more stable than the structure with the loopright connections.[26]

To compare the topology of the current structure for Syn-F4-Link with the parental homo-dimeric protein, we constructed two models for the loop connections, indicated by loopleft and loopright in **Figure 3A**. Both models were subjected to MD simulations. The results indicate that for the linked protein, loopleft and loopright have similar overall stabilities (**Figure 3B**). However, in the loopleft simulation, residues 57-101 in the second helix are fragile, allowing a break in the helix in traj09 frame101 (**Figure 3C, D**).

Furthermore, the results suggest that the linked protein with loopleft and loopright (average RMSD: ~2 Å) are less dynamic than the unlinked parental protein with loopleft and loopright (average RMSD: ~3 Å and ~3.5 Å, respectively) (**SI Figure 4**). This change in dynamics may affect the active site and thereby account for the loss of catalytic activity associated with the covalent linkage of two subunits.

We also probed the binding of Syn-F4-Link to Fe(III)-Ent by docking simulation. We found that while the linked protein could bind to Fe(III)-Ent in its hydrophilic pocket, the distribution of Fe^{3+} in Syn-F4-Link is distant from the catalytic residues H74 and H186 (**SI Figure 5**).

Together, the MD simulations suggest that Syn-F4-Link adopts a different loop connection and is less dynamic comparing to the parental Syn-F4 protein. In addition, while Syn-F4-Link still shows binding to Fe(III)-Ent, it could not position the substrate properly in the active site. These differences between Syn-F4-Link and Syn-F4 could impact the catalytic activity of the linked protein.

Covalent linkage of two Syn-F4 chains destroys enzymatic activity

To probe the biological activity of Syn-F4-Link, we transformed a plasmid encoding Syn-F4-Link into Δ*fes* cells and attempted to grow the transformed cells in minimal media. We discovered that the Syn-F4-Link covalent dimer rescues Δ*fes* cells in iron-deficient medium, although at a slower rate than the unlinked dimer, Syn-F4 (**Figure 1B**). Next, we purified the Syn-F4-Link protein and assessed its enzymatic activity *in vitro*. Surprisingly, we found that purified Syn-F4-Link does *not* hydrolyze Fe(III)-Ent. As shown in **Figure 1C**, the parental

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

protein Syn-F4 leads to nearly a complete loss of Fe(III)-Ent (rightmost peak in the chromatogram). However, samples containing Syn-F4-Link retain the same amount of Fe(III)-Ent as the buffer control, with breakdown products observed only at the level as seen for spontaneous hydrolysis in the control. In addition, the soluble lysate of Δ*fes* overexpressing Syn-F4-Link cannot hydrolyze Fe(III)-Ent at a rate faster than background hydrolysis (**SI Figure 6**). Thus, although the linked protein can still rescue Δ*fes* cells *in vivo*, surprisingly, it has no detectable enzymatic activity *in vitro*.

The peptide linker prevents catalytic activity of the linked protein

To further probe the surprising impact of the linker on catalytic activity, we engineered a FactorXa protease site into the linker sequence, allowing us to break the covalent linkage in the purified Syn-F4-Link protein (**Figure 4A**), and thereby enable direct comparison of enzymatic activity pre- and post-cleavage. This new variant is called Syn-F4-Link-FactorXa, and contains two copies of Syn-F4 joined by the 14-residue linker Gly-Gly-Gly-Gly-Ser-Ile-Glu-Gly-Arg-Gly-Gly-Gly-Gly-Ser, containing the (underlined) recognition site for the FactorXa protease.

Enzyme activity assays of this new protein showed that like the original Syn-F4-Link, purified Syn-F4-Link-FactorXa cannot hydrolyze Fe(III)-Ent (**Figure 4B**). However, after cleaving the linker with FactorXa protease, the cleaved product gained partial catalytic activity (**SI Figure 7**). We surmised that the partial (rather than complete) return of activity might indicate that although the linker was cut, most of the dimeric protein was kinetically trapped in the conformation it had assumed prior to proteolytic cleavage. To test this possibility, we disassembled and denatured the cleaved products by running the material over reverse phase HPLC and eluting with organic solvents. We then lyophilized the sample to remove acetonitrile and refolded the protein in a non-denaturing buffer (75mM HEPES, 1M NaCl, pH 8.0). We found that refolding the cleaved products fully restored the enzymatic activity (**Figure 4B**). These results demonstrate that covalently linking two chains of Syn-F4 – either with the original decapeptide linker or with the modified 14 residue peptide linker with a FactorXa cleavage site – destroys the catalytic activity of the *de novo* protein.

The linked protein upregulates enterobactin biosynthesis in *E. coli*

The parental Syn-F4 protein rescues Δ*fes* cells *in vivo*, and *in vitro* it catalyzes the same reaction as the natural Fes enzyme. On the other hand, Syn-F4-Link rescues Δ*fes* cells *in vivo*, but *fails* to catalyze the hydrolysis of Fe(III)-Ent. Therefore, the linked protein must enable the growth of Δ*fes* cells by some alternative – perhaps regulatory – mechanism. To probe this possibility, we performed global transcriptomic analysis of *E. coli* cells expressing Syn-F4 or Syn-F4-Link and compared these results to cells expressing the natural Fes protein. RNA-seq analysis of Δ*fes* cells expressing Syn-F4-Link revealed that expression of the covalently linked dimer upregulates enterobactin biosynthesis genes, including *entA*, *entB*, *entE*, *entF*, and *entH* (**Figure 5**). Importantly, this upregulation is also observed in a pseudo-wildtype strain (Keio parent strain, rather than Keio Δ*fes*) expressing the natural Fes protein. This last result

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

demonstrates that expression of Syn-F4-Link – not just background starvation for iron – causes the observed upregulation of genes related to enterobactin biosynthesis and iron assimilation (**SI Figure 8**).

These transcriptomic results indicate that Syn-F4-Link leads to overproduction of enterobactin. Therefore, expression of Syn-F4-Link in Δ*fes* cells, which cannot degrade the siderophore, should lead to an accumulation of Fe(III)-Ent. To test this expectation, we grew cultures of Δ*fes* cells overexpressing Syn-F4-Link in iron-enriched media and compared them to the same cells harboring the natural *fes* gene. As shown in **Figure 6A**, Δ*fes* cells expressing Syn-F4-Link produce red cultures, consistent with the accumulation of Fe(III)-Ent. Moreover, quantification of the siderophore by HPLC showed that Syn-F4-Link induces the synthesis of substantially more Fe(III)-Ent than does the natural Fes protein (**Figure 6B)**.

The parental homodimeric *de novo* protein Syn-F4 has two biological functions

After discovering that Syn-F4-Link is not enzymatically active, but instead rescues Δ*fes* cells by a regulatory mechanism that induces overproduction of enterobactin, we questioned whether the enzymatically active parental protein, Syn-F4, might also have a secondary activity upregulating the biosynthesis of enterobactin. To address this question, we analyzed the effect of Syn-F4 expression on the transcriptome. The results, shown in **Figure 7 and SI Figure 9**, indicate that Syn-F4 does indeed induce increased expression of enterobactin biosynthesis genes. Thus, the original Syn-F4 homodimer is a bifunctional *de novo* protein. It can rescue Δ*fes* by performing the deleted enzymatic function to hydrolyze Fe(III)-Ent, and it also alters gene expression and upregulates enterobactin biosynthesis.

DISCUSSION

Several years ago, we described Syn-F4 as the first *de novo* protein that shares no ancestry with natural enzymes, but nonetheless provides a life-sustaining catalytic function *in vitro* and *in vivo*.[25] Subsequent determination of the crystal structure of Syn-F4 revealed a 4-helix bundle comprising two α-helical hairpins arranged in a *syn* (as compared to *anti*) topology.[26] Because this orientation placed the C-terminus of one chain close to the N-terminus of the second chain, we reasoned that linking these termini would enhance (or at least not destroy) activity by creating a single chain that could fold into its active conformation without paying the entropic price of dimerization.

As described above, the linked protein was constructed and shown to fold into a structure that closely resembled the original homodimer. Moreover, as expected, it rescued the growth of Δ*fes* cells in minimal media. Surprisingly, however, the purified Syn-F4-Link protein was *not* enzymatically active and failed to hydrolyze ferric enterobactin. Instead, we found that Syn-F4-Link rescues Δ*fes* cells by using a regulatory, rather than enzymatic, mechanism.

This surprising change in the mechanism of biological rescue led to a closer examination of the structure and dynamics of Syn-F4-Link relative to its progenitor Syn-F4. Specifically, MD simulations of Syn-F4-Link with either of two possible loop connections (loopleft or loopright)

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

showed that a loopleft connection like that in the unlinked Syn-F4 homodimer would cause the second helix to be unstable. This finding suggests that the engineered covalent linkage favors folding into the inactive loopright conformation. In this loopright conformation, the hydrogen-bonding network among the active site residues is disrupted, which would account for the loss of catalytic activity. In addition, the MD simulation indicated Syn-F4-Link was less dynamic than the unlinked Syn-F4, which would hinder the ability of the linked chain from sampling the catalytically active confirmation. This finding suggests that for this *de novo* enzyme, as for many natural enzymes, structural dynamics are important for catalytic activity.[27,28]

It was surprising that linking the two chains abrogated enzymatic activity. It would be even more surprising if this linkage simultaneously generated an entirely new regulatory activity that was not present in the parental homodimer. Therefore, we questioned whether the Syn-F4 homodimer might also upregulate enterobactin biosynthesis. As shown in **Figure 7**, this turned out to be true: The parental Syn-F4 homodimer – like the engineered Syn-F4-Link protein – upregulates enterobactin biosynthesis. This regulatory function of Syn-F4 was not observed previously because it is latent when the *de novo* rescuer acts as an enzyme, but was revealed when enzymatic activity was destroyed by engineering a covalent linkage joining the two chains.

Our initial isolation of numerous *de novo* proteins that rescue Δ*fes* cells,[18] followed by subsequent genetic, biochemical and structural characterization of the linked and unlinked versions of Syn-F4 reveal that it is not difficult to find non-natural proteins that enable the growth of Δ*fes* cells in iron-limited conditions. Importantly, these studies reveal that novel proteins that did not arise in nature can use either (or both) catalytic or regulatory mechanisms to sustain the growth of living organisms.

