Bioinformatics classification of the MgtE Mg^{2+} channel and de novo protein design for the stabilization of its novel subclass

Zhixuan Zhao[1], Kimiho Omae[2], Wataru Iwasaki[2,3], Ziyi Zhang[1], Fazhi Pan[1], Eun-Jin Lee[4], Motoyuki Hattori[1,*]

[1]*State Key Laboratory of Genetics and Development of Complex Phenotypes, Collaborative Innovation Center of Genetics and Development, Department of Physiology and Neurobiology, School of Life Sciences, Fudan University, Shanghai China;*

[2]*Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan*

[3]*Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan*

[4]*Department of Life Sciences, Korea University, Seoul, South Korea*

***Correspondence address.** E-mail: hattorim@fudan.edu.cn (M.H.)

Abstract

MgtE channels play crucial roles in Mg^{2+} homeostasis and are implicated in bacterial survival under antibiotic exposure. Previous structural and biophysical studies have predominantly focused on the *Thermus thermophilus* MgtE, leaving the structural and mechanistic diversity of MgtE family proteins largely unexplored. In this study, using a genome mining approach, we identified diverse MgtE homologs, including a novel subclass termed the "mini-N type," which lacks the canonical cytoplasmic N and CBS domains but possesses a unique small N-like domain. Despite extensive expression screening, mini-N type homologs could not be stably purified. To address this issue, we designed a series of de novo proteins and determined their crystal structures. A selected de novo protein was fused to a mini-N type MgtE, enabling successful purification and preliminary cryo-EM imaging. Our findings demonstrate that de novo designed protein fusions can serve as powerful tools for stabilizing and purifying otherwise unstable membrane proteins, opening new avenues for the structural and functional studies of otherwise inaccessible membrane proteins.

Key words: protein design, artificial intelligence, genome mining, bioinformatics

Introduction

MgtE represents a widely conserved Mg^{2+} channel in bacteria that plays a critical role in Mg^{2+} homeostasis [1-4] and contributes to bacterial survival under antibiotic exposure

by counteracting hyperpolarization [5]. In *Pseudomonas aeruginosa*, MgtE inhibits transcription of the type III secretion system-a critical virulence factor-by promoting the expression of the translation-inhibitory sRNAs RsmY and RmsZ [6, 7]. In humans, MgtE homologues are referred to as SLC41A1 to SLC41A3 and have been implicated in various physiological functions and diseases, including Parkinson's disease [8-11]. Accordingly, the MgtE/SLC41 family of proteins has attracted considerable interest as a potential therapeutic target.

To date, extensive structural and mechanistic studies have been conducted on MgtE [12-14]. These studies have revealed that MgtE functions as a "Mg^{2+}-gated Mg^{2+} channel," where Mg^{2+} binding to the cytoplasmic domain leads to channel inactivation [3, 15-17]. High-resolution structural analyses of the MgtE transmembrane (TM) domain have elucidated the mechanism of Mg^{2+} selectivity [18, 19], while the cryo-EM structure of MgtE in the absence of Mg^{2+} have provided insights into the gating mechanism [20].

However, these studies have predominantly focused on the *Thermus thermophilus* MgtE (TtMgtE), and structural and biochemical data for MgtE homologues from other species are extremely limited [21, 22]. As a result, the structural and functional diversity of the broadly conserved MgtE family remains largely unexplored. In parallel, the recent explosion in genomic data has made it increasingly feasible to investigate the diversity of specific protein families using genome mining approaches [23, 24].

One major challenge in studying membrane proteins such as MgtE lies in their generally low expression levels and poor stability in detergents [25], which significantly hampers structural and functional analyses. Recently, AI-based de novo protein design

technologies, such as ProteinMPNN [26] and RFdiffusion [27], have rapidly advanced [28] and are increasingly expected to facilitate research on membrane proteins [29]. De novo designed proteins typically exhibit high stability, and fusing them to target membrane proteins may enhance expression and stability, potentially enabling the preparation of large quantities of stable membrane protein samples.

In this study, we conducted a bioinformatics analysis of the MgtE family and identified a novel subclass of MgtE proteins with unique domain architectures. Despite extensive expression screening, these proteins could not be stably purified. To overcome this challenge, we performed de novo protein design and successfully determined the crystal structures of several designed proteins. Furthermore, by generating fusion constructs with the de novo designed protein and the novel MgtE subclass, we achieved successful expression and purification. Overall, this study highlights a promising strategy for stabilizing and purifying membrane proteins using a de novo designed protein fusion approach.

Materials and Methods

Bioinformatics analysis of MgtE proteins

The prokaryotic proteomes and taxonomic information were downloaded from Genome Taxonomy Database Release 06-RS202 [30]. Only the high-quality (HQ) genomes (completeness ≥95% and contamination ≤5%) based on CheckM [31] were used for the analysis. The genome quality of p__Patescibacteria was estimated by using custom marker set provided by CheckM. The hidden Markov model profile of Pfam PF01769

(Divalent cation transporter, MgtE) was used as the query for hmmsearch to find MgtE homologs from the proteomes [32, 33]. The domain architecture of MgtE was determined by InterProScan search [34]. The InterPro IDs IPR006667, IPR046342 and IPR006668 were used as proxies for transmembrane (TM), CBS and N domains of MgtE, respectively.

Data analysis and visualization were performed using R 4.3.2 [35] with the following packages: tidyverse [36], plyranges [37], Biostrings [38] and patchwork [39]. The code for the analysis is provided in the GitHub repository https://github.com/0mae/mgte_short.

Expression screening of MgtE by FSEC-Nb

MgtE genes and their mutants were synthesized by Azenta (Suzhou, China) and subcloned into a pETNb-nALFA vector [40] containing an N-terminal 8×His tag, an ALFA tag [41], and an HRV 3C protease cleavage site. *Escherichia coli* Rosetta (DE3) cells were transformed with the expression vectors and cultured in 8mL LB medium containing 50 μg/mL kanamycin and 30 μg/mL chloramphenicol at 37°C with the addition of 0.5 mM IPTG when the OD_{600} reached ~0.6. *E. coli cells* were then cultured at 18°C for 16 hours and harvested by centrifugation. The *E. coli* cell pellets were suspended in 400 μl of buffer TBS (50 mM Tris-HCl pH 8.0, 150 mM NaCl) containing 1 mM phenylmethylsulfonyl fluoride (PMSF) and disrupted by sonication. The cell debris was removed by centrifugation (20,000 × g, 10 min). The supernatant was solubilized with 500 μl of buffer TBS containing 2% (w:v) DDM and 0.5 mM PMSF for 1 hour, followed by ultracentrifugation (200,000 × g, 20 min). 1 μg of EGFP-tagged NbALFA was added to the half amount of the supernatant and incubated for 30 min. After

centrifugation (20,000 × g, 10 min), 50 μl of the sample was applied to a Superdex 200 Increase 10/300 GL column or a Superose 6 Increase 10/300 GL column (Cytiva, USA) equilibrated with buffer TBS containing 0.05% (w:v) DDM for FSEC analysis [42, 43]. Fluorescence was monitored using an RF-20Axs fluorescence detector (Shimadzu, Japan) (excitation: 480 nm, emission: 512 nm).

Expression and purification of MgtE

The large-scale expression of MgtE proteins followed the same method as small-scale expression screening and utilized 6 L of LB medium for expression. After harvesting by centrifugation, the *E. coli* cell pellets were resuspended and disrupted using a microfluidizer in TBS buffer. After centrifugation to remove cell debris, the supernatant was ultracentrifuged at 200,000 × g for 1 hour to isolate the membrane fraction. The membrane was solubilized in S buffer [50 mM Tris (pH 8.0), 150 mM NaCl, 2% n-Dodecyl-β-D-maltoside (DDM), 1 mM PMSF, and 1 mM β-Mercaptoethanol (β-ME)] for 1 hour. After ultracentrifugation at 200,000 × g for 1 hour, the supernatant was loaded onto Ni-NTA agarose beads (QIAGEN, Germany). The beads were washed with 10 column volumes of buffer A (50 mM Tris-HCl [pH 8.0], 150 mM NaCl, 20 mM imidazole, 0.03% DDM, 1 mM β-ME), and proteins were eluted using buffer B (50 mM Tris-HCl [pH 8.0], 150 mM NaCl, 250 mM imidazole, 0.03% DDM, 1 mM β-ME). The eluates were dialyzed against buffer C (20 mM HEPES [pH 7.0], 150 mM NaCl, 0.03% DDM, 1 mM β-ME) for 16 hours at 4°C. The protein was injected onto a Superdex 200 Increase 10/300 GL size-exclusion chromatography column (Cytiva, USA) and eluted using buffer

