PNAS
www.pnas.org

Supplementary Information for

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

Shane Caldwell, Ian Haydon, Nikoletta Piperidou, Po-Ssu Huang, Matthew Bick, H. Sebastian Sjöström, Donald Hilvert, David Baker*, and Cathleen Zeymer*

*David Baker and Cathleen Zeymer
Email: dabaker@uw.edu and cathleen.zeymer@tum.de

This PDF file includes:

Supplementary Text

Methods

Gene synthesis, protein expression and purification

Proteins TFD-EE as well as the sTIM-EEEE protein and TFD-HE prepared for CD were encoded in synthetic genes optimized for expression in *E. coli.* The genes were ordered from TWIST Bioscience and subcloned in a pETM-11 vector in order to produce proteins with N-terminal, TEV-cleavable His-tag (construct: His_6 – Linker – TEV site – Protein sequence). Expression was done in *E. coli* BL21 Gold overnight at 18 °C after induction with 0.5 mM IPTG. The fusion proteins were purified by Ni-NTA chromatography under gravity flow, following a standard protocol with respective wash buffer (25 mM HEPES pH 7.5, 300 mM NaCl, 20 mM imidazole) and elution buffer (25 mM HEPES pH 7.5, 300 mM NaCl, 300 mM imidazole). 0.2 mg TEV protease per ml protein solution was added after elution to cleave the N-terminal His-tag. TEV cleavage was performed during overnight dialysis at 4 °C against 25 mM HEPES pH 7.5, 300 mM NaCl, 1 mM DTT, 0.5 mM EDTA. A second Ni-NTA chromatography step was performed to remove the His-tag with linker, uncut fusion protein, and the TEV protease, which itself possesses a His-Tag. The untagged protein of interest was then dialyzed against 25 mM HEPES pH 7.5 and concentrated to 15 mg/ml. The protein concentration was determined by absorption at 280 nm.

DeNovoTIM15 and TFD-HE prepared for crystallography were separately encoded in synthetic genes optimized for expression in *E. coli*. These constructs were ordered from GenScript, cloned directly into pET29b vectors (Novagen) for expression in *E. coli*. An N-terminal histidine tag was included in all constructs, as well as a TEV-cleavage site and tryptophan residue for ease of quantification (Full sequences provided in Supplemental Figure S3). Plasmids were transformed into *E. coli* Lemo21 (DE3) strain (NEB, cat. No. C2528J) for expression in autoinduction media (1). Overnight cultures in ZYP-0.8G were subcultured into 2.0 mL ZYP-0.8g and grown at 37°C

for 1 h before inoculation of 500mL of ZYP-5052 with 1.0 mL of this starter culture. Cultures were grown for 6h at 37°C prior to 22°C expression overnight. Cells were harvested by centrifugation at 5,000 g for 20 min, frozen to -80°C and then thawed and resuspended in lysis buffer (300 mM NaCl, 20mM Tris, 20 mM Imidazole, 5mM PMSF, 100 ug/mL Lysozyme, 10 ug/mL DNase, pH 8.0) and mixed to homogeneity before lysis in a microfluidizer. The lysate was clarified by centrifugation for 30min at 25,000 g before loading to a Nickel-NTA gravity flow column equilibrated with wash buffer (300 mM NaCl, 20mM Tris, 20 mM Imidazole, pH 8.0). Following washing with 10 column volumes of wash buffer, the protein was eluted in 5 CV of elution buffer (300 mM NaCl, 20mM Tris, 500 mM Imidazole, pH 8.0). Protein was concentrated in an Amicon centrifugal spin filter with a 3 kDa cutoff (Millipore Sigma) before running on an equilibrated SEC column (GE Healthcare, Superdex 75 Increase 300/10 on Akta Pure, buffer 10 mM NaCl, 20 mM Tris pH 8.0). Pure protein representing a dimeric peak was pooled and concentrated in a 3kDa Amicon to ~10 mg/mL.

ICP-MS measurements

Element screening and calcium quantification was carried out using microwave coupled atmospheric plasma (MICAP) mass spectrometer (MS) systems. The instruments are prototype instruments developed in the Laboratory for Inorganic Chemistry at ETH Zürich, where the conventional Ar-based inductively coupled plasma ion source was exchanged for a microwave sustained nitrogen plasma. The primary advantage of the new source in these experiments is the reduction of spectral interference by $^{40}Ar^{+}$, allowing for the use of the most abundant isotope of calcium (^{40}Ca) for quantification.

An initial element screening was carried out on a prototype MICAP-TOFMS unit as described by Schild et al., allowing for a rapid, comprehensive screening of element abundances. Operating conditions are listed in Table 2. Quantitative Ca-analysis was carried out using the same source

coupled to a quadrupole-based mass spectrometer, comprising the vacuum system, ion optics and MS part from an Elan 6100DRC^{plus} instrument (Perkin Elmer Sciex, Canada). Both setups used a concentric nebulizer (PFA ST Elemental Scientific, USA) and cyclonic spray chamber (Glass Expansion, Australia) for sample introduction. Operating conditions were adjusted to optimize ion signal intensities while minimizing oxide-ion abundances and instrumental background. A semi-quantitative approach was used to estimate element concentrations from the MICAP-TOFMS spectra by comparing ion signals of the elements that could unambiguously identified based on the isotope abundance pattern Concentration estimates are based on the quantified Ca-concentrations and using relative sensitivity factors. Reported values are in this case considered to be accurate within a margin of 50 %. Quantification on the MICAP-QMS was carried out by external calibration between 0 and 500 µg/L Ca, with Sc as internal standard. Calibration curves for $^{40}Ca^{+}$ and $^{44}Ca^{+}$ were linear across the calibration range and matrix effects, based on the sensitivity observed for the internal standard, were insignificant (< 5% sensitivity variation).

ICP-MS was performed to identify the metal that bound to TFD-EE during production and/or purification of the protein. A protein 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, an ICP-MS element screen was performed at 45 µM TIM-EE monomer. Subsequently, the amount of calcium in the sample was determined quantitatively using a MICAPMS experiment.

Rosetta Remodel input

The TFD model was generated using Rosetta remodel with the following input command line options, blueprint and constraint files. For a model that adds two 3-residue linkers, the following files are used:

```
command-line arguments for rosetta remodel (flags):

-constraints:cst_file constraints/constraints_3_3.cst
-symmetry:symmetry_definition input/symdef.sym
-remodel:blueprint blueprints/blueprint_3_3.bp
-out:prefix TFD_3_3
-s input/half_TIM_input.pdb
-remodel:domainFusion:insert_segment_from_pdb
input/FD_monomer_input.pdb
-remodel:use_cart_relax
-set_weights pro_close 0
-num_trajectory 5
-nstruct 5
-chain A
-max_linear_chainbreak 1
-overwrite
-jd2:no_output
-ex1
-ex2
```

Description of Remodel Flags file contents

-constraints:cst_file constraints/constraints_3_3.cst

File that specifies constraints that will be applied to the design. In this case, these constraints are inter-chain distances within the dimeric ferredoxin between the inserted domain and it's symmetric fusion partner. See the .cst file below

-symmetry:symmetry_definition input/symdef.sym

File that specifies symmetry of the system, in this case C2 or two-fold cyclically symmetric see the .sym file below

-remodel:blueprint blueprints/blueprint_3_3.bp

File that specifies The numbers and nature of residues built into the output model, and their flexibility to sample rotamers during model building. See the .bp file below.

-out:prefix TFD_3_3

Prefix applied to output files. Optional.

-s input/half_TIM_input.pdb

PDB file of the domain in which the insertion will be added (monomer, in this case a half-TIM barrel)

-remodel:domainFusion:insert_segment_from_pdb input/FD_monomer_input.pdb

PDB file of the inserted domain (monomer)

-remodel:use_cart_relax

Allow Remodel to minimize and repack residues in cartesian space throughout model building

-set_weights pro_close 0

Modification to rosetta score function to avoid triggering bug.

-num_trajectory 5

Number of independent remodelling trajectories to carry out per cycle. Best model will be returned.

-nstruct 5

Number of cycles to build models per instantiation

-chain A

Specify model building of chain A of input structure.

-max_linear_chainbreak 1

Terminate trajectory if the first gap is not successfully closed.

-overwrite

Replace previous outputs. Optional.

-jd2:no_output

Silence output from jd2 functions. Optional.

-ex1

-ex2

During relax steps, sample extended sets of rotamers to find better energy minima.

constraints_3_3.cst:

```
AtomPair O 86 N 257 HARMONIC 2.8 0.2
AtomPair N 86 O 257 HARMONIC 2.8 0.2
AtomPair O 88 N 255 HARMONIC 2.8 0.2
AtomPair N 88 O 255 HARMONIC 2.8 0.2
AtomPair CA 73 CA 242 HARMONIC 7.1 0.5
```

```
symdef.sym:
symmetry_name reTIM26_C2_input_2
E = 2*VRT0_base + 1*(VRT0_base:VRT1_base)
anchor_residue COM
virtual_coordinates_start
xyz VRT0 0.2343902,-0.9719931,-0.0170486 0.9721236,0.2342398,0.0103677
0.0096813,0.0910492,-0.3335388
xyz VRT0_base 0.2343902,-0.9719931,-0.0170486
0.9721236,0.2342398,0.0103677 -1.9238571,8.1092418,-0.1929011
xyz VRT1 -0.2343902,0.9719931,0.0170486 -0.9721236,-0.2342398,-
0.0103677 0.0096813,0.0910492,-0.3335388
xyz VRT1_base -0.2343902,0.9719931,0.0170486 -0.9721236,-0.2342398,-
0.0103677 1.9432197,-7.9271433,-0.4741765
xyz VRT 0.0000000,-0.9998194,-0.0190038 0.9999815,-0.0001156,0.0060828
0.0096813,-0.9087517,-0.3525422
virtual_coordinates_stop
connect_virtual JUMP0_to_com VRT0 VRT0_base
connect_virtual JUMP0_to_subunit VRT0_base SUBUNIT
connect_virtual JUMP1_to_com VRT1 VRT1_base
connect_virtual JUMP1_to_subunit VRT1_base SUBUNIT
connect_virtual JUMP0 VRT VRT0
connect_virtual JUMP1 VRT0 VRT1
set_dof JUMP0_to_com x(8.24922789260257) angle_x
set_dof JUMP0_to_subunit angle_x angle_y angle_z
set_jump_group JUMPGROUP2 JUMP0_to_com JUMP1_to_com
set_jump_group JUMPGROUP3 JUMP1_to_subunit JUMP0_to_subunit
```

blueprint_3_3.bp:

```
 1 D .
 2 I .
 3 L .
 4 I .
 5 V .
 6 N .
 7 A .
 8 T .
 9 K .
10 V .
11 D .
12 E .
13 M .
14 L .
15 K .
16 Q .
17 V .
18 E .
19 I .
20 L .
21 R .
22 R .
23 L .
24 G .
25 A .
26 K .
27 Q .
28 I .
29 A .
30 V .
31 N .
32 S .
33 S .
34 D .
35 W .
36 R .
37 I .
38 L .
39 Q .
40 E .
41 A .
42 L .
43 K .
44 K .
45 G .
46 G .
47 D .
48 I .
49 L .
```

```
50 I .
51 V .
52 N D PIKAA N
0 x D
0 x D
0 x D
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
```

0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x I NATAA
0 x D
0 x D
0 x D
53 D D PIKAA D
54 V .
55 D .
56 E .
57 M .
58 L .
59 K .
60 Q .
61 V .
62 E .
63 I .
64 L .
65 R .
66 R .
67 L .
68 G .
69 A .
70 K .
71 Q .
72 I .
73 A .
74 V .

75 N .
76 S .
77 D .
78 D .
79 W .
80 R .
81 I .
82 L .
83 Q .
84 E .
85 A .
86 L .
87 K .
88 K .
89 G .
90 G .

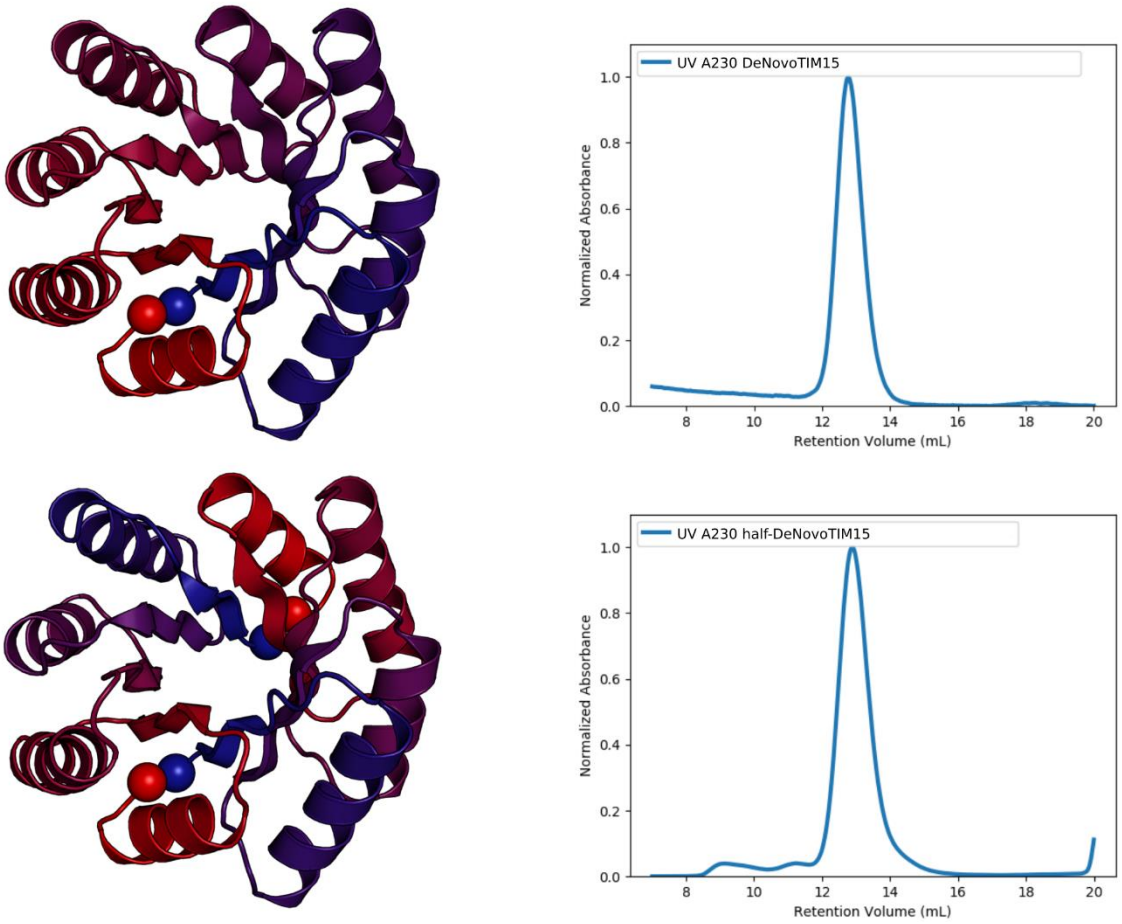

Fig S1: DeNovoTIM15 and homodimeric half-construct of DeNovoTIM15 (half-DeNovoTIM15). Both proteins elute at the same retention volume following size-exclusion chromatography on a Superdex 75 Increase column.

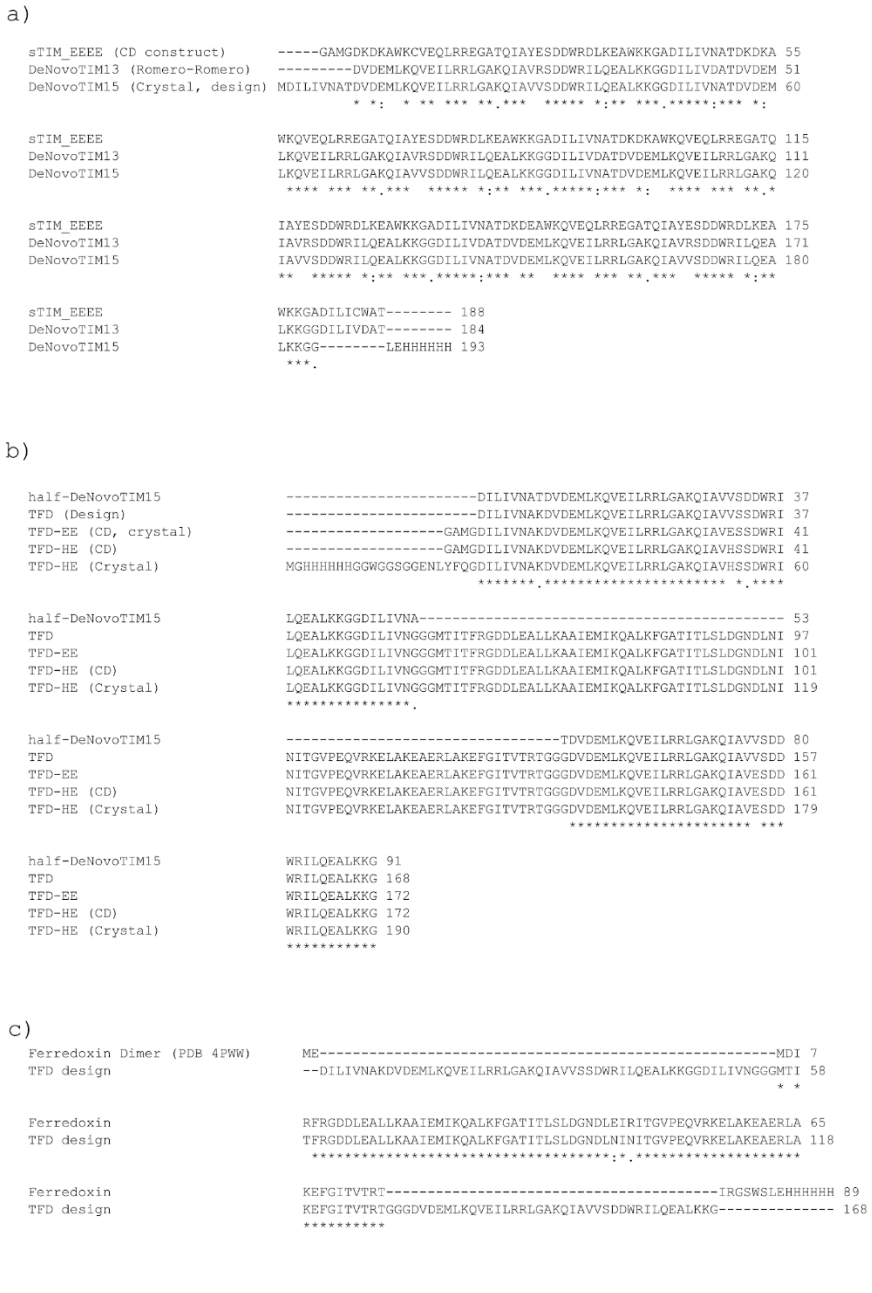

Fig S2: Sequences of the proteins studied and reported in this paper, and alignments thereof. a) *de novo* TIM barrels discussed or used in this paper b) TFD constructs aligned to the scaffold homodimeric half-TIM barrel. c) The insert ferredoxin domain aligned to the successfully built TFD protein. Alignments were constructed with Clustal Omega and manually edited for clarity.

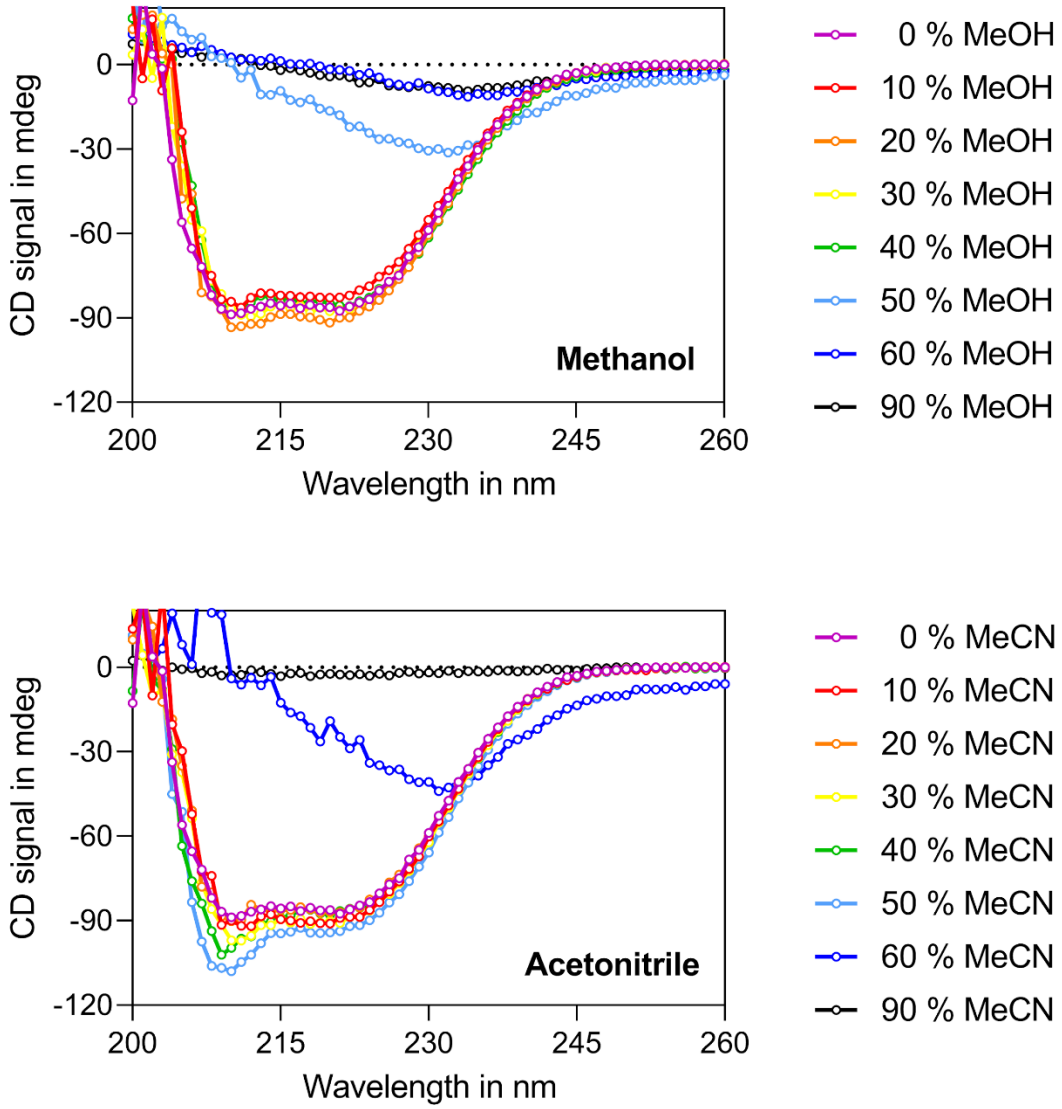

Fig S3: Stability of the TFD-EE scaffold in the presence of co-solvents. CD spectra were recorded after incubating the protein in 0-90 % methanol or acetonitrile for 20 min at 25 °C. The protein remained fully folded in up to 40 % methanol and 50 % acetonitrile. Precipitation was observed at > 60% co-solvent.

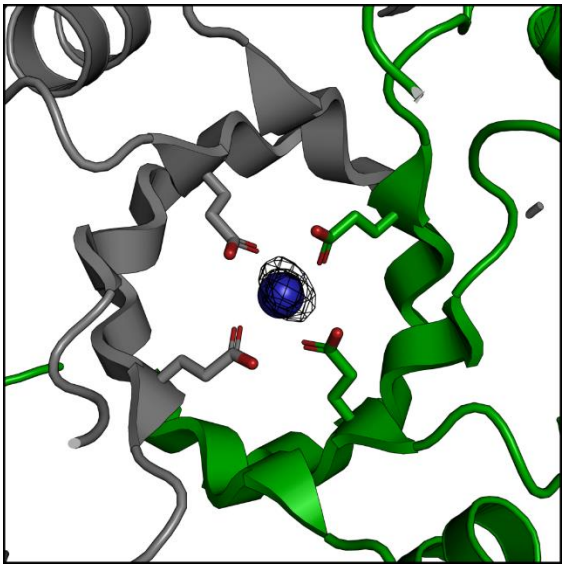

Fig S4: Electron density from the TFD-EE crystal structure, indicating that a metal ion (modelled as Ca^{2+}) is bound in the protein coordinated by four glutamate residues, at positions 31 and 154 of the homodimer. Electron density represents an omit map at 5.0 σ, carved at 6.0 Å from the modelled ion.

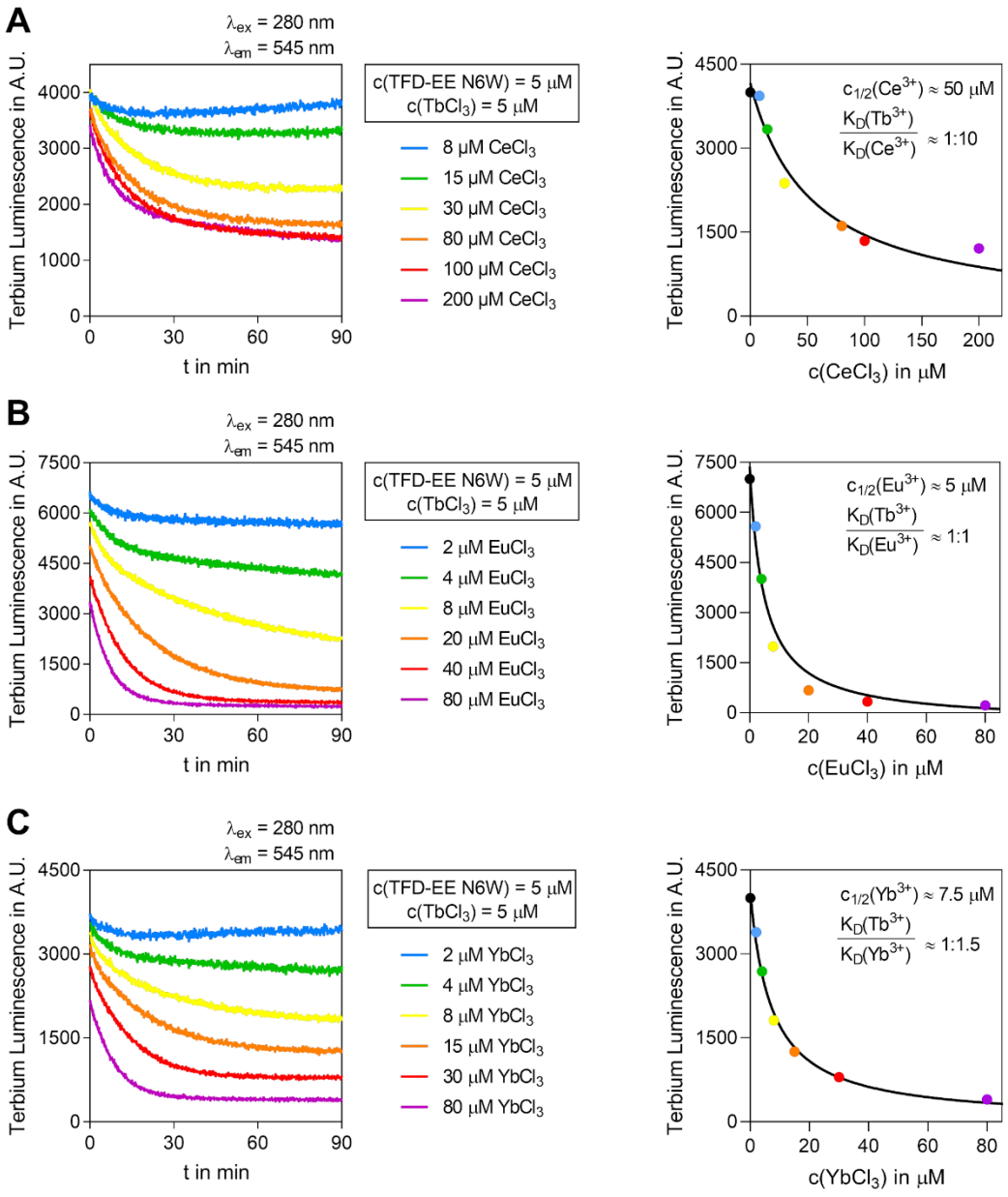

Fig S5: Binding of other lanthanides, such as Ce(III), Eu(III), and Yb(III), was measured in displacement titrations using Tb(III)-bound TFD-EE N6W and observing the tryptophan-enhanced Tb(III) luminescence decrease over time. The K_D ratio was estimated from the extracted equilibrium curve. Whereas Eu(III) and Yb(III) bind with affinities comparable to Tb(III), Ce(III) binds approximately 10-fold weaker to TFD-EE N6W.

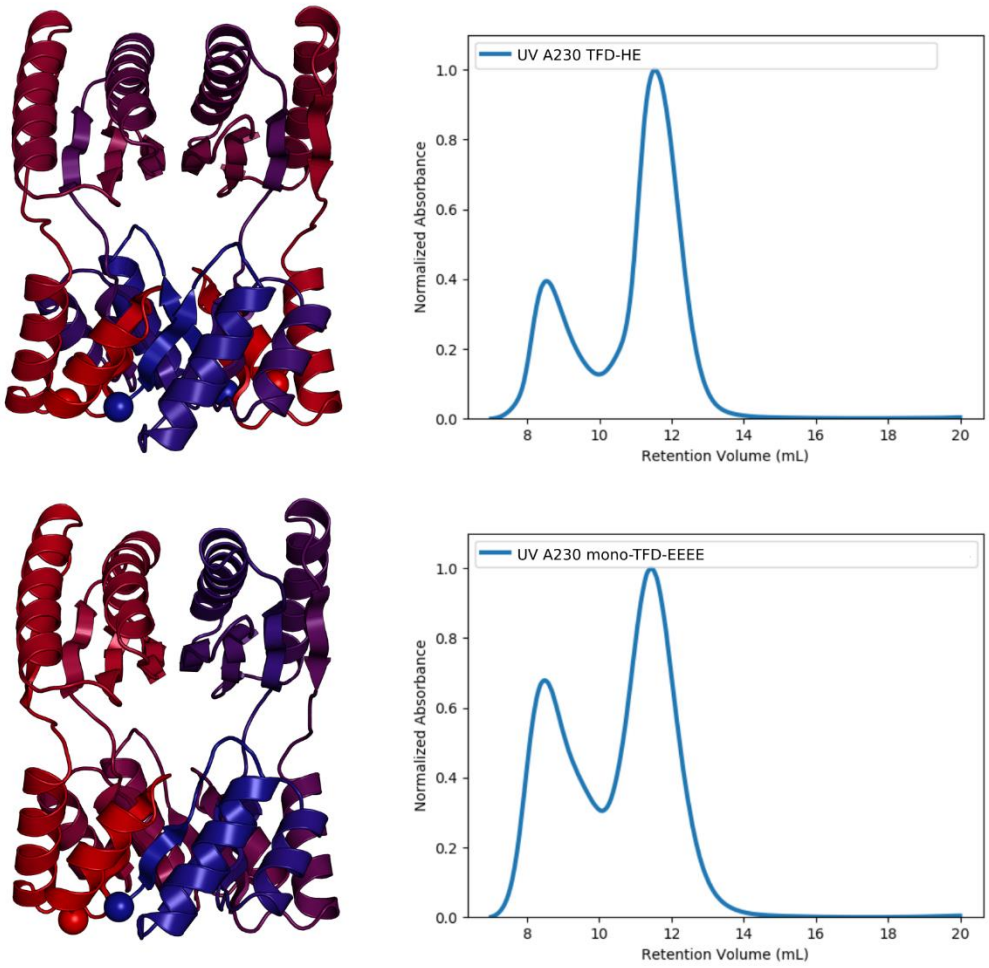

Fig S6: Model of the homodimeric TFD protein (top) and the asymmetric, monomeric variant mono-TFD created by joining the ends of the homodimer together, allowing for fully asymmetric customization of the protein (bottom). Elution of the TFD-HE variant from size-exclusion chromatography (Superdex 75 increase) followed by Nickel-NTA purified single-chain asymmetric monomer, mono-TFD-EEEE. Both chromatograms are from intermediate protein preparations and contain some aggregate in the void volume (~8mL), but the main chromatographic peak for the dimeric and fused monomer show equivalent retention volumes of ~11.75 mL. No free half-TIM barrel construct is observed; all is consistent with the full TFD homodimeric protein.

Table S1: Operating conditions of MICAP-TOFMS und MICAP-QMS

	MICAP-TOFMS			MICAP-QMS	
MW power, W	1500			1500	
Coolant Gas N_2, L/min	14			14	
Auxiliary Gas N_2, L/min	1.1			1.0	
Nebulizer Gas N_2, L/min	0.9			1.2	
Sample uptake, mL/min	0.3			0.28	
Integration time, ms	0.03			1000	
Replicates	100			10	
CeO^+/Ce^+, %	3.1			2.9	
Instr. Backgr., cps	< 10			< 100	
Sensitivity, kcps/(g/L)	$^{59}Co^+$	$^{133}Cs^+$	$^{238}U+$	$^{40}Ca^+$	$^{45}Sc^+$
	0.11	0.75	0.68	8.2	12.7

Table S2: Concentrations of ions measured for TFD-EE sample and buffer

	TFD-EE		Buffer	
	Conc.	CI (95%)	Conc.	CI (95%)
Element	g/L	g/L	g/L	g/L
Ca	181.5	1.7	46.1	0.5
Screening*				
Mg	100		1	
K	300		40	
Fe	1		<0.05	
Ni	5		<0.05	
Cu	0.5		<0.05	
Zn	2		<0.05	
Rb	0.3		0.02	
Sr	0.6		0.1	
Ag	0.2		<0.01	
Cd	1		<0.02	
Sb	0.5		<0.01	
Ba	0.2		0.02	
La	0.02		<0.005	
Ce	0.02		<0.005	
Eu	<0.005		0.1	
Pb	0.02		0.01	

SI References

1. F. W. Studier, Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).