While Syn-F4-Link significantly upregulates enterobactin biosynthesis, this alone is not sufficient to explain how the cell acquires sufficient Fe^{3+} to survive in iron deficient media. We hypothesize that binding of Syn-F4-Link to Fe(III)-Ent could disrupt the precise coordination between enterobactin and Fe^{3+}, thereby releasing iron for cellular metabolism. Likely, this process is inefficient. Therefore, only when Fe(III)-Ent is overexpressed could the cell gain enough iron for survival, which explains the importance of Syn-F4-Link in upregulating the biosynthesis of Fe(III)-Ent. Future studies characterizing the binding of enterobactin to iron under the effect of Syn-F4-Link could enhance our understanding of the *de novo* protein's life-sustaining mechanism.

ACKNOWLEDGEMENTS

We thank J. Eng for assistance with mass spectrometry and I. Pelczer for assistance with protein NMR. We thank all members of the Hecht lab for helpful discussion. This work was funded by NSF MCB-1947720.

AUTHOR CONTRIBUTIONS

G.L., S.T. and, M.H.H. initiated the project idea. S.T. constructed Syn-F4-Link. G.L. performed *in vivo* rescue, *in vitro* protein biophysics and enzymology, RNA-seq, and metabolomics assay. M.N., K.K., and R.A. conducted crystallography and SAXS. K.U.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

conducted MD and docking simulations. G.L. and M.H.H. drafted the manuscript, with input from all the authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MATERIALS AND METHODS

Gene construction

To generate the Syn-F4-link construct, in which two Syn-F4 coding sequences are joined by a flexible linker (GGGGSGGGGS), two overlapping DNA fragments were amplified by PCR using the Syn-F4-containing plasmid (Syn-F4 in p3glar) as the template. The first fragment was amplified with primers F-1 and F-2, introducing the linker sequence at the 5′ end of the Syn-F4 coding region. The second fragment was amplified using primers V-1 and V-2; this removed the stop codon of Syn-F4 and appended the same linker sequence to its 3′ end. Additionally, the second fragment included vector-derived sequences flanking the Syn-F4-linker region, allowing it to serve as the backbone for plasmid assembly. The two fragments were assembled using Gibson Assembly (NEB), transformed into *E. coli* DH5α, and verified by colony PCR and Sanger sequencing.

F-1	CATCGCGGCGGCGGCGGCTCGGGCGGCGGCGGCTCGATGTATGGCAAGTTG AACCAGC
F-2	GCTCAGCTAATTAAGTTAGCGATGTACAAG
V-1	CTTGTACATCGCTAACTTAATTAGCTGAGC
V-2	CATACATCGAGCCGCCGCCGCCCGAGCCGCCGCCGCCGCGATGTACAAGGT TGTCCAG

Protein expression and purification

Syn-F4 (K4T), Syn-F4-Link (K4Tx2) and Syn-F4-Link (FactorXa) genes were cloned into the expression vector p3glar, with chloramphenicol resistance. Proteins were overexpressed in BL21 (New England Biolabs), where expression was induced with 0.1 mM IPTG at OD_{600} ~0.5 and incubated at 18 ^{o}C for 18 hours. Cells were harvested by centrifugation, resuspended in purification buffer (75 mM HEPES, 1 M NaCl, pH 8.0), and lysed by sonication (10 seconds on, 50 seconds off, total sonication time of 3 minutes 30 seconds). The soluble lysate was separated from the insoluble fraction by centrifugation and filtered through a 0.45 µm syringe filter (Millipore). The soluble lysate containing overexpressed protein was loaded onto a nickel column (HisTrap HP, Cytiva), washed with purification buffer containing 50 mM imidazole, and eluted with 375 mM imidazole. The eluted fraction was then loaded on a size exclusion chromatographic column (HiLoad 26/600 Superdex 75 pg, Cytiva) and fractions corresponding to the expected molecular weight of the protein were combined and stored at -80 ^{o}C. The purity of protein was confirmed by SDS-PAGE.

Crystallography and structure determination

For crystallization, the Syn-F4-Link (K4Tx2) protein was purified by immobilized metal ion affinity chromatography with COSMOGEL 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%

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

glycerol and 250 mM imidazole). The protein was further purified by anion exchange chromatography (25 mM Tris-HCl buffer (pH 8.5) containing 10% glycerol with a linear gradient of NaCl from 0 to 1.5 M) with a RESOURCE Q 6 mL column (Cytiva) and size-exclusion chromatography (25 mM Tris-HCl buffer (pH 9.0) containing 300 mM NaCl, 10% glycerol and 200 mM Arg-HCl with a Superdex 75 Increase 10/300 GL column (Cytiva).

Plate-like crystals were obtained in a drop composed of 1 μL of the protein solution (9.2 mg/mL) and 1 μL of the reservoir solution (110 mM sodium malonate, pH 7.0, 9.5% w/v polyethylene glycol 3,350) by the hanging drop vapor diffusion method against 500 μL of the reservoir solution with a 24-well plate at 20°C in a week.

X-ray diffraction data were collected at the Photon Factory, BL-17A (KEK, Tsukuba, Japan) with a EIGER X 16M detector (Dectris, Baden, Switzerland). The data collection was carried out at 95 K with 30% glycerol as a cryoprotectant. Diffraction data were processed with the programs XDS[29,30] and AIMLESS[31]. The structure was solved by molecular replacement method using Phaser with a model structure of Syn-F4 (K4T) (PDB ID: 8H7D)[26]. The model was corrected with the program COOT[32] and was refined with the program REFMAC5[33] in the CCP4 suite[34]. The quality of the model was inspected by the programs PROCHECK[35] and MolProbity[36]. All statistics are presented in **SI Table 1**. The atomic coordinates and the structure factors have been deposited in the Protein Data Bank with the accession code 9UGR.

Small-angle X-ray scattering (SAXS)

SAXS measurements were performed for samples (~5 mg/mL) of Syn-F4-Link (K4Tx2) and 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 (λ = 1.3 Å) at the Photon Factory BL-10C beamline[37] (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[38], 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 [39] using AUTORG in ATSAS[40] with SAngler[38]. Forward scattering intensity normalised 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 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 modelling programs in the ATSAS program suite[40]. Calculations of rapid *ab initio* shape determination were performed ten times by DAMMIF[41] without a symmetry constraint, and the generated models were aligned and averaged by DAMAVER[42]. The averaged model was modified with the fixed core by DAMSTART and further refinement of the model was performed by DAMMIN[43]. Superimposing the dummy atom model and crystal structures was performed by SUPCOMB[44]. The SAXS data and dummy-atom models of Syn-F4-Link (K4Tx2) and Syn-F4 (K4T) have been deposited in Small Angle Scattering Biological Data Bank[45] (SASBDB accession codes: xxxxxxx, yyyyyyyy).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Molecular dynamics and docking simulations

All-atom MD simulations were performed for the two Syn-F4-Link constructs with different loop connections (loop-left and loop-right) in a rectangular periodic box. The initial conformations of the loop regions were generated by MODELLAR[46]. The force field of the protein was taken from Amber ff14SB[47], and the TIP3P water model[48] was used. Potassium and chloride ions were added to neutralize the system and achieve a salt concentration of 100 mM. Following energy minimization, the system was heated linearly from 1 K to 310 K for 1 ns. The initial velocities of atoms were randomly assigned according to the Maxwell-Boltzmann distribution. The system was then equilibrated for 10 ns under NPT conditions (1 bar, 310 K). Production run was conducted for 1.0 μs. Snapshots were saved every 100 ps, yielding 10,000 frames per simulation. To enhance sampling, this procedure was repeated ten times with different initial velocities, generating a total of ten independent trajectories. Finally, 100,000 snapshots were collected for subsequent analyses, including secondary structure population and root-mean-square deviation (RMSD) from the initial structure. A subset of 1,000 snapshots, extracted every 10 ns, was used for the subsequent docking simulations. All simulations and analyses were carried out using the Amber18 and AmberTools18 software packages[49].

Enzymatic assay for hydrolysis of Fe(III)-Ent

Fe(III)-Ent was prepared as described by Donnelly *et al.*[25] To assay catalytic activity, 25 μM of purified protein and 120 μM Fe(III)-Ent were incubated in 75 mM HEPES, 1 M NaCl, pH 8.0 at 37 ^{o}C. Aliquots were collected at various time points, quenched by 0.5 volume of 2.5 M HCl in methanol, and Fe(III)-Ent and its hydrolysis products were extracted by 2 volumes of ethyl acetate. The organic layer was analyzed by reverse-phase HPLC, and the enterobactin and hydrolysis products were monitored by absorbance at 316 nm.

RNA-seq

Δ*fes* or Keio parent *E. coli* cells from the Keio collection[50] harboring Syn-F4, Syn-F4-Link, natural *fes* from the ASKA collection[51] or empty vector were incubated in M9-glucose minimal media at 37 ^{o}C. Cells were collected at OD_{600} 0.2-0.5 and resuspended in 0.5 mL DEPC treated water (Invitrogen) and kept frozen. The total RNA from cells were released by cryomill and purified by Direct-zol RNA Miniprep kits (Zymo Research). cDNA libraries were prepared by ribo-depletion followed by RNA-Seq Directional Library Prep on Apollo 324 Robot. Next-generation sequencing of the library was done with NovaSeq SP 100nt Lane v1.5 (Princeton University). The sequencing data was processed on Galaxy.[52] The quality of RNA-seq data were tested by FastQC. The data were then processed by Trimmomatic to trim the adapters[53], mapped to the *E. coli* K12 strain BW25113 reference genome[54] using BWA-MEM[55], and the genes counts were quantified by htseq-count[56]. Differential expression analysis was conducted by DESeq2[57] with R language. The volcano plots were created by EnhancedVolcano package[58] on R.

Quantification of Fe(III)-Ent in culture

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Δ*fes* cells expressing either Syn-F4-Link, Syn-F4, natural *fes* from the ASKA collection or S824 (a *de novo* 4-helix bundle protein with no biological function)[59] were incubated in M9-glucose minimal media supplemented with 100 μM $FeCl_3$ at 37 ^{o}C. After turbidity was observed, the cells and liquid culture were separated by centrifugation. Liquid culture was acidified by hydrochloric acid, and enterobactin was extracted by ethyl acetate. The organic layers were analyzed by reverse-phase HPLC to quantify the amount of enterobactin. The elution peak corresponding to enterobactin was collected and analyzed on mass spectrometry to confirm the identity of the molecule.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1 Hecht, M. H., Zarzhitsky, S., Karas, C. & Chari, S. Are natural proteins special? Can we do that? *Current Opinion in Structural Biology* **48**, 124-132 (2018). https://doi.org/https://doi.org/10.1016/j.sbi.2017.11.009

2 Beasley, J. R. & Hecht, M. H. Protein Design: The Choice of de Novo Sequences*. *Journal of Biological Chemistry* **272**, 2031-2034 (1997). https://doi.org/https://doi.org/10.1074/jbc.272.4.2031

3 Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320-327 (2016). https://doi.org/10.1038/nature19946

4 Kuhlman, B. *et al.* Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **302**, 1364-1368 (2003). https://doi.org/doi:10.1126/science.1089427

5 Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089-1100 (2023). https://doi.org/10.1038/s41586-023-06415-8

6 Ingraham, J. B. *et al.* Illuminating protein space with a programmable generative model. *Nature* **623**, 1070-1078 (2023). https://doi.org/10.1038/s41586-023-06728-8

7 Sakuma, K. *et al.* Design of complicated all-α protein structures. *Nature Structural & Molecular Biology* **31**, 275-282 (2024). https://doi.org/10.1038/s41594-023-01147-9

8 Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-195 (2008). https://doi.org/10.1038/nature06879

9 Shen, H. *et al.* De novo design of pH-responsive self-assembling helical protein filaments. *Nature Nanotechnology* **19**, 1016-1021 (2024). https://doi.org/10.1038/s41565-024-01641-1

10 Madani, A. *et al.* Large language models generate functional protein sequences across diverse families. *Nature Biotechnology* (2023). https://doi.org/10.1038/s41587-022-01618-2

11 Langan, R. A. *et al.* De novo design of bioactive protein switches. *Nature* **572**, 205-210 (2019). https://doi.org/10.1038/s41586-019-1432-8

12 Goverde, C. A. *et al.* Computational design of soluble and functional membrane protein analogues. *Nature* **631**, 449-458 (2024). https://doi.org/10.1038/s41586-024-07601-y

13 Xu, C. *et al.* Computational design of transmembrane pores. *Nature* **585**, 129-134 (2020). https://doi.org/10.1038/s41586-020-2646-5

14 Butterfield, G. L. *et al.* Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **552**, 415-420 (2017). https://doi.org/10.1038/nature25157

15 Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. Protein design by binary patterning of polar and nonpolar amino acids. *Science* **262**, 1680 (1993). https://doi.org/10.1126/science.8259512

16 West, M. W. *et al.* De novo amyloid proteins from designed combinatorial libraries. *Proceedings of the National Academy of Sciences* **96**, 11211 (1999). https://doi.org/10.1073/pnas.96.20.11211

17 Hoegler, K. J. & Hecht, M. H. A de novo protein confers copper resistance in Escherichia

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

coli. *Protein Sci* **25**, 1249-1259 (2016). https://doi.org/10.1002/pro.2871

18 Fisher, M. A., McKinley, K. L., Bradley, L. H., Viola, S. R. & Hecht, M. H. De Novo Designed Proteins from a Library of Artificial Sequences Function in Escherichia Coli and Enable Cell Growth. *PLoS ONE* **6** (2011). https://doi.org/10.1371/journal.pone.0015364

19 Digianantonio, K. M. & Hecht, M. H. A protein constructed de novo enables cell growth by altering gene regulation. *Proceedings of the National Academy of Sciences* **113**, 2400 (2016). https://doi.org/10.1073/pnas.1600566113

20 Digianantonio, K. M., Korolev, M. & Hecht, M. H. A Non-natural Protein Rescues Cells Deleted for a Key Enzyme in Central Metabolism. *ACS Synthetic Biology* **6**, 694-700 (2017). https://doi.org/10.1021/acssynbio.6b00336

21 Raymond, K. N., Dertz, E. A. & Kim, S. S. Enterobactin: An archetype for microbial iron transport. *Proceedings of the National Academy of Sciences* **100**, 3584 (2003). https://doi.org/10.1073/pnas.0630018100

22 Lin, H., Fischbach, M. A., Liu, D. R. & Walsh, C. T. In Vitro Characterization of Salmochelin and Enterobactin Trilactone Hydrolases IroD, IroE, and Fes. *Journal of the American Chemical Society* **127**, 11075-11084 (2005). https://doi.org/10.1021/ja0522027

23 Brickman, T. J. & McIntosh, M. A. Overexpression and purification of ferric enterobactin esterase from Escherichia coli. Demonstration of enzymatic hydrolysis of enterobactin and its iron complex. *Journal of Biological Chemistry* **267**, 12350-12355 (1992).

24 Joyce, A. R. *et al.* Experimental and Computational Assessment of Conditionally Essential Genes in *Escherichia coli*. *Journal of Bacteriology* **188**, 8259 (2006). https://doi.org/10.1128/JB.00740-06

25 Donnelly, A. E., Murphy, G. S., Digianantonio, K. M. & Hecht, M. H. A de novo enzyme catalyzes a life-sustaining reaction in Escherichia coli. *Nature Chemical Biology* **14**, 253-255 (2018). https://doi.org/10.1038/nchembio.2550

26 Kurihara, K. *et al.* Crystal structure and activity of a de novo enzyme, ferric enterobactin esterase Syn-F4. *Proc Natl Acad Sci U S A* **120**, e2218281120 (2023). https://doi.org/10.1073/pnas.2218281120

27 Benkovic, S. J. & Hammes-Schiffer, S. A Perspective on Enzyme Catalysis. *Science* **301**, 1196-1202 (2003). https://doi.org/doi:10.1126/science.1085515

28 Klinman, J. P. & Kohen, A. Hydrogen Tunneling Links Protein Dynamics to Enzyme Catalysis. *Annual Review of Biochemistry* **82**, 471-496 (2013). https://doi.org/https://doi.org/10.1146/annurev-biochem-051710-133623

29 Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010). https://doi.org/10.1107/S0907444909047337

30 Brehm, W., Trivino, J., Krahn, J. M., Uson, I. & Diederichs, K. XDSGUI: a graphical user interface for XDS, SHELX and ARCIMBOLDO. *J Appl Crystallogr* **56**, 1585-1594 (2023). https://doi.org/10.1107/S1600576723007057

31 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallographica Section D-Biological Crystallography* **69**, 1204-1214 (2013).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

https://doi.org/10.1107/S0907444913000061

32 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010). https://doi.org/10.1107/S0907444910007493

33 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367 (2011). https://doi.org/10.1107/S0907444911001314

34 Agirre, J. *et al.* The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr D Struct Biol* **79**, 449-461 (2023). https://doi.org/10.1107/S2059798323003595

35 Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* **26**, 283-291 (1993). https://doi.org/10.1107/S0021889892009944

36 Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315 (2018). https://doi.org/10.1002/pro.3330

37 Shimizu, N. *et al.* BL-10C, the small-angle x-ray scattering beamline at the photon factory. *AIP Conf. Proc.* **2054**, 060041 (2019). https://doi.org/10.1063/1.5084672

38 Shimizu, N. *et al.* Software development for analysis of small-angle X-ray scattering data. *AIP Conf. Proc.* **1741**, 050017 (2016). https://doi.org/10.1063/1.4952937

39 Glatter, O. & Kratky, O. *Small-Angle X-Ray Scattering*. (Academic Press, 1982).

40 Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **50**, 1212-1225 (2017). https://doi.org/10.1107/S1600576717007786

41 Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* **42**, 342-346 (2009). https://doi.org/10.1107/S0021889809000338

42 Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* **36**, 860-864 (2003). https://doi.org/10.1107/S0021889803000268

43 Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal* **76**, 2879-2886 (1999). https://doi.org/10.1016/S0006-3495(99)77443-6

44 Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* **34**, 33-41 (2000).

45 Kikhney, A. G., Borges, C. R., Molodenskiy, D. S., Jeffries, C. M. & Svergun, D. I. SASBDB: Towards an automatically curated and validated repository for biological scattering data. *Protein Sci* **29**, 66-75 (2020). https://doi.org/10.1002/pro.3731

46 Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **54**, 5.6.1-5.6.37 (2016). https://doi.org/https://doi.org/10.1002/cpbi.3

47 Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Parameters from ff99SB. *Journal of Chemical Theory and Computation* **11**, 3696-3713 (2015). https://doi.org/10.1021/acs.jctc.5b00255

48 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983). https://doi.org/10.1063/1.445869

49 Case, D. *et al*. *Amber 2018*. (2018).

50 Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* **2**, 2006.0008 (2006). https://doi.org/https://doi.org/10.1038/msb4100050

51 Kitagawa, M. *et al.* Complete set of ORF clones of Escherichia coli ASKA library (A Complete S et of E. coli K -12 ORF Archive): Unique Resources for Biological Research. *DNA Research* **12**, 291-299 (2005). https://doi.org/10.1093/dnares/dsi012

52 Community, T. G. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research* **52**, W83-W94 (2024). https://doi.org/10.1093/nar/gkae410

53 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014). https://doi.org/10.1093/bioinformatics/btu170

54 Grenier, F., Matteau, D., Baby, V. & Rodrigue, S. Complete Genome Sequence of Escherichia coli BW25113. *Genome Announcements* **2**, 10.1128/genomea.01038-01014 (2014). https://doi.org/doi:10.1128/genomea.01038-14

55 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* (2013).

56 Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2014). https://doi.org/10.1093/bioinformatics/btu638

57 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8

58 Blighe, K., Rana, S. & Turkes, E. *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.24.0*, <https://github.com/kevinblighe/EnhancedVolcano> (2024).

59 Wei, Y., Kim, S., Fela, D., Baum, J. & Hecht, M. H. Solution structure of a de novo protein from a designed combinatorial library. *Proceedings of the National Academy of Sciences* **100**, 13270 (2003). https://doi.org/10.1073/pnas.1835644100

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

FIGURES

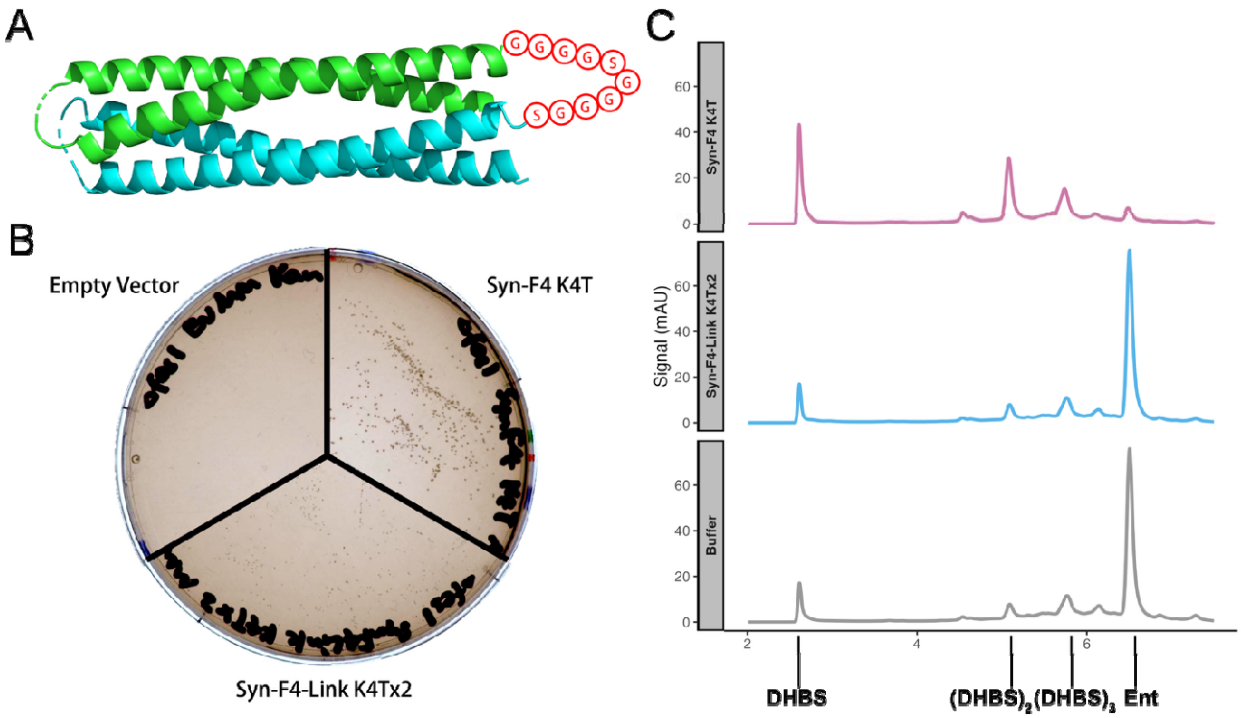


Figure 1. Syn-F4-Link rescues Δ*fes* but cannot hydrolyze Fe(III)-Ent *in vitro*.
(A) Design of Syn-F4-Link. The linker residues are shown in red circles.
(B) Rescue of *E. coli* growth by Syn-F4, Syn-F4-Link, with empty vector as a negative control.
(C) Hydrolysis of Fe(III)-Ent by purified Syn-F4, Syn-F4-Link (and buffer as negative control) monitored by reverse phase HPLC. Enterobactin (Ent), and its hydrolysis products (DHBS, $(DHBS)_2$ and $(DHBS)_3$) are labeled.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

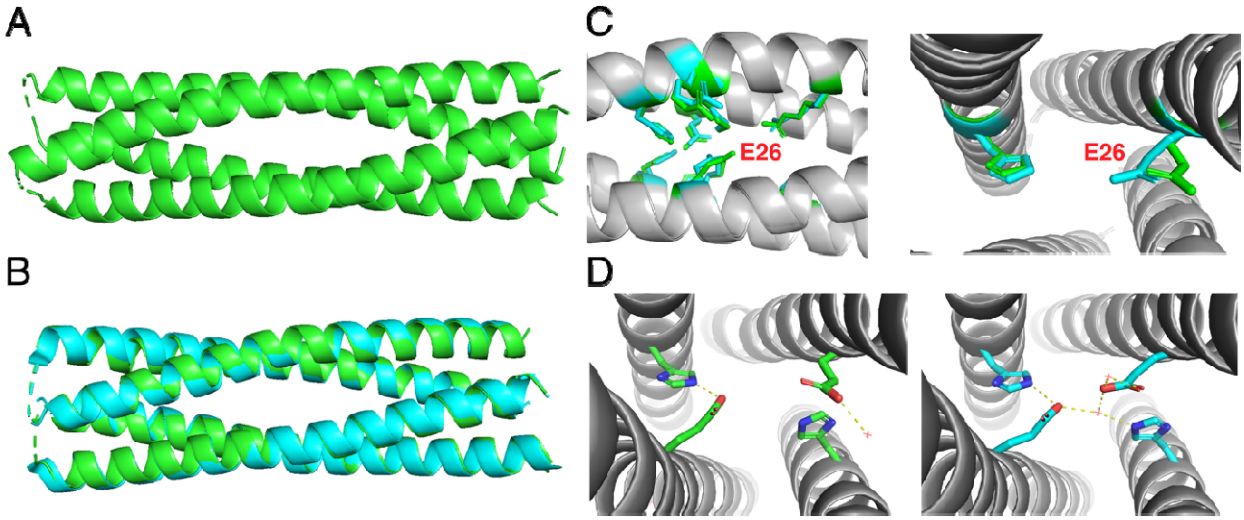


Figure 2. Structure of Syn-F4-Link.

(A) Crystal structure of Syn-F4-Link K4Tx2.

(B) Syn-F4-Link (green) and Syn-F4 (cyan) have an overall C-alpha RMSD of 0.45 Å.

(C) Catalytic residues, with Syn-F4-Link residues in green and Syn-F4 residues in cyan. Glu26 residue is highlighted in red.

(D) H-bond and salt bridge network among Glu26, His74 and water molecules in Syn-F4-Link (left) and unlinked Syn-F4 (right).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

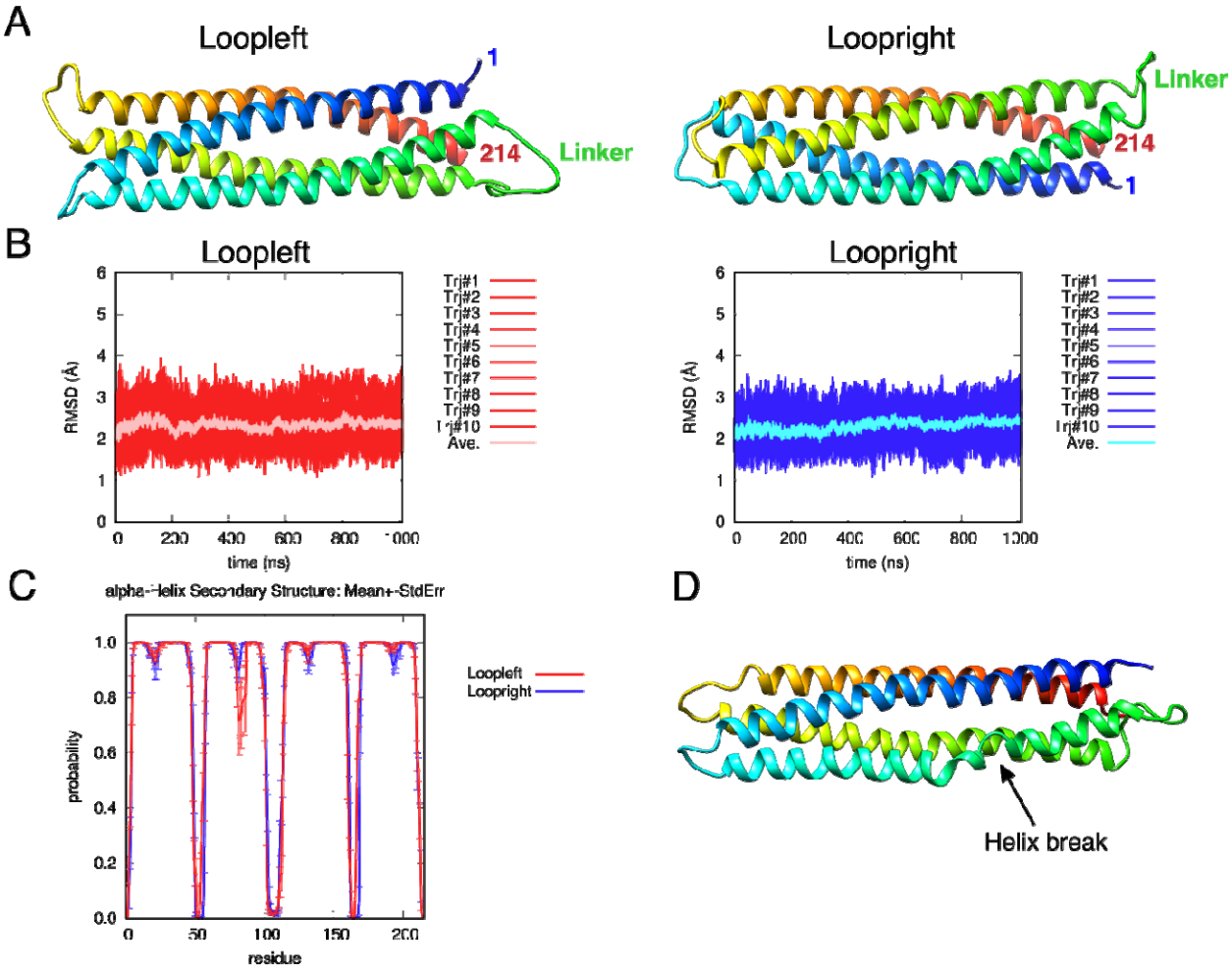


Figure 3. MD simulations of Syn-F4-Link.

(A) Syn-F4-Link with different loop connections. N-terminus of the protein is in blue, and C-terminus is in red. Residue 1 and 214 in the chains are indicated. The covalent flexible linker is shown in green.

(B) RMSD of Syn-F4-Link structures from the original models in the simulation.

(C) Probability of forming an α-helical structure for each residue in MD simulations.

(D) The second helix in loopleft conformation is fragile.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

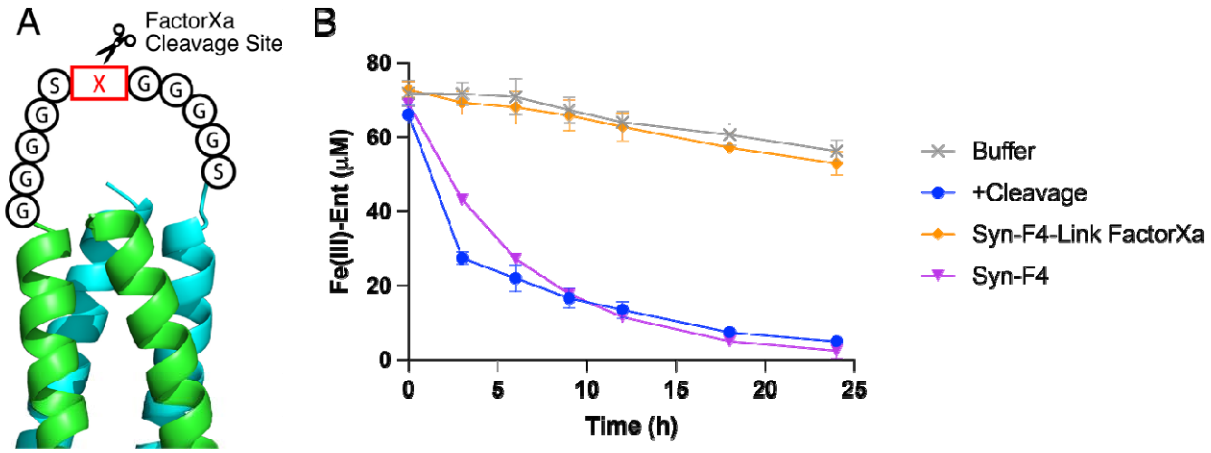


Figure 4. Cleavage and refolding of the linker of Syn-F4-Link derivative restores the Fe(III)-Ent hydrolysis activity.

(A) Design of Syn-F4-Link with FactorXa cleavable site.

(B) Hydrolysis of Fe(III)-Ent over 24 hours by Syn-F4-Link without cleavage, with cleavage and refolding, unlinked Syn-F4 and buffer.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

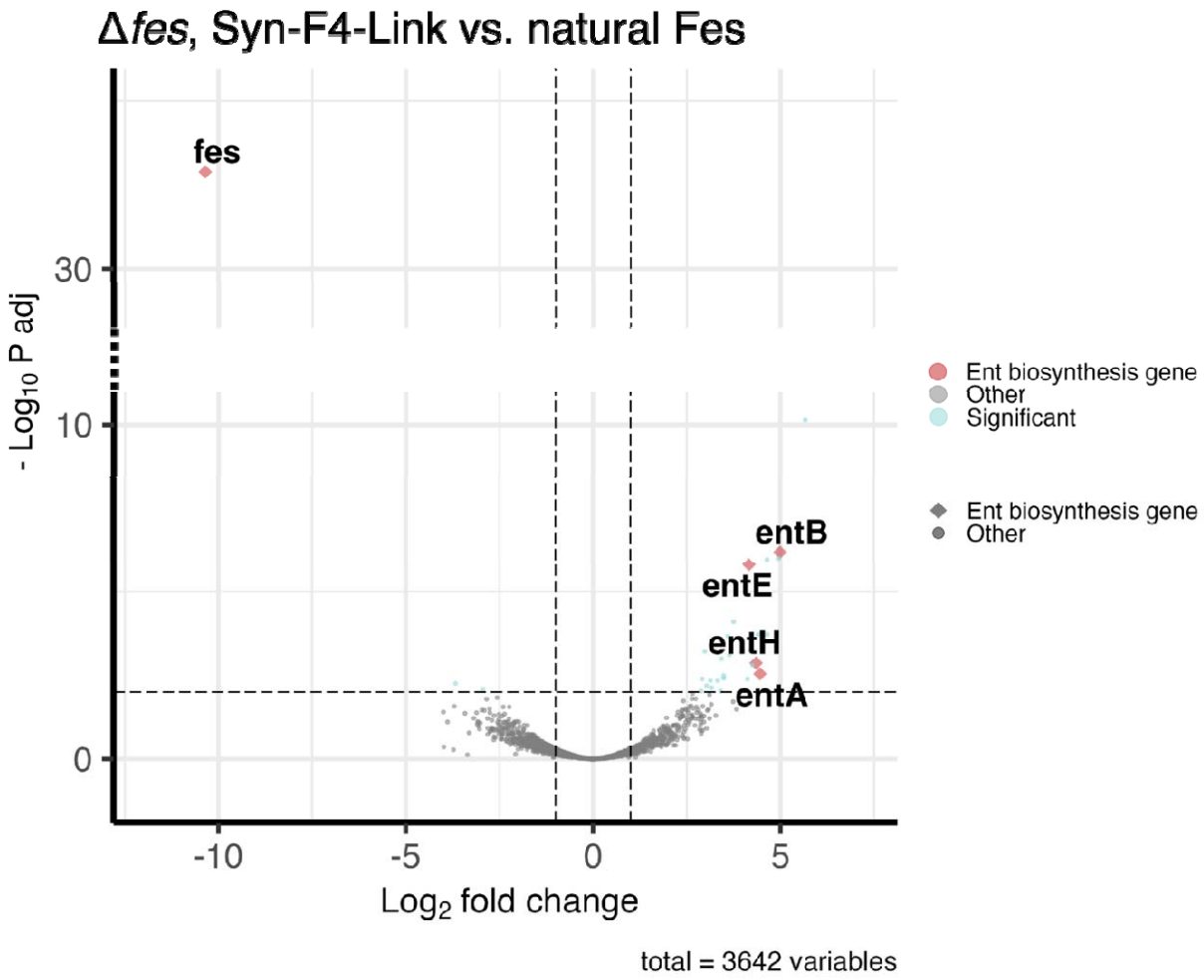


Figure 5. Syn-F4-Link significantly upregulates expression of enterobactin synthesis genes. The effect of Syn-F4-Link relative to natural Fes on the transcriptome of Δ*fes* while growing in minimal media is shown. Enterobactin synthesis-related genes, and *fes* gene as an internal control, are highlighted.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

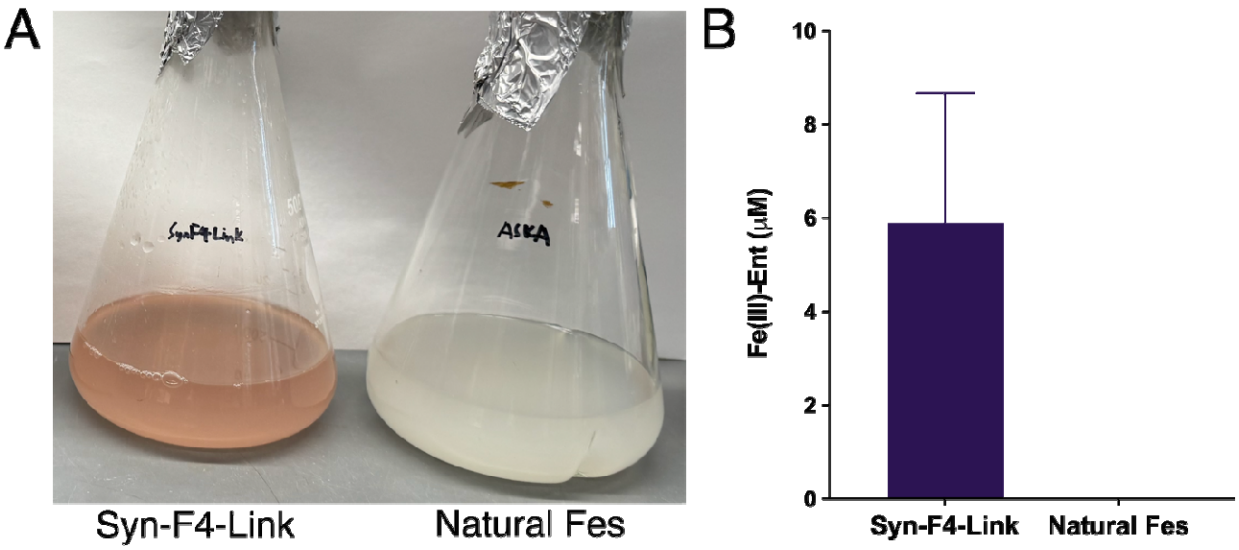


Figure 6. Syn-F4-Link upregulates enterobactin production.

(A) Growth of Δ*fes* Syn-F4-Link and Δ*fes* natural Fes in minimal media supplied with 100 μM Fe^{3+}.

(B) Enterobactin abundance in media determined by reverse-phase HPLC.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

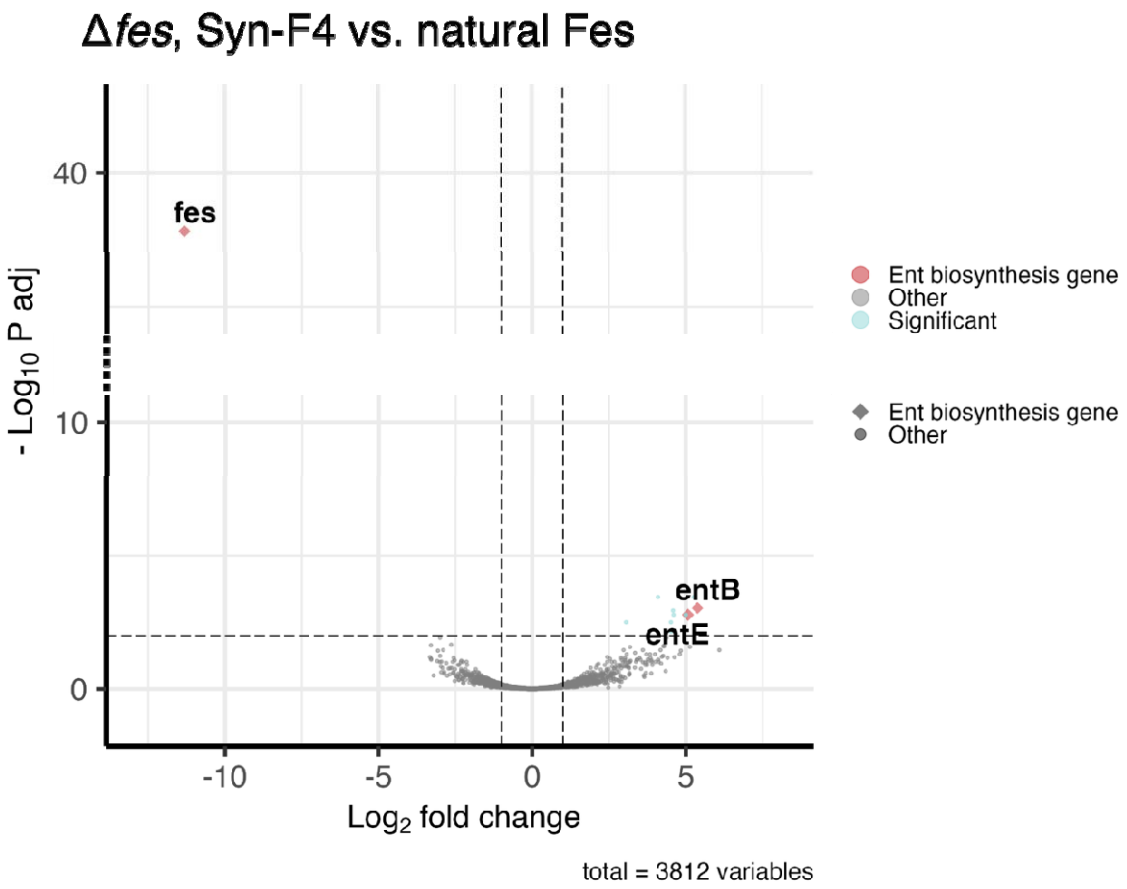
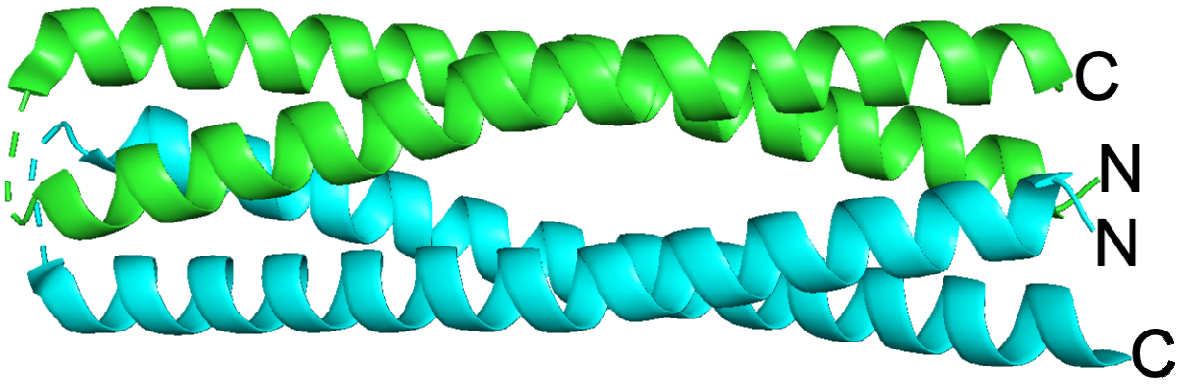


Figure 7. Syn-F4 significantly upregulates expression of enterobactin synthesis genes. The effect of Syn-F4 relative to natural Fes on the transcriptome of Δ*fes* in minimal media is shown. *entB* and *entE* genes, and the internal control gene *fes*, are highlighted.

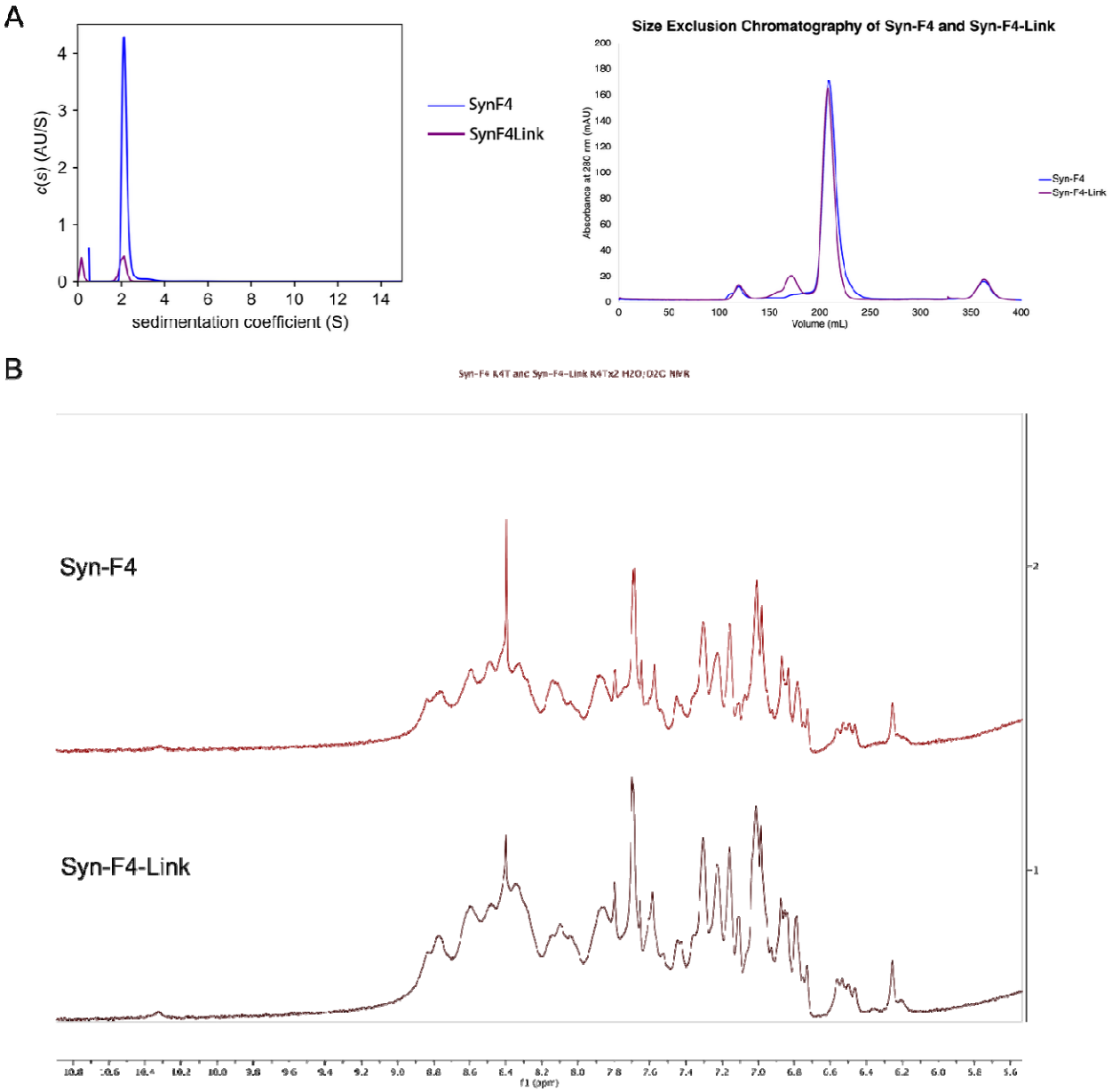
bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Information



SI Figure 1. Crystal structure of Syn-F4 K4T (PDB 8H7D). N- and C-termini of each chain (green and cyan) are labeled.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

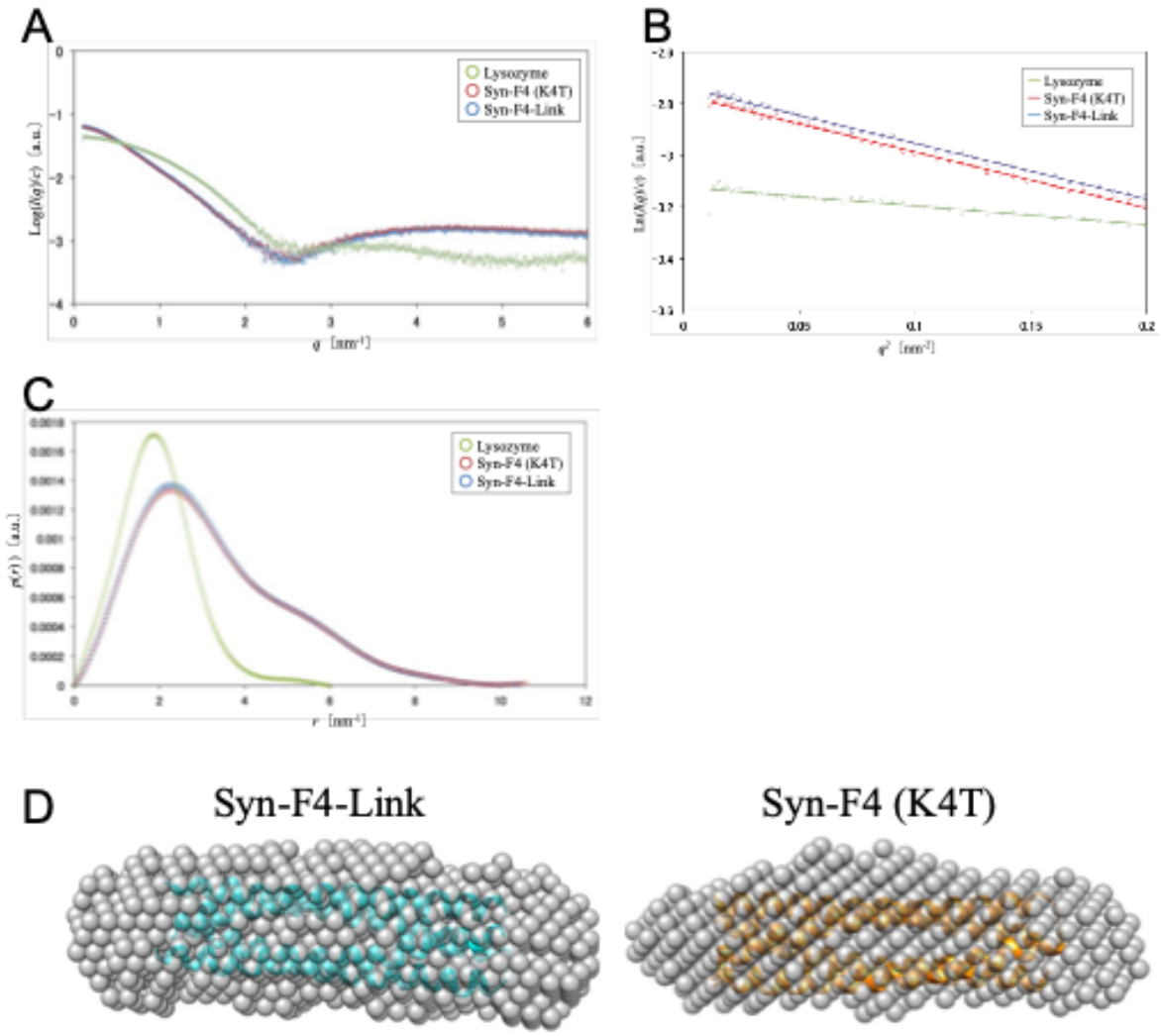


SI Figure 2. Biophysical characterization of Syn-F4-Link.

(A) Analytical ultracentrifuge (left) and size exclusion chromatography (right) suggest Syn-F4-Link and Syn-F4 has similar apparent molecular weight in solution.

(B) 1D NMR of Syn-F4 and Syn-F4-Link. The amide proton and tryptophan peaks are shown.

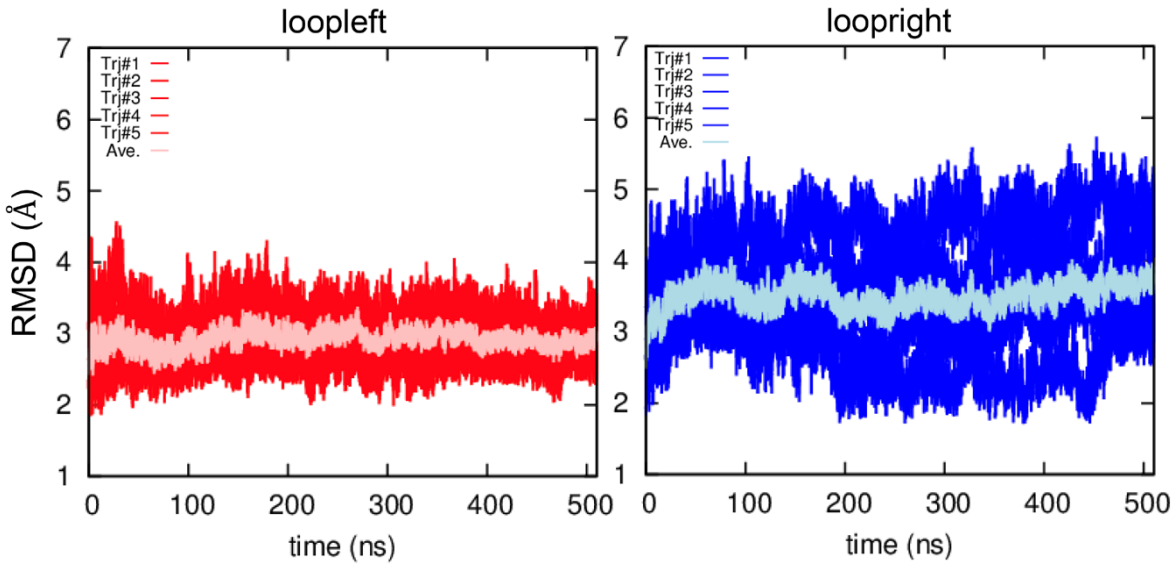
bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



SI Figure 3. Small-angle X-ray scattering (SAXS) analyses of Syn-F4-Link and Syn-F4 (K4T).

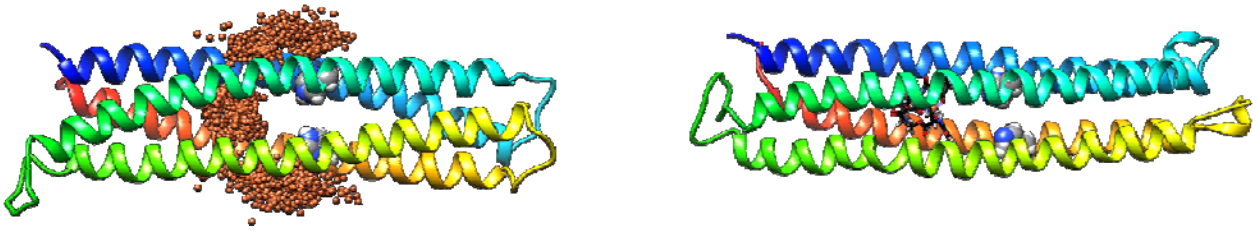
(A) Concentration-normalized scattering intensities. (B) Guinier plots. (C) Pair-distance distribution functions. Lysozyme was used as a molecular mass reference standard. SAXS analysis result and estimated molecular mass are summarized in SI Table 2. (C) Dummy atom models of Syn-F4-Link and Syn-F4 (K4T) calculated from the SAXS data using *ab initio* modeling programs, DAMMIF and DAMMIN. For reference, the crystal structures (ribbon representation) are superimposed on the dummy atom models.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



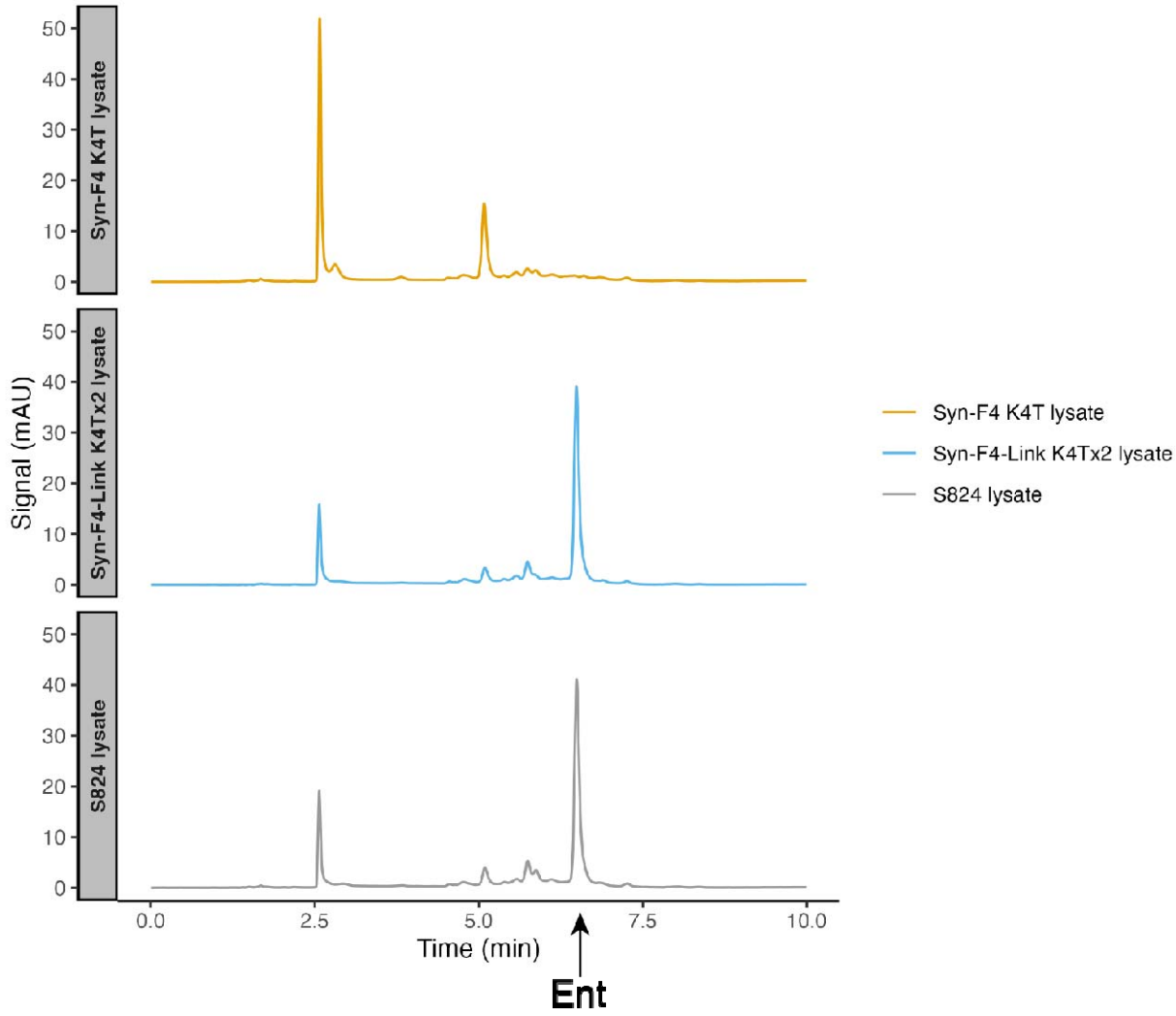
SI Figure 4. MD simulation of Syn-F4 in loopleft and loopleft conformations. Figure adopted from Kurihara et al.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



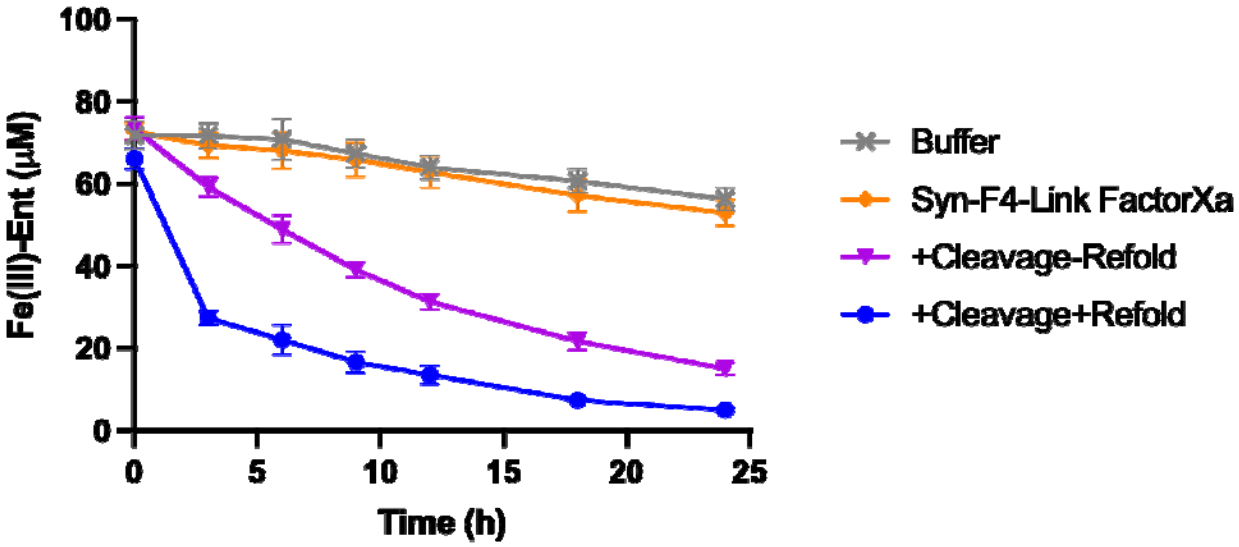
SI Figure 5. Docking simulation of Fe(III)-Ent onto Syn-F4-Link with loopright conformation. Left: distribution of Fe^{3+} in docking simulations (orange spheres). Right: the docking pose with the lowest docking score. Fe(III)-Ent molecule is shown in ribbon model. H74 and H186 are shown in blue and grey spheres.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



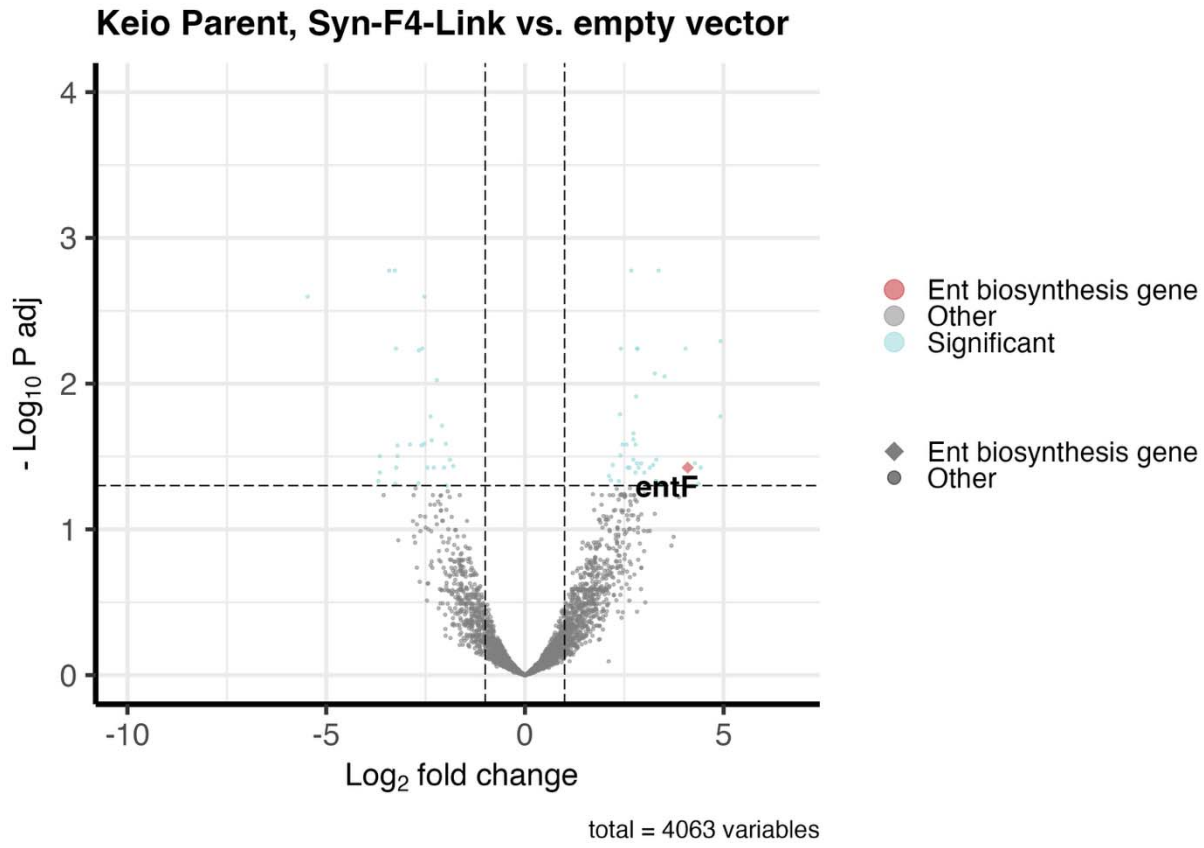
SI Figure 6. HPLC chromatogram of enterobactin and its hydrolysis products by the soluble lysate of Δ*fes* cells expressing Syn-F4, Syn-F4-Link and S824 (negative control). Peak corresponding to enterobactin (Ent) is labeled.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



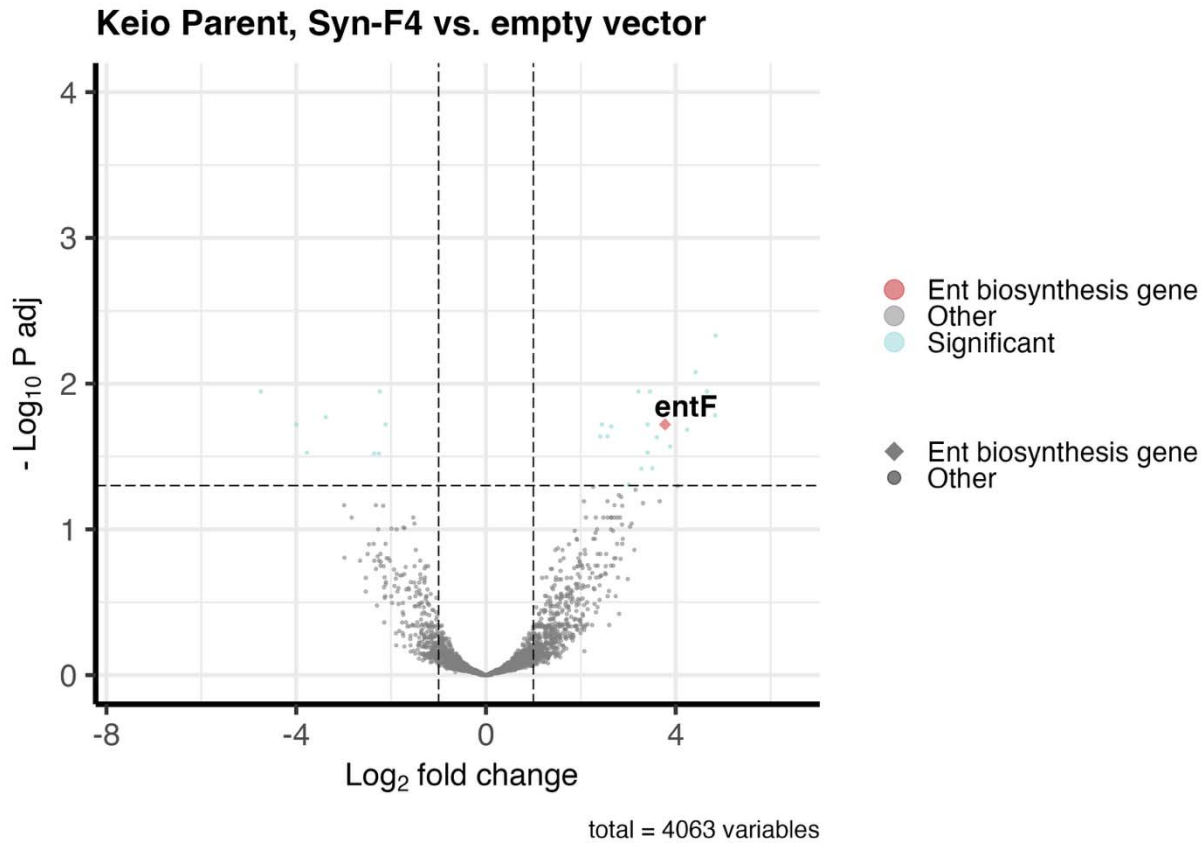
SI Figure 7. Hydrolysis of Fe(III)-Ent over 24 hours by modified Syn-F4-Link without cleavage (Syn-F4-Link FactorXa), with cleavage only (+Cleavage-Refold), with cleavage and refolding (+Cleavage+Refold) or buffer.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



SI Figure 8. RNA-seq of Syn-F4-Link vs. empty vector in Keio parent strain. The expression of *entF* gene is significantly upregulated.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



SI Figure 9. RNA-seq of Syn-F4 vs. empty vector in Keio parent strain. The expression of *entF* gene is significantly upregulated.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SI Table 1. X-ray data collection and refinement statistics.

Crystal	Syn-F4-Link (K4Tx2)
Data collection	
Space group	*C*2
Unit-cell parameters	*a* = 61.44 Å, *b* = 47.39 Å, *c* = 70.13 Å α = 90.00°, β = 103.63°, γ = 90.00°
X-ray source	KEK Photon Factory (PF) BL-17A
Wavelength (Å)	0.980
Resolution (Å)	37.12–2.30 (2.38–2.30)
Unique reflections	8823 (866)
Average redundancy	6.4 (6.3)
Completeness (%)	99.6 (99.6)
$I / \sigma I$	6.9 (2.0)
R_{meas}	0.228 (1.246)
$CC_{1/2}$	0.982 (0.646)
Refinement	
Resolution (Å)	37.12–2.30
No. of reflections	8367
No. of protein atoms	1566
No. of water molecules	29
R_{work}	0.274
R_{free} [1]	0.326
RMSD bond length (Å)	0.012
RMSD bond angles (°)	1.499
Average B-factor (Å^2)	42.76
Ramachandran plots	
in favored region (%)	98.3
in allowed region (%)	1.7
in outlier region (%)	0.0
PDB accession code	9UGR

All numbers in parentheses represent the last outer shell statistics.
[1] R_{free} is calculated for 5% of randomly selected reflections excluded from refinement.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.09.693225; this version posted December 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SI Table 2. Summary of small-angle X-ray scattering (SAXS) results.

Sample	$I(q\rightarrow 0)/c$ [a.u.]	D_{max} [nm]	R_g [nm]	M_w [kDa]	Calc. M [kDa] (Monomer)	Oligomeric State [mer]
Syn-F4_Link	0.01350	10.5	2.50	23.8	25.5	1
Syn-F4(K4T)	0.01314	10.6	2.58	23.2	12.5	2
Lysozyme	0.00812	6.0	1.51	14.3	14.3	1