D [20 mM HEPES (pH 7.0), 150 mM NaCl, 0.03% DDM, 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP)].

De novo protein design

All de novo proteins in this study were designed using the ColabDesign framework (https://github.com/sokrypton/ColabDesign). Briefly, the backbone structures of ZZ1–ZZ3 proteins were generated using RFdiffusion [27] with a specified length of 200 amino acid residues and an unconditional folding mode. The corresponding sequences were then generated using ProteinMPNN [26], and the designed protein sequences were subsequently validated by AlphaFold [44]. The ZZ4-ZZ7 proteins were similarly designed but using the conditional folding mode.

Expression screening of de novo designed proteins

The DNA sequence encoding each de novo designed protein was synthesized by Azenta (Suzhou, China) and subcloned into a pETNb-cALFA vector [40] containing an HRV 3C protease cleavage site, an ALFA tag, and an 8×His tag at the C-terminus. *E. coli* Rosetta (DE3) cells were transformed with the expression vectors and cultured in 8 mL of LB medium containing 50 μg/mL ampicillin and 30 μg/mL chloramphenicol at 37°C. When the OD_{600} reached ~0.6, 0.5 mM IPTG was added to induce protein overexpression. The *E. coli* cells were then cultured at 18°C for 16 hours and harvested by centrifugation. The *E. coli* cell pellets were resuspended and disrupted by ultrasonication in 400 μL of TBS buffer [50 mM Tris-HCl (pH 8.0), 150 mM NaCl] supplemented with 1 mM PMSF.

After ultracentrifugation (200,000 × g, 20 min), 1 μg of EGFP-tagged NbALFA was added to the supernatant and incubated for 30 min. The mixture was then centrifuged at 20,000 × g for 10 min, and 50 μL of the sample was applied to a Superdex 200 Increase 10/300 GL column (Cytiva, USA) equilibrated with TBS buffer for the FSEC assay.

Expression and purification of de novo designed proteins

The large-scale expression of de novo designed proteins followed the same method as small-scale expression screening, utilizing 6 L of LB medium for expression. After harvesting by centrifugation, the *E. coli* cell pellets were resuspended in TBS buffer. Following centrifugation, the supernatant was loaded onto Ni-NTA agarose beads and washed with 10 column volumes of buffer E (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 20 mM imidazole). Proteins were then eluted using buffer F (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 250 mM imidazole). After elution, the ALFA tag and 8×His tag were cleaved by HRV 3C protease while the eluates were dialyzed against buffer G (20 mM HEPES, pH 7.0, 150 mM NaCl) for 16 h at 4°C. The sample was then reapplied to Ni-NTA beads, and the flow-through fractions were injected onto either a Superdex 200 Increase 10/300 GL or a Superdex 75 Increase 10/300 GL size-exclusion chromatography column (Cytiva, USA) in buffer G. The peak fractions were concentrated to 10 mg/mL for crystallization.

Crystallization

One microliter of ZZ1 protein was mixed with 1 μl of crystallization reservoir buffer H [0.05 M zinc acetate, 0.05 M MES, pH 6.0, 7.6% PEG 8000], while 1 μl of ZZ4 protein was mixed with 1 μl of crystallization reservoir buffer I [0.2 M $MgCl_2$, 0.1 M KCl, 0.025 M sodium citrate, pH 4.0, 24.5% PEG 400] for crystallization using the vapor diffusion method at 18°C. Crystals typically appeared within one week and were subsequently harvested in reservoir solutions supplemented with a final concentration of 30% glycerol for ZZ1 or 38% PEG 400 for ZZ4, respectively before flash-freezing in liquid nitrogen for X-ray diffraction experiments.

X-ray data collection and structure determination

X-ray diffraction experiments for ZZ1 and ZZ4 were performed at beamline BL02U1 of the Shanghai Synchrotron Radiation Facility (SSRF) [45]. Data processing was carried out using XDS [46]. The initial phase was determined by molecular replacement using Phaser [47], utilizing AlphaFold [44] models as templates. Manual model building and iterative refinement were performed with Coot [48] and PHENIX [49], respectively. Structural figures were prepared using PyMOL (https://pymol.org).

Construct design of ZZ1-fused VhMgtE

The ZZ1-fused VhMgtE harbors a 3- or 4-residue truncation at their C-terminus and are fused via one or two EK2R1 (AEEEKRK) helices [50]. The construct was validated by AlphaFold [44] before wet experiments.

Cryo-EM of ZZ1-fused MgtE

The purified ZZ1-fused MgtE protein was mixed with Amphipol A8-35 (Anatrace, USA) at a mass ratio of 1:10 and incubated for 16 hours. Bio-Beads SM-2 (Bio-Rad, USA) were then added to the mixture and incubated at 4°C for 4 hours to remove detergents. The complex was further purified by size-exclusion chromatography using a Superdex 200 10/300 column in buffer J (20 mM HEPES pH 7.0, 150 mM NaCl, 0.5mM TCEP). The peak fractions were concentrated to 2.0 mg/mL. Freshly prepared samples were incubated with 2 mM $MgCl_2$ on ice for 30 minutes before vitrification. Cryo-grid preparation was performed at 4°C and 100% humidity using a Vitrobot Mark IV (Thermo Fisher, USA). A 2.5 μL sample was applied to each freshly glow-discharged Au 300-mesh grid (Quantifoil, Germany, R1.2/1.3). The grids were then plunge-frozen in liquid ethane. Cryo-grids were screened using a 200 kV Glacios 2 microscope equipped with a Falcon 4i camera (Thermo Fisher, USA).

Data and code availability

The atomic coordinates of the de novo designed protein structures were deposited in the Protein Data Bank under accession code ID: 9LNI (ZZ1) and 9LX4 (ZZ2). All relevant data are included in the paper, its supplemental information or deposited in Mendeley

Data (https://data.mendeley.com/datasets/kdns5hnrmw/1) and GitHub (https://github.com/0mae/mgte_short).

Results and Discussion

Bioinformatics analysis of MgtE proteins identified the mini-N type of MgtE

We conducted an extensive search against prokaryotic genome data sets containing 27,742 genomes to identify genes encoding the TM domain of MgtE, which revealed a variety of domain architectures (Figure 1). The canonical type of MgtE (e.g., *Thermus thermophilus* MgtE) contains all three domains: the N domain, the CBS domain, and the TM domain (N-CBS-TM; Figure 1A). However, some MgtE proteins lack the N domain (CBS-TM), while others lack both the N and CBS domains, retaining only the TM domain (TM; Figure 1B). These proteins are not negligible in number (~2,600), and they were conserved in both bacteria and archaea (Table 1). Notably, some proteins lacking both the N and CBS domains possess tandem-repeat TM domains (TM-TM), similar to the eukaryotic MgtE homolog, the SLC41 family (Figure 1C).

Intriguingly, we also identified a novel type of MgtE gene that lacks both the N and CBS domains but contains an additional domain structurally similar to, yet smaller than, the canonical N domain (Figure 1C and S1). We designated this class of MgtE as the "mini-N" type (Figure 1D). The discovery of the mini-N-type MgtE was surprising, as in the canonical MgtE, the CBS domain is known to be essential for channel regulation, while the N domain itself does not appear to have a direct role in gating regulation[3].

Expression screening of the mini-N type MgtE proteins

To structurally and functionally characterize these various types of MgtE proteins, we initially selected mini-N-type MgtE proteins for expression screening before undertaking large-scale expression and purification. We expressed 12 genes encoding mini-N-type MgtE proteins in E. coli at a small scale and evaluated their expression profiles using fluorescence-detection size-exclusion chromatography (FSEC) [42, 51], specifically through nanobody (Nb)-based FSEC (FSEC-Nb)[40] (Figure 2). In the FSEC-Nb method, targeted membrane proteins are fused to an ALFA peptide tag [41], and solubilized whole-cell extracts were mixed with anti-ALFA tag Nb (NbALFA) fused to GFP for subsequent FSEC analysis to detect the complex of the target membrane protein and the Anti-ALFA tag NB fused to GFP. FSEC-Nb screening of mini-N-type MgtE homologs showed that multiple homologs (Nos. 5–9 in Figure 2) exhibited sharp, high-intensity peaks, indicating the high level of expression and monodispersity, so that they might be suitable for further large-scale expression and purification. However, in further expression and purification trials, these proteins showed a tendency to aggregate or precipitate during purification.

Design of de novo proteins

To overcome these challenges in the purification of mini-N-type MgtE homologs, we designed a series of de novo proteins (Figure 3 and Table S1) to serve as potential fusion partners. Protein fusion strategies have been widely employed for membrane protein

stabilization and structural studies [52, 53], as fusion partners can not only enhance the stability of membrane proteins but also facilitate crystallization and particle alignment during cryo-EM data processing. Furthermore, de novo designed proteins generally exhibit extremely high stability [27, 28], potentially making them particularly suitable as fusion partners for membrane proteins. Using RFdiffusion [27], we designed seven de novo proteins: three (ZZ1–ZZ3) using unconditional folding mode with a fixed length of 200 amino acid residues, and four (ZZ4–ZZ7) using conditional folding mode with lengths close to 200 residues. Validation through AlphaFold2 prediction yielded high pLDDT and pTM scores, with small root mean square deviation (RMSD) values ranging from approximately 0.4 to 0.6 Å (Table S1). In the unconditional folding mode, we generated all-α (ZZ1), α/β (ZZ2), and α+β (ZZ3) structures (Figure 3A-C). In the conditional folding mode, besides an all-α structure (ZZ5), we aimed to design β-barrel architectures containing internal α-helix elements (ZZ4, ZZ6, and ZZ7), reminiscent of the GFP fold [54]. Intriguingly, despite their relatively simple topologies, none of these designed proteins exhibited significant overall structural similarity to proteins in the PDB100 database as analyzed by FoldSeek [55] (minimum structural alignment overlap of 90% and an E-value cutoff of 0.01 as a threshold). These observations support the idea that the structural space accessible by de novo designed proteins is far broader than the space currently represented by naturally occurring protein structures [56].

Experimental verification of de novo designed proteins

The designed proteins were screened for expression in *E. coli* using the FSEC-Nb

method [40] prior to their structural analyses (Figure 4A). Although FSEC techniques are primarily utilized for screening membrane protein expression, they can also be effectively applied to soluble proteins [51]. All of seven designed proteins showed high and sharp peaks corresponding to the complex of the ALFA-tagged these proteins and NbALFA fused to GFP in FSEC (Figure 4A), and were subsequently purified for crystallization trials. Notably, the elution profiles of these proteins corresponded to monomeric molecular weights, as anticipated from our design. Crystals of ZZ1 and ZZ4 diffracted to resolutions of 1.61 and 2.03 Å, respectively, and their structures were successfully determined by molecular replacement using AlphaFold2 models as templates (Figure 4B and Table S2). Structural superposition revealed RMSD values of 0.65 Å and 0.83 Å for the Cα atoms of ZZ1 and ZZ4, respectively, between the AlphaFold-predicted models and crystal structures, demonstrating the high accuracy of our designs (Figure 4B).

Fusion of the de novo designed protein enabled the mini-N type MgtE purification

To fuse our de novo designed proteins to the mini-N type MgtE, we selected ZZ1 based on its α-helix bundle architecture, which is highly compatible with fusion to membrane proteins in various configurations, as well as its demonstrated stability evidenced by successful X-ray crystallography (Figure 4). For the mini-N type MgtE, we chose the homolog from Vibrio harveyi (VhMgtE) due to its favorable peak profile observed in FSEC-Nb analysis (No. 9 in Figure 2).

To fuse ZZ1 to VhMgtE, we employed an EK2R1 (AEEEKRK) helical linker (Figure 5A), known for its high stability [57] and utility in membrane protein structural

determination by fusion strategies [58]. We designed three constructs, fusing ZZ1 to the C-terminus of VhMgtE: Construct 1 (3-residue truncation at the C-terminus of VhMgtE with one EK2R1 linker), Construct 2 (3-residue truncation at the C-terminus of VhMgtE with two EK2R1 linkers), and Construct 3 (4-residue truncation at the C-terminus of VhMgtE with two EK2R1 linkers), as validated by AlphaFold2 structure prediction (Figure S1). These constructs were then screened for expression using FSEC-Nb (Figure 5B).

All three constructs exhibited expression levels similar to that of the wild type (Figure 5B). We initially selected Construct 3 for purification. Construct 3 was successfully purified (Figure 5C), reconstituted into amphipol A8-35, and showed a sharp, monodisperse peak in SEC analysis (Figure 5D). Furthermore, preliminary cryo-EM imaging of amphipol-reconstituted ZZ1-fused VhMgtE, using a 200 kV Glacios 2 microscope, showed monodisperse particles (Figure 5E).

Taken together, our results showed that the fusion strategy utilizing de novo designed proteins effectively enabled the purification of this novel subclass of MgtE protein.

Concluding remarks

Our bioinformatics pipeline identified a novel "mini-N" subclass of MgtE Mg^{2+} channels with atypical domain architectures, suggesting previously unrecognized structural diversity within the MgtE family. While we encountered challenges in expression and purification of the mini-N type MgtE, we tackled these technical issues using AI-driven de novo protein design and protein engineering. We designed stable de

novo proteins using state-of-the-art deep learning tools and determined the crystal structures of selected constructs, validating the design approach. Fusion of one such de novo protein (ZZ1) to a mini-N type MgtE homolog significantly enabled purification suitable for structural studies. Preliminary cryo-EM analysis further confirmed the structural homogeneity of the fusion construct. Overall, our work highlights the utility of AI-assisted de novo protein design in overcoming key bottlenecks in membrane protein research. The fusion strategy we developed may be broadly applicable to other challenging targets, paving the way for deeper insights into membrane protein structure and function.

Acknowledgments

We thank the staff from BL02U1 at Shanghai Synchrotron Radiation Facility (SSRF) for assistance during X-ray data collection. The diffraction experiments were performed at SSRF BL02U1 (Proposal No. 2020-SSRF-PT-014702). We appreciate Ruiliang Jia in Hattori lab for the technical support regarding the FSEC screening of de novo designed proteins. We also appreciate the kind support of Dr. Wataru Iwasaki as a former senior mentor to Kimiho Omae. Computation time was provided by the SuperComputer System, Institute for Chemical Research, Kyoto University.

Funding

This work was supported by funding from the National Natural Science Foundation of China (32411540020, 32471247, 32271244) to M.H.

Figure legends

Fig. 1. Bioinformatics analysis of MgtE proteins identified various types of MgtE.

(A) The canonical MgtE structure. The structure of *Thermus thermophilus* MgtE (PDB ID: 2ZY9). (B) The number and length of MgtE proteins identified by bioinformatics analysis. MgtE proteins are classified based on the domain components. (C) The length distribution of MgtE gene that lacks both the N and CBS domains. MgtE proteins are classified based on the number of TM domains. The novel MgtE protein is indicated by the red arrow. (D) The structure of the novel "mini-N" type MgtE. The structure of *Vibrio harveyi* MgtE (UniProt ID: A0A0D0IA63) predicted by AlphaFold2.

Fig. 2. Expression screening of mini-N type MgtE proteins using FSEC-Nb.

FSEC traces of unpurified ALFA-tagged mini-N type MgtE proteins mixed with EGFP-tagged NbALFA on a Superdex 200 Increase 10/300 GL column, detected via EGFP fluorescence. Expression of mini-N type MgtE proteins from the following species was screened: 1: *Endozoicomonas elysicola* (WP_081622142.1), 2: BD1-7 clade bacterium (CAA0079619.1), 3: *Endozoicomonas sp*. OPT23 (WP_228550828.1), 4: *Endozoicomonