Tight and specific lanthanide binding in a de novo TIM barrel with a large internal cavity designed by symmetric domain fusion

Shane J. Caldwell[a,b,1], Ian C. Haydon[a,b,1], Nikoletta Piperidou[c], Po-Ssu Huang[a,b,d], Matthew J. Bick[a,b], H. Sebastian Sjöström[c], Donald Hilvert[c], David Baker[a,b,e,2], and Cathleen Zeymer[c,f,2]

[a]Department of Biochemistry, University of Washington, Seattle, WA 98195; [b]Institute for Protein Design, University of Washington, Seattle, WA 98195; [c]Laboratory of Organic Chemistry, Eidgenössische Technische Hochschule (ETH) Zürich, 8093 Zürich, Switzerland; [d]Department of Bioengineering, Stanford University, Shriram Center for Bioengineering and Chemical Engineering, Stanford, CA 94305; [e]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195; and [f]Department of Chemistry, Technische Universität München, 85747 Garching, Germany

Edited by James A. Wells, University of California, San Francisco, CA, and approved October 13, 2020 (received for review May 5, 2020)

De novo protein design has succeeded in generating a large variety of globular proteins, but the construction of protein scaffolds with cavities that could accommodate large signaling molecules, cofactors, and substrates remains an outstanding challenge. The long, often flexible loops that form such cavities in many natural proteins are difficult to precisely program and thus challenging for computational protein design. Here we describe an alternative approach to this problem. We fused two stable proteins with C2 symmetry—a de novo designed dimeric ferredoxin fold and a de novo designed TIM barrel—such that their symmetry axes are aligned to create scaffolds with large cavities that can serve as binding pockets or enzymatic reaction chambers. The crystal structures of two such designs confirm the presence of a 420 cubic Ångström chamber defined by the top of the designed TIM barrel and the bottom of the ferredoxin dimer. We functionalized the scaffold by installing a metal-binding site consisting of four glutamate residues close to the symmetry axis. The protein binds lanthanide ions with very high affinity as demonstrated by tryptophan-enhanced terbium luminescence. This approach can be extended to other metals and cofactors, making this scaffold a modular platform for the design of binding proteins and biocatalysts.

lanthanides | de novo protein | protein engineering | metalloprotein | protein design

The TIM barrel fold is among the most versatile enzyme scaffolds in nature (1, 2). It consists of eight parallel β-strands surrounded by eight α-helices in a $(\beta/\alpha)_8$ topology. The parallel β-strands confer a polarity to the domain—a "top" and "bottom." Enzyme active sites in nature are usually located at the center of the top (3), where side chains from all eight strands are directed toward the center of the barrel. Additional substrate-binding components are typically formed from irregular, extended loops that connect the strands and helices (4). Below the active site, an extensive hydrophobic core between the β-barrel and the α-helices provides considerable stability, allowing TIM barrel active sites to diversify with minimal effects on overall fold stability (5).

The engineering of natural TIM barrels, both by directed evolution and design, has resulted in enzymes with altered substrate specificities, enhanced reaction rates, and even new-to-nature chemistries (6–13). The de novo design of catalytic functions, however, has been challenged in part by the difficulty in programming flexible loops; the most successful approaches to date have reused loops found in nature (14, 15). An artificial TIM barrel designed with atomic-level accuracy was created previously using the Rosetta macromolecular and design platform (16, 17). This protein (sTIM11), which has no native function and no natural sequence homologs, was created by repeating four copies of a 46-amino acid $(\beta/\alpha)_2$ unit about a central symmetry axis. It was expressed as a single chain, but the internal symmetry left open the possibility of converting the barrel into a homotetramer or homodimer by expressing one or two 46-residue repeats, then allowing these parts to self-assemble in solution to form a complete TIM barrel. More recently, this scaffold has been reengineered to improve thermodynamic and kinetic stability, yielding a series of improved variants dubbed DeNovoTIMs (18). De novo protein design is inspired by but not limited to nature's solutions. To generate a protein scaffold that might support the incorporation of catalytic cofactors, we aimed for an alternative to loop design: fusing ordered protein domains directly on top of a DeNovoTIM variant.

Many natural proteins require a metal ion to carry out their function. Such metalloproteins are involved in numerous biological processes, including metal ion storage and transport, sensing and signaling, electron transfer pathways, and catalysis. Computationally designed proteins may serve as a starting point to engineer artificial metalloproteins with specific binding sites for a metal ion of choice. Thus, in addition to engineering a suitable lid atop DeNovoTIM, we set out to introduce a de novo metal-binding site for large, trivalent lanthanide cations, which possess unique electronic and photophysical properties (19). The f-block elements exhibit long-lived luminescence, which can be

Significance

Despite considerable advances in de novo protein design in recent years, it still remains challenging to engineer proteins with large internal cavities that can be functionalized to become biotechnological tools, such as specific binders, sensors, or catalysts. In this work, we outline a computational strategy to combine multiple de novo designed domains into symmetric protein assemblies that enclose large internal chambers. The high stability of de novo scaffolds enables ready functionalization of these chambers; for instance, with specific metal-binding sites, as demonstrated here by generating a lanthanide-binding protein with ultra-high affinity.

Author contributions: S.J.C., I.C.H., P.-S.H., D.H., D.B., and C.Z. designed research; S.J.C., I.C.H., N.P., H.S.S., and C.Z. performed research; P.-S.H. and C.Z. contributed new reagents/analytic tools; S.J.C., N.P., P.-S.H., M.J.B., D.H., D.B., and C.Z. analyzed data; and S.J.C., I.C.H., D.B., and C.Z. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]S.J.C. and I.C.H. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: dabaker@uw.edu or cathleen.zeymer@tum.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2008535117/-/DCSupplemental.

enhanced through energy transfer from a close-by antenna molecule, an approach that is exploited in so-called lanthanide-binding tags (20, 21). De novo designed peptides were shown previously to form lanthanide-binding coiled coils with potential applications as paramagnetic MRI contrast agents (22, 23). Their strong Lewis acidity also makes lanthanides powerful catalysts for various synthetically useful reactions (24), although their biological relevance was only recently discovered, when lanthanide-dependent methanol dehydrogenases were isolated from certain methylotrophic and methanotrophic bacteria (25, 26). Here, we introduce a de novo lanthanide-binding protein with ultra-high affinity, which provides a starting point for the future engineering of lanthanide-dependent biotechnological tools and artificial enzymes.

Results

DeNovoTIM15. We combined stabilizing sequence changes (18) with circular permutation and substitution of residues at the top of the TIM barrel to produce stable and modular starting TIM barrel scaffolds (Fig. 1*A*). These proteins, dubbed DeNovoTIMs, were expressed at high levels in *Escherichia coli*. They are stable beyond 95 °C as assessed by circular dichroism (CD) and have a guanidinium chloride unfolding midpoint of ~2M (Fig. 1*B*). One of the variants, DeNovoTIM15, was crystallized and the structure solved using the Rosetta model as input for molecular replacement at a resolution of 2.3 Å (Table 1). The design model and crystal structure were very similar, with an all-atom Cα root-mean-square deviation (rmsd) of 1.65 Å (Fig. 1*C*). Exploiting the symmetry of the de novo TIM system, we were able to express a half fragment of DeNovoTIM15, containing only the first two repeat units of the protein. This protein was expressed in *E. coli* and purified. It was shown to be a homodimer by size exclusion chromatography (SEC) (*SI Appendix*, Fig. S1), consistent with a C2-symmetric homodimeric assembly. Relocating the termini from the top to the bottom of the barrel by circular permutation provided positions at the top of the TIM barrel where loops could be substituted with inserted domains. We inserted domains in one of the loops that is structurally equivalent to the prepermutation termini: the β-α loop connecting β3 and α3.

Symmetric Fusion Design. To create large, engineerable reaction chambers at the top of the TIM barrel, we used a computational approach that takes advantage of the C2 symmetry of many de novo designed homodimers. C2-symmetric de novo designed homodimer proteins were fused to the top of the half-DeNovoTIM15 using a method that preserves symmetry of the resulting assembly. Two different symmetric homodimers were considered for the top of the reaction chamber: a de novo designed ferredoxin (FD) fold (27) and cFr (28), derived from Top7 (29), a de novo designed protein with a fold not previously seen in nature. Both of these dimers have continuous β-sheets extending across the dimer interface. The domain insertion function of RosettaRemodel (30) was used to connect each "lid" monomer to a monomer of the

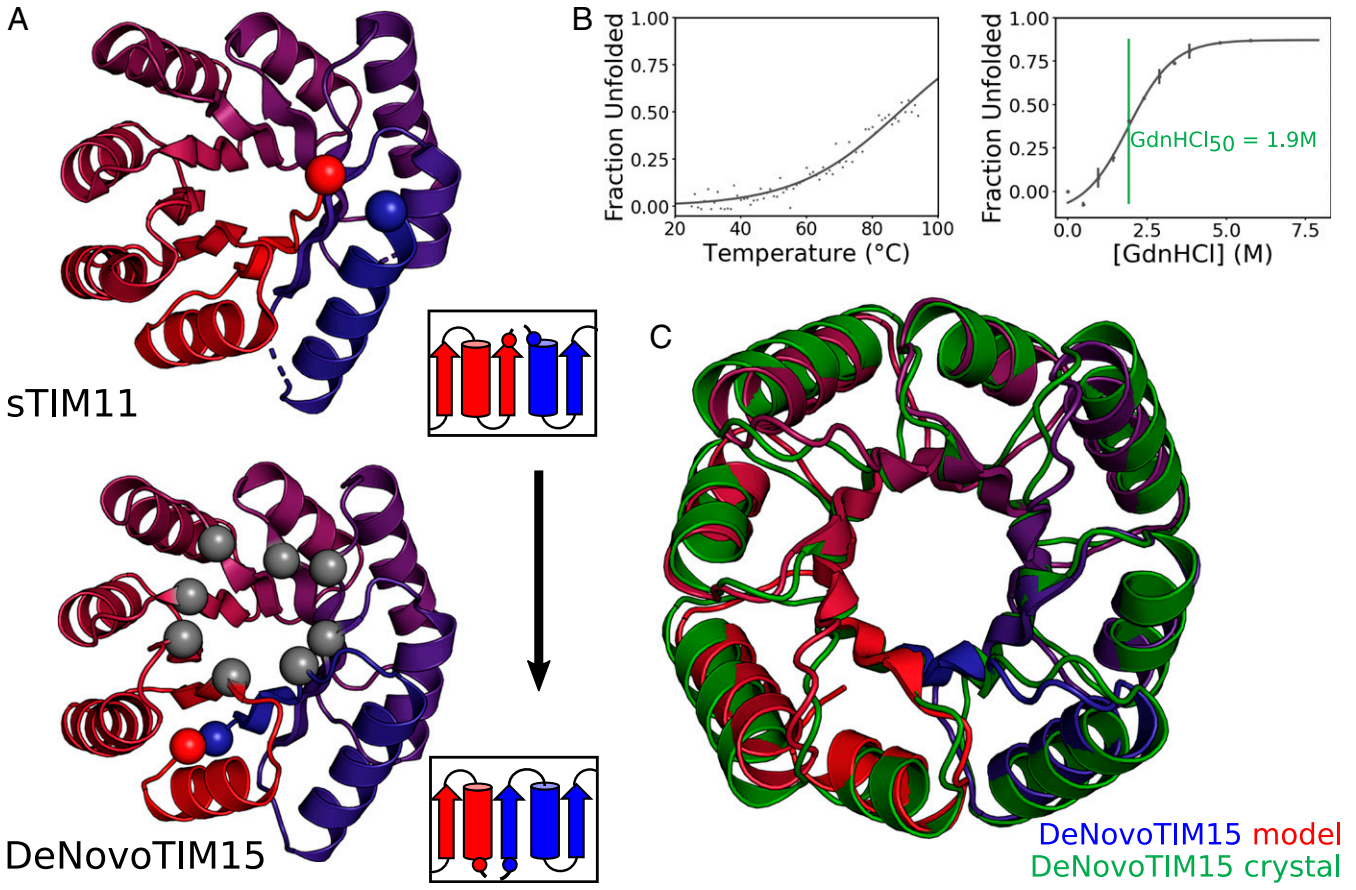

Fig. 1. Design and characterization of a hyperstable de novo TIM barrel—DeNovoTIM15. (*A*) DeNovoTIM15 represents a circular permutation of sTIM11, with termini relocated to the bottom of the TIM barrel. Blue and red spheres represent N- and C-termini, respectively. In addition to stabilizing changes engineered in the DeNovoTIM series, select surface sites were also changed on the top of DeNovoTIM15, shown in light gray. (*B*) Thermal and chemical denaturation profiles of a re-engineered TIM barrel (sTIM-EEEE) using circular dichroism. (*C*) Superimposition of DeNovoTIM15 crystal structure (green) with design model (blue-red).

Table 1. Data collection and refinement statistics

	DeNovoTIM15	TFD-HE	TFD-EE	TFD-EE N6W + Tb(III)[†]
PDB ID	6WVS	6WXO	6WXP	6ZV9
Wavelength	1.00	1.00	1.00	1.00
Resolution range	37.44 to 2.20 (2.28 to 2.20)	32.95 to 1.41 (1.46 to 1.41)	46.44 to 2.5 (2.59 to 2.5)	30.00 to 1.85 (1.95 to 1.85)
Space group	$P\ 3_2$	$P\ 2_1$	$P\ 2_1$	$P\ 2_1$
Unit cell	53.20 53.20 64.27	48.85 62.51 48.95	43.91 101.16 117.73	43.97 56.47 71.28
	90 90 120	90 95.27 90	90 95.50 90	90 101.69 90
Total reflections	93,205 (7,829)	202,977 (19,737)	109,529 (10,001)	172,450 (24,339)
Unique reflections	10,220 (946)	55,043 (5,447)	34,814 (3,455)	56,427 (8,189)
Multiplicity	9.1 (8.2)	3.7 (3.6)	3.1 (2.9)	3.1 (3.0)
Completeness (%)	99.54 (97.44)	97.03 (95.08)	97.88 (96.05)	98.4 (97.9)
Mean I/sigma(I)	8.32 (2.31)	7.39 (0.77)	14.17 (2.10)	13.12 (1.94)
Wilson B-factor	50.02	16.91	56.35	31.37
R-merge	0.204 (2.76)	0.104 (1.79)	0.0524 (0.505)	0.034 (0.339)
$CC_{1/2}$	0.988 (0.558)	0.995 (0.220)	0.999 (0.776)	0.998 (0.836)
CC*	0.997 (0.846)	0.999 (0.601)	1 (0.935)	1 (0.954)
Reflections used in refinement	10,220	54,890	34,799	29,150
Reflections used for R-free	982	1,179	1,736	1,456
R-work	0.200	0.182	0.227	0.195
R-free	0.216	0.212	0.281	0.229
Number of nonhydrogen atoms	1,310	2,847	7,319	2,649
Macromolecules	1,291	2,658	7,208	2,600
Ligands		21	51	5
Solvent	19	168	60	44
Protein residues	182	352	932	335
RMSD (bonds)	0.004	0.012	0.002	0.003
RMSD (angles)	0.94	1.10	0.47	1.18
Ramachandran favored (%)	98.33	99.43	97.07	99.70
Ramachandran allowed (%)	1.67	0.57	2.93	0.30
Ramachandran outliers (%)	0.00	0.00	0.00	0.00
Rotamer outliers (%)	2.73	0.37	1.70	3.65
Clashscore	3.91	4.98	6.17	1.87
Average B-factor	55.91	23.29	85.40	50.40
Macromolecules	55.97	22.26	85.79	50.57
Ligands		56.27	63.68	30.28
Solvent	51.81	35.41	57.92	42.83

Statistics for the highest-resolution shell are shown in parentheses.
[†]Friedel pairs were treated as different reflections.

homodimeric TIM barrel. Linker loops between zero and six residues in length were generated to connect the domains (Fig. 2*A*), and C2 symmetry as well as distance constraints were imposed during design to arrive at fusion proteins in which both the TIM barrel dimer interface and lid dimer interface were intact. We discarded all models for which complete chain closure was unsuccessful and used Rosetta to assign the sequences of the linker loops. Limited sequence changes were made at the interface between domains to avoid generating structures with stability dependent on these interactions. Changes to the insert (four per monomer) and TIM (two per monomer) domains were solely made to reduce the charge lining the chamber between domains (*SI Appendix*, Fig. S2). Following successful construction of the fused homodimer, we also considered additional features that could provide function to the otherwise inert scaffold. Visual inspection of the top of the TIM barrel suggested that the introduction of C4-symmetric polar residues could generate a C4-symmetric metal ion-binding site. To explore this, we created fusion protein variants in which glutamate or histidine residues replaced select valines at the top of one of the β-strands, at positions 31 and 154 of the full-length construct. In each case, there were four altered residues, corresponding to the C4 symmetry of the TIM barrel.

Characterization of TFD Designs. The most successful fusion contained a C2-symmetric FD atop DeNovoTIM15, with three-residue glycine linkers closing both junctions. This protein, dubbed TIM-FD dimer (TFD) is remarkably stable, soluble, and robust. Two variants of this scaffold were characterized. The first, TFD-HE, contains a histidine in position 31 and a glutamate in position 154, while the second, TFD-EE, contains two glutamate substitutions in these positions at the top of the beta-barrel, for a total of four potentially metal-coordinating residues in the two-fold symmetric assembly. Both TFD-HE and TFD-EE are remarkably stable protein scaffolds (Fig. 3*A*). They were subjected to the same CD experiments as the lid-less TIM barrel variant detailed in Fig. 1*B*, and we observed an even greater resilience to thermal denaturation, as 80% of the secondary structure remained intact at 95 °C. Furthermore, the scaffolds showed a two-step melting behavior in response to chemical denaturation with guanidinium chloride (GdnHCl), reflecting the unfolding of the TIM domain (2.5, 2.1 M GdnHCl) and the FD insert domain (7.0, 6.9 M GdnHCl), respectively. The TFD scaffolds also remain fully folded in substantial amounts of cosolvent, such as 50% acetonitrile (*SI Appendix*, Fig. S3).

The proteins were first crystallized in the absence of metal. The structure of TFD-HE was determined at 1.4 Å. This protein construct contained an N-terminal affinity tag that remained in the protein that was crystallized. The tag was well resolved in the electron density, partially packing beside the C-terminal helix, which explained our difficulty in cleaving it with tobacco etch virus (TEV) protease. This structure validated the connectivity and topology of the TIM barrel and FD domains. However,

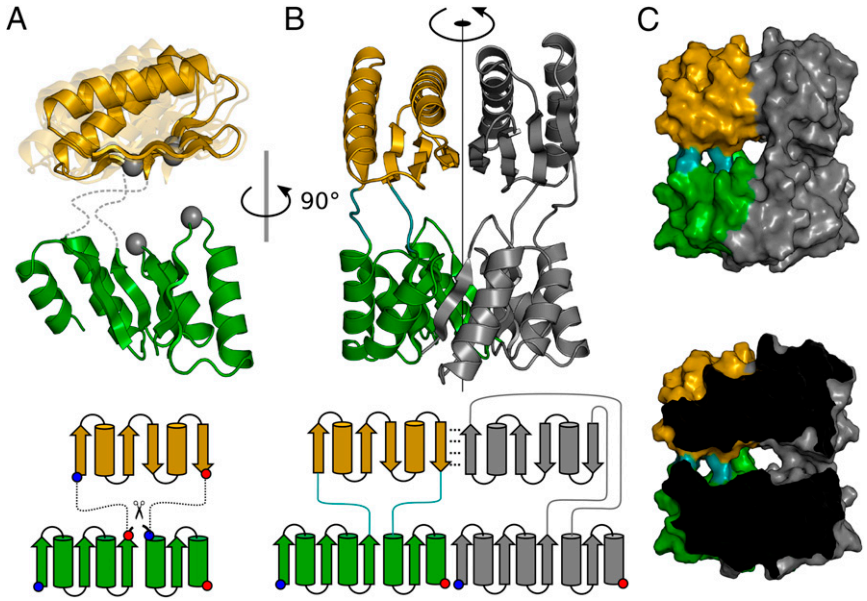

Fig. 2. Construction of the TFD scaffolds. (*A*) The insert domain (yellow) was inserted into a loop of the TIM barrel (green) using RosettaRemodel. Select sites were mutated to reduce the surface charge of the inserted domains (gray spheres). Using constraints between the inserted domain and its symmetric partner, the symmetry of the assembly was maintained, yielding a C2-symmetric homodimer following successful construction. (*B*) In the dimeric assembly, the two C2-symmetric domains have aligned symmetry axes, with contiguous beta-sheets. Linking peptides are shown in teal. (*C*) The successfully built homodimeric Rosetta model.

there was a rotation of the two domains relative to each other, deviating by 10° around the symmetry axis when compared to the design model (Fig. 3 *B–E*). The TIM domain and FD domain enclose an internal cavity of 420 Å^3. The inserted H31 and E154 residues of this protein did not adopt a conformation that would support metal binding as the histidines in the H-E-H-E symmetric site are constrained to hydrogen-bond networks.

The structure of the second variant, TFD-EE, was determined at 2.5 Å resolution and contains three dimers in the asymmetric unit. The rotation of the two domains relative to each other was more pronounced, deviating by 17 to 34°. Despite this, the TIM and FD dimers align to their parental domains with Cα rmsds of 3.16 Å and 1.78 Å, respectively, with deviations in the TIM domain predominantly explained by shifts of the C-terminal helix, remote from the domain-domain interface. In the central symmetric metal-binding site, we observed electron density consistent with a bound metal ion, even though no specific metal was added to the protein sample (*SI Appendix*, Fig. S4). We believe this metal copurified with the protein from the intracellular environment or lysate. The coordination sphere resembles the situation in natural calcium-binding proteins, which led us to hypothesize that the bound metal is likely a Ca^{2+} ion. Microwave inductively coupled atmospheric pressure plasma mass spectrometry (MICAP-MS) experiments supported this hypothesis (*SI Appendix*, Tables S1 and S2), so a Ca^{2+} ion was modeled into the crystal structure at the center of each dimer.

Lanthanide Binding. We hypothesized that the four glutamate residues at the center of TFD-EE would provide a suitable coordination environment for large trivalent cations, such as lanthanides, as these are typically oxophilic and can interact directly with eight or more ligand atoms. Inspired by the design of lanthanide-binding tags (21), we installed an intrinsic antenna residue adjacent to the metal-binding site (N6W mutation); this tryptophan serves as a site-specific sensitizer for lanthanide luminescence (Fig. 4 *A–C*). When the protein was mixed with $TbCl_3$, strong tryptophan-enhanced terbium(III) luminescence was observed upon excitation at 280 nm, enabling quantitative measurement of lanthanide binding through a specific spectroscopic readout. Furthermore, we determined the crystal structure of terbium(III)-bound TFD-EE N6W at 1.85 Å resolution and observed strong anomalous electron density at the metal-binding site originating from specific binding of the heavy ion (Fig. 4*B*). The C2 symmetry of the scaffold is lost in this structure due to a translation of the FD dimer with respect to the TIM barrel, further indicating the flexibility of the linkers that connect both domains.

We titrated metal-free protein with preincubated mixtures of $TbCl_3$ and ethylenediaminetetraacetic acid (EDTA) and monitored the enhanced terbium signal over time (Fig. 4*D*). The protein slowly removes the metal from the EDTA complex, which has a binding constant of 1.6×10^{-18} M (31), suggesting ultra-high affinity. A binding curve based on the estimated free metal concentration in these mixtures indicates a binding affinity of TFD-EE for Tb(III) in the subfemtomolar range. Notably, when the Tb(III)/protein complex is formed first, even 100-fold excess EDTA does not quantitatively remove the metal from the scaffold after 1 wk of incubation. Displacement experiments with other lanthanides showed that Tb(III) bound to TFD-EE N6W can be exchanged on the timescale of hours. Eu(III) and Yb(III) bind with comparable affinity, while Ce(III) binds ∼10-fold weaker (*SI Appendix*, Fig. S5).

Discussion

By fusing two de novo designed homodimeric proteins such that their C2 symmetry axes are preserved, we have created a homodimeric protein with a TIM barrel base and a dimeric FD lid surrounding an enclosed chamber. Fusion of de novo designed building blocks has been used previously to create proteins with extended beta sheets (32). We show here that extended enclosed cavities can be created by designed juxtoposition of de novo C2 building blocks. Some examples of similar fusions creating cavities exist in nature, such as AAA+-type molecular machines (33) or ATP-binding cassette (ABC) membrane transporters (34). However, to our knowledge, TFD is the most compact protein to enclose a central cavity in this manner.

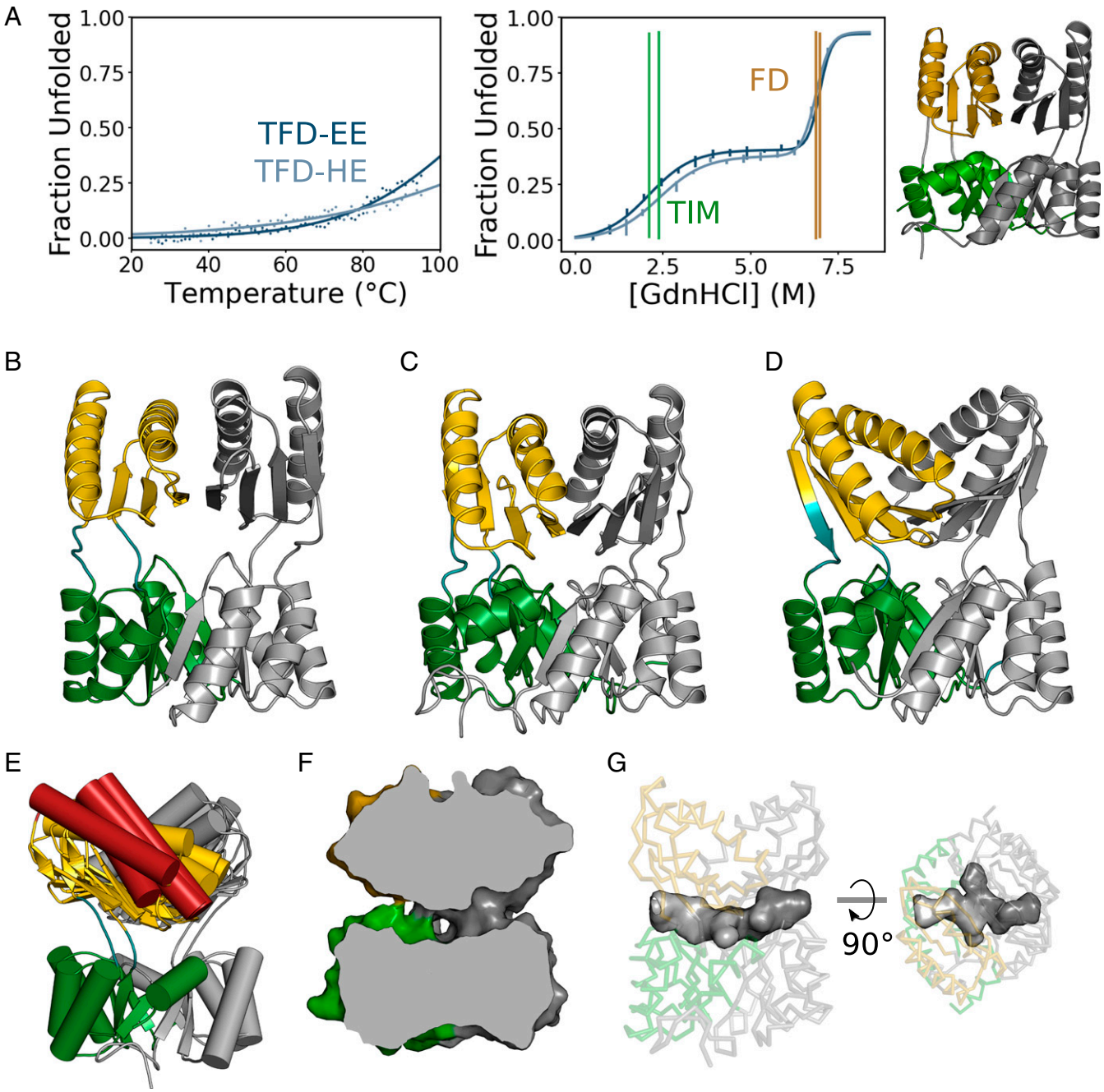

Fig. 3. Stability and structure of TFD proteins. (*A*) Thermal and chemical denaturation profiles measured by circular dichroism for TFD-EE and TFD-HE. Both proteins are stable beyond 95 °C, and possess two inflection points upon chemical denaturation, reflecting the TIM and FD denaturation points at around 2 M and 7 M guanidinium chloride. A slight change in the TIM melting is observed due to the Glu-His substitution of TFD-HE relative to TFD-EE. (*B–E*) The alignment of TFD design model and crystal structures indicate the variability in interdomain orientation between the FD and TIM domain. (*B*) TFD design model. (*C*) TFD-HE structure, aligned to the TIM barrel of the design model in *B*. The upper FD domain carries a 10° twist relative to the TIM. (*D*) TFD-EE structure, aligned to the same TFD design model. This structure carries a 34° twist. (*E*) Superimposition of all three dimers of the asymmetric unit of the TFD-EE structure, and the dimer from the TFD-HE structure illustrates the different interdomain conformations accessed TFD. Helix 2 of the first FD colored red for emphasis. (*F*) TFD encloses a large internal cavity that is solvent accessible from both sides but enclosed between the top and bottom domains. TFD-HE surface displayed at 1.9 Å solvent radius, cross-section that includes molecular symmetry axis. (*G*) In the TFD-HE structure, this chamber encloses about 420 Å^3. Rendered at 1.9 Å solvent radius cavity mode in PyMol.

Our TIM-FD design has attractive features as a scaffold for future de novo enzyme design. First, a large chamber is formed at the interface between two structured domains (Fig. 3 *F* and *G*). In TFD, there are 9 × 2 residues pointing into this chamber from the TIM barrel, and 9 × 2 from the FD, for a total of 36 functionalizable residues. Some of these residues could be substituted with hydrophobic amino acids to increase the hydrophobic character of the chamber. Second, the high stability of TFD suggests it should tolerate considerable modification of the residues surrounding the chamber. TFD and the derivative

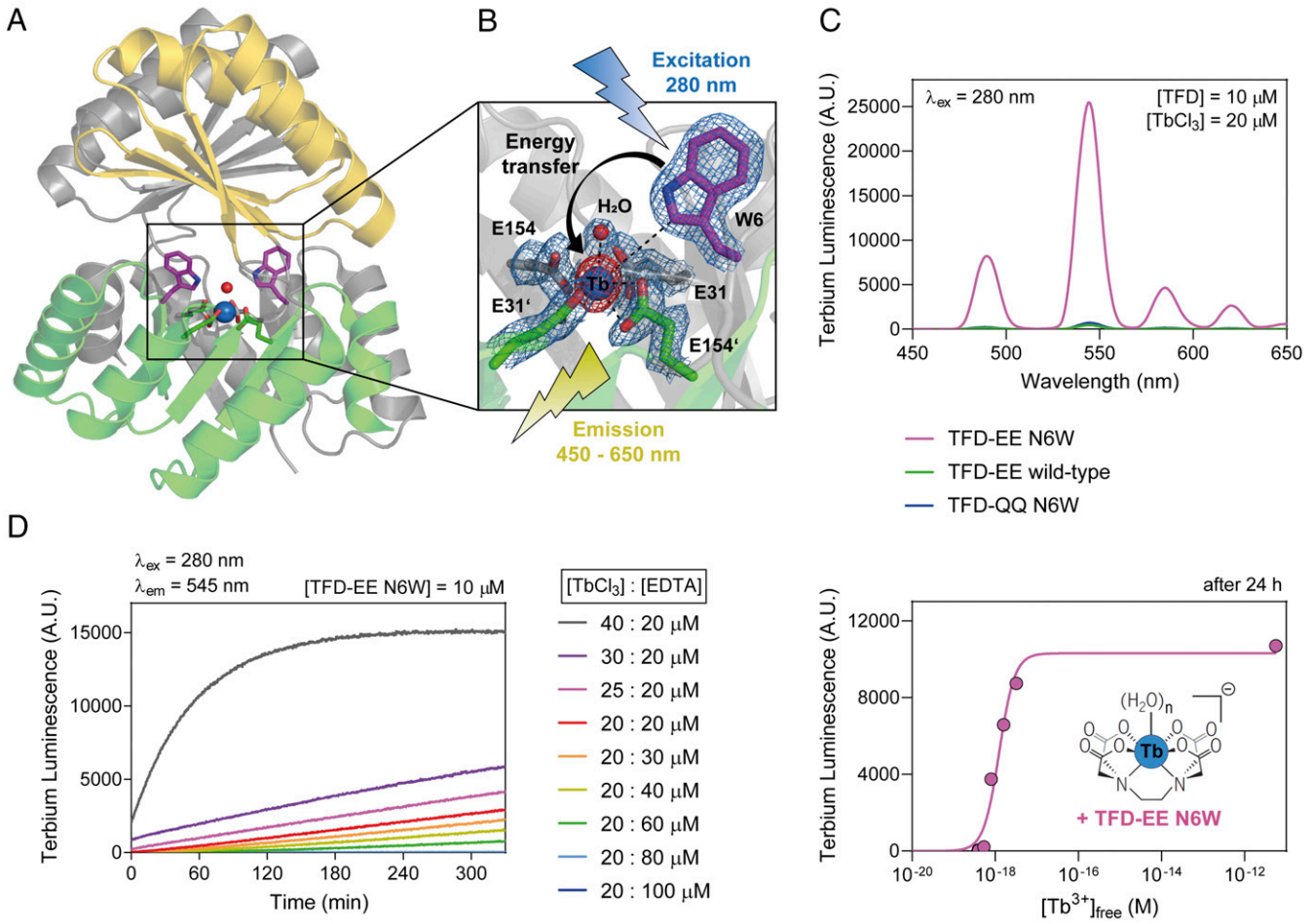

Fig. 4. Structure and lanthanide-binding properties of TFD-EE. (*A*) Crystal structure of the Tb(III)-bound TFD-EE N6W dimer at 1.85 Å resolution. The protein was equipped with an intrinsic tryptophan antenna (N6W, pink) to enhance Tb(III) luminescence through energy transfer. The initial C2 symmetry of the dimeric scaffold is lost, as the FD lid domain is both rotated and translated compared to the structures shown above. Each monomer exhibits a distinct conformation of the flexible linkers that connect both domains. (*B*) Close-up view of the metal-binding site, which consists of 2 × 2 glutamate residues (green and gray). $2F_o$-F_c electron density is shown as a blue mesh contoured at 1.0 σ. Anomalous electron density is shown as a red mesh at 15.0 σ, indicating specific binding of the lanthanide ion. The antenna tryptophan (pink) is in close proximity (<5.0 Å), allowing for efficient energy transfer. The distance between Tb(III) and the coordinating carboxylate oxygen atoms is 2.3 to 2.5 Å and the distance between Tb(III) and the axial water molecule is 2.7 Å. (*C*) Tryptophan-enhanced Tb(III) luminescence upon excitation at 280 nm. Control proteins with either no tryptophan antenna (TFD-EE wild-type) or knocked-out binding site (Glu to Gln mutations, TFD-QQ N6W) show only background signal. (*D*) Lanthanide-binding measurements using the specific luminescence readout. Direct titrations were not possible due to the very high affinity and slow binding kinetics. However, titrating TFD-EE N6W with preincubated mixtures of EDTA and $TbCl_3$ allows for the estimation of the Tb(III)-binding affinity of the protein scaffold in the subfemtomolar range, based on $K_D(Tb^{3+}/EDTA) = 1.6 \times 10^{-18}$ M as the reference.

designs presented here are homodimeric proteins, but the modular design makes it readily convertible to an asymmetric monomeric protein through the simple closure of the protein chain between two repeats (*SI Appendix*, Fig. S6). Control of the interdomain configuration of the TIM domain and FD remains a challenge. While the homodimeric domains were both internally unaffected by fusion into a two-domain homodimeric assembly, their relative orientations were not exactly as predicted. Controlling this relative position will be important for construction of binding sites that use elements from both domains. Future design will address the relative orientations through engineering of the contacting surfaces and linking loops. It should also be noted that this mobility, if controlled, could prove beneficial in the longer term as dynamics are often important to enzyme function.

In this work, the TIM barrel component of the highly stable TFD scaffold was equipped with a specific binding site for lanthanide ions. As lanthanides find applications in various technologies, including magnets, lasers, luminescent materials, and optoelectronics, there is an increasing demand in industry (35). However, conventional mining of these elements is problematic, and other sources of low lanthanide concentration must be considered. A biomacromolecular extraction concept has recently been suggested using lanmodulin, a natural lanthanide-binding protein (36). Our de novo lanthanide-binding scaffold, TFD-EE, exhibits even higher lanthanide affinity and tolerates heat and cosolvents to a remarkable extent. It may thus be utilized in biotechnological approaches to lanthanide recycling as well as lanthanide-dependent catalysis. While the FD lid domain of TFD-EE is not involved in metal binding, it will be crucial for further engineering; for example, providing residues for specific substrate recognition inside the cavity. We are currently pursuing the design of enzyme active sites in the lanthanide-bound TFD scaffolds—for both hydrolytic and carbon-carbon bond forming reactions that can capitalize on the strong Lewis acidity of lanthanide ions. Our approach can be extended to various other metals by changing the type and geometry of the coordinating

residues. Large cofactors may also be inserted through specific interactions with the TIM domain or FD domain. We believe the internal chamber-containing scaffolds will provide a modular platform for de novo enzyme engineering, contributing to the longstanding goal of designing novel enzymes from scratch.

Methods

Protein Engineering and Fusion Protein Design. DeNovoTIM15 was developed combining stabilizing changes introduced to de novo TIM barrel sTIM11 by Romero-Romero, et al. (18), with select mutations at the top of the TIM barrel and a circular permutation (Fig. 1*A*). DeNovoTIM15 corresponds to the sequence of DeNovoTIM13 with a change of eight residues replacing charged residues atop the TIM barrel. A C2-symmetric half-TIM construct was created by truncating the four-repeat protein to two repeats. This construct was cloned and expressed and found to self-assemble into a homodimer as measured by size-exclusion chromatography (*SI Appendix*, Fig. S1). RosettaRemodel was used to design the connections (or "linkers") between half-DeNovoTIM15 and each inserted domain (Fig. 2). We used the design feature that facilitates the introduction of a domain as a rigid body insertion while allowing the connecting loops to be designed. The inserted domain has both its termini connected within a loop of a hosting structure; RosettaRemodel uses random fragments to sample the linkers and loop closure criteria to screen plausible fusions of the inserted domain and the host structure. By sampling the linker conformations, the relative position of the inserted domain with respect to the host structure is also sampled (Fig. 2*A*). Linker lengths between zero and six amino acids were tested, though only a subset produced fully connected fusion proteins. For this project, we enabled symmetry with the domain insertion feature in RosettaRemodel, with the C2 symmetry of the host structure (TIM barrel) held constant while the rigid-body conformations of the FD or cFr domains move synchronously across the interface during sampling. To ensure that the dimer interfaces of the inserted domains are maintained, restraints corresponding to the FD and cFr dimers were used, defined as distance constraints harvested from atom-pairs from the dimer structures. The RosettaRemodel flags and the calculation setups as blueprint files are given in the *SI Appendix*.

Gene Synthesis, Protein Expression, and Purification. Genes were synthesized by commercial vendors and cloned into vectors for protein expression in *E. coli*. Protein was expressed in shaking liquid cultures using either isopropyl thiogalactoside or autoinduction methods. Cells were lysed using high-pressure homogenization or sonication and clarified by centrifugation. This lysate was purified by gravity-flow nickel-affinity chromatography and further purified using either size-exclusion chromatography or TEV protease digestion, followed by a secondary nickel separation step. As these proteins have shown themselves to be extremely robust, we do not believe alterations in the expression and purification protocols to be substantive to the results. For more detailed information for specific protein constructs see *SI Appendix, Methods*.

Thermal and Chemical Protein Denaturation Measured by CD Spectroscopy. CD experiments were performed in quartz cuvettes with 2 mm path length using a JASCO J-710 CD instrument. The proteins were diluted in pure water or 0.5 to 8.0 M guanidinium chloride solution to a final concentration of 20 μM TFD monomer, with a 30 min incubation time at room temperature prior to measurement. Thermal denaturation curves were measured in the absence of denaturing agent at 222 nm from 25 to 95 °C with a scan speed of 30 K/h. Chemical denaturation curves were generated by measuring CD spectra (260 to 200 nm, averaging five scans) at various concentrations of guanidinium chloride at 25 °C. The signal at 222 nm was converted into mean residue molar ellipticity, and then to fraction of total protein unfolded. All points were measured in triplicates. The curves were fit with a sigmoidal equation to determine the transition points.

Protein stability in cosolvents was measured by recording CD spectra (260 to 200 nm, five averages) of TFD-EE (c = 20 μM monomer) after incubation in 0 to 90% methanol or acetonitrile for 20 min at 25 °C.

Crystallization and Structure Determination. All proteins were crystallized by vapor diffusion. The reengineered DeNovoTIM15 was crystallized in Morpheus screen solution D9 (37): 10% wt/vol polyethylene glycol (PEG) 20000, 20% vol/vol PEG monomethyl ether 550 0.02 M each of each 1,6-hexanediol, 1-butanol, (*RS*)-1,2-propanediol, 2-propanol, 1,4-butanediol, 1,3-propanediol, 0.1 M bicine/Trizma base pH 8.5. TFD-HE was crystallized in 0.1 M citric acid pH 4.0 and 2.4 M ammonium sulfate. Small rod-shaped crystals were found after 6 mo of incubation. Crystals of TFD-EE were grown in a sitting drop vapor diffusion setup in 96-well plates at 20 °C. The protein solution (10 mg/mL in 25 mM Hepes pH 7.5) was mixed 1:1 with the reservoir solution, which contained 0.1 M Hepes pH 7.5, 10 to 15% PEG 4000, and 5 to 10% isopropanol using microseeding (38). Rod-shaped crystals grew within 7 d. TFD-EE N6W was mixed 1:1 with $TbCl_3$ and concentrated to 8 mg/mL. Crystals grew in 0.1 M Hepes pH 7.5, 20% PEG 4000, and 10% isopropanol, containing Gly_3 as an additive. The cryo-protectant solutions for all TFD-EE crystals consisted of the reservoir solution supplemented with 20% ethylene glycol. Crystals were cryo-cooled in liquid nitrogen.

Diffraction data from DeNovoTIM15 and TFD-HE were collected at the Advanced Light Source (ALS) Beamline 8.2.1. TFD-EE data were collected at the synchrotron beamline X06SA (PXI) and Tb(III)-bound TFD-EE N6W at X06DA (PXIII) at the Swiss Light Source. In all cases, diffraction was collected at 1.0 Å with the crystals kept at 100 K. The program XDS was used for initial data processing (39).

All structures were refined and built using the Phenix and CCP4 software suites (40, 41), using the program PHASER (42) to phase the structures by molecular replacement. DeNovoTIM15 was phased using the unaltered Rosetta design model. TFD-EE and TFD-HE structures were phased by searching separately for the TIM domain and FD domain. In the case of the third TFD-EE dimer and TFD-HE structure, the FD dimer could not be placed by phaser, but was manually built by modeling helices into F_o-F_c difference density and iterative manual model building. Further model building and refinement were done in iterative cycles using the programs Coot (43) and phenix.refine (40) or Refmac (44), including translation–libration–screw refinement (45) and reference model restraints to the Rosetta models. The model quality was validated using MolProbity (46). Structure illustrations were prepared with PyMOL (47), while estimates of volume of enclosed cavity were obtained using the CASTp server (48), based upon models with bulk solvent-accessible channels plugged with dummy atoms. Structures were deposited to the Protein Data Bank (PDB) with accession numbers 6WVS, 6WXP, 6WXO, and 6ZV9. Summary statistics are presented in Table 1.

ICP-MS Measurements. Element screening and Ca quantification was carried out using MICAP-MS systems. The instruments are prototype instruments developed in the Laboratory for Inorganic Chemistry at ETH Zürich specialized for measuring metal ions (49). A TFD-EE sample was buffer-exchanged with 1 mM Hepes pH 7.5 using a PD10 desalting column. The same buffer was used as the reference sample. First, a MICAP-MS element screen was performed at 45 μM TIM-EE monomer to generate semiquantitative estimates of element concentrations. Subsequently, the amount of calcium in the sample was determined quantitatively using a MICAP-MS experiment, using an external calibration between 0 and 500 μg/L Ca, with Sc as an internal standard. Further experimental details are described in the *SI Appendix*.

Lanthanide-Binding Measurements. Tryptophan-enhanced terbium luminescence was measured in 96-well plates using a Varioskan Flash plate reader in time-resolved fluorescence mode (TRF, delay time: 50 μs, integration time: 1 ms). This mode allows for suppression of all background signals, collecting only the long-lived terbium luminescence in a pulsed experiment. All experiments were performed in 25 mM Hepes pH 7.5, 100 mM NaCl, and treated with Chelex100 resin. TFD variants were also dialyzed in the presence of Chelex100 resin prior to measurement to remove residual bound metals. TFD-QQ N6W served as a negative control (metal-binding site knockout, E to Q mutations). Emission spectra (450 to 650 nm) were collected upon excitation of the tryptophan antenna at 280 nm, after 3 h incubation of 10 μM TFD dimer and 20 μM $TbCl_3$. Time courses were collected with the same plate reader setup monitoring the Tb(III) luminescence peak (λ_{ex} = 280 nm, λ_{em} = 545 nm) over time. Titrations were performed with 10 μM TFD-EE N6W and various preincubated mixtures of EDTA and $TbCl_3$. The titration curve was derived from spectra recorded after 24 h incubation, estimating the concentration of free Tb(III) based on the K_D(Tb/EDTA) = 1.6×10^{-18} M as a reference (31). Binding of other lanthanide ions was measured in displacement titrations of 5 μM preincubated Tb(III)/TFD-EE N6W complex with increasing amounts of $CeCl_3$, $EuCl_3$ or $YbCl_3$.

Data Availability. The crystal structures data of DeNovoTIM15, TFD-HE, TFD-EE, and Tb(III)-bound TFD-EE N6W have been deposited in the PDB with accession codes 6WVS, 6WXP, 6WXO, and 6ZV9.

ACKNOWLEDGMENTS. This work was supported by the Howard Hughes Medical Institute, ETH Zürich, and the Swiss National Science Foundation. This work was supported by Defense Threat Reduction Agency grant HDTRA1-19-1-0003. C.Z. is supported by an Ambizione Grant (PZ00P2_179887) of the

Swiss National Science Foundation. S.J.C. is supported by Canadian Institutes of Health Research postdoctoral fellowship MFE-152581 #382353. We thank the staff members of the Protein Crystallization Center at the University of Zurich and the Swiss Light Source for help with TFD-EE protein crystallization and diffraction data collection, respectively, as well as Michael Groll and Peer Mittl for help with structure refinement. We also thank Alex Kang, Asim Bera at the Institute for Protein Design for crystallization of TFD-HE, and the staff at the ALS Beamline 8.1.2 for their help in collecting diffraction data. We thank the Günther group at ETH Zürich for MICAP-MS data collection.

1. R. Sterner, B. Höcker, Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold. *Chem. Rev.* **105**, 4038–4055 (2005).
2. R. K. Wierenga, The TIM-barrel fold: A versatile framework for efficient enzymes. *FEBS Lett.* **492**, 193–198 (2001).
3. N. Nagano, C. A. Orengo, J. M. Thornton, One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.* **321**, 741–765 (2002).
4. A. Tóth-Petróczy, D. S. Tawfik, The robustness and innovability of protein folds. *Curr. Opin. Struct. Biol.* **26**, 131–138 (2014).
5. E. Dellus-Gur, A. Toth-Petroczy, M. Elias, D. S. Tawfik, What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. *J. Mol. Biol.* **425**, 2609–2621 (2013).
6. R. Blomberg *et al*., Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **503**, 418–421 (2013).
7. L. Giger *et al*., Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9**, 494–498 (2013).
8. O. Khersonsky *et al*., Optimization of the in-silico-designed kemp eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.* **407**, 391–412 (2011).
9. R. Obexer *et al*., Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50–56 (2017).
10. H. K. Privett *et al*., Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3790–3795 (2012).
11. P. Sharma, P. Kaila, P. Guptasarma, Creation of active TIM barrel enzymes through genetic fusion of half-barrel domain constructs derived from two distantly related glycosyl hydrolases. *FEBS J.* **283**, 4340–4356 (2016).
12. J. Wang *et al*., Recurring sequence-structure motifs in ($\beta\alpha$)8-barrel proteins and experimental optimization of a chimeric protein designed based on such motifs. *Biochim. Biophys. Acta Proteins Proteom.* **1865**, 165–175 (2017).
13. N. Preiswerk *et al*., Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8013–8018 (2014).
14. O. Khersonsky, S. J. Fleishman, Why reinvent the wheel? Building new proteins based on ready-made parts. *Protein Sci.* **25**, 1179–1187 (2016).
15. G. Lapidoth *et al*., Highly active enzymes by automated combinatorial backbone assembly and sequence design. *Nat. Commun.* **9**, 2780 (2018).
16. P.-S. Huang *et al*., De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12**, 29–34 (2016).
17. A. Leaver-Fay *et al*., "Rosetta3: An object-oriented software suite for the simulation and design of macromolecules" in *Methods in Enzymology*, M. L. Johnson, L. Brand, Eds. (Computer Methods, Part C, Academic Press, 2011), pp. 545–574.
18. S. Romero-Romero *et al*, Epistasis on the stability landscape of de novo TIM barrels explored by a modular design approach. *bioRxiv*:10.1101/2020.09.29.319103 (1 October 2020).
19. T. Cheisson, E. J. Schelter, Rare earth elements: Mendeleev's bane, modern marvels. *Science* **363**, 489–493 (2019).
20. E. G. Moore, A. P. S. Samuel, K. N. Raymond, From antenna to assay: Lessons learned in lanthanide luminescence. *Acc. Chem. Res.* **42**, 542–552 (2009).
21. L. J. Martin, B. Imperiali, "The best and the brightest: Exploiting tryptophan-sensitized Tb3+ luminescence to engineer lanthanide-binding tags" in *Peptide Libraries: Methods and Protocols*, R. Derda, Ed. (Methods in Molecular Biology, Springer, 2015), pp. 201–220.
22. M. R. Berwick *et al*., De novo design of Ln(III) coiled coils for imaging applications. *J. Am. Chem. Soc.* **136**, 1166–1169 (2014).
23. M. R. Berwick *et al*., Location dependent coordination chemistry and MRI relaxivity, in *de novo* designed lanthanide coiled coils. *Chem. Sci. (Camb.)* **7**, 2207–2216 (2016).
24. H. Pellissier, Recent developments in enantioselective lanthanide-catalyzed transformations. *Coord. Chem. Rev.* **336**, 96–151 (2017).
25. J. A. Cotruvo Jr, The chemistry of lanthanides in biology: Recent discoveries, emerging principles, and technological applications. *ACS Cent. Sci.* **5**, 1496–1506 (2019).
26. L. J. Daumann, Essential and ubiquitous: The emergence of lanthanide metallobiochemistry. *Angew. Chem. Int. Ed. Engl.* **58**, 12795–12802 (2019).
27. Y.-R. Lin, N. Koga, S. M. Vorobiev, D. Baker, Cyclic oligomer design with de novo $\alpha\beta$-proteins. *Protein Sci.* **26**, 2187–2194 (2017).
28. G. Dantas *et al*., Mis-translation of a computationally designed protein yields an exceptionally stable homodimer: Implications for protein engineering and evolution. *J. Mol. Biol.* **362**, 1004–1024 (2006).
29. B. Kuhlman *et al*., Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
30. P.-S. Huang *et al*., RosettaRemodel: A generalized framework for flexible backbone protein design. *PLoS One* **6**, e24109 (2011).
31. D. C. Harris, *Quantitative Chemical Analysis* (WH Freeman, ed. 9, 2015).
32. I. C. King *et al*., Precise assembly of complex beta sheet topologies from de novo designed building blocks. *eLife* **4**, e11012 (2015).
33. J. P. Erzberger, J. M. Berger, Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 93–114 (2006).
34. S. Wilkens, Structure and mechanism of ABC transporters. *F1000Prime Rep.* **7**, 14 (2015).
35. V. Balaram, Rare earth elements: A review of applications, occurrence, exploration, analysis, recycling, and environmental impact. *Geosci. Front.* **10**, 1285–1303 (2019).
36. G. J.-P. Deblonde *et al*., Selective and efficient biomacromolecular extraction of rare-earth elements using lanmodulin. *Inorg. Chem.* **59**, 11855–11867 (2020).
37. F. Gorrec, The MORPHEUS protein crystallization screen. *J. Appl. Cryst.* **42**, 1035–1042 (2009).
38. T. Bergfors, Seeds to crystals. *J. Struct. Biol.* **142**, 66–76 (2003).
39. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
40. D. Liebschner *et al*., Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).
41. M. D. Winn *et al*., Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
42. A. J. McCoy *et al*., Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
43. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
44. G. N. Murshudov *et al*., REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
45. M. D. Winn, M. N. Isupov, G. N. Murshudov, Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 122–133 (2001).
46. I. W. Davis *et al*., MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
47. W. L. DeLano, *The PyMOL Molecular Graphics System* (Schrödinger, LLC, New York, 2010).
48. W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018).
49. M. Schild *et al*., Replacing the argon ICP: Nitrogen microwave inductively coupled atmospheric-pressure plasma (MICAP) for mass spectrometry. *Anal. Chem.* **90**, 13443–13450 (2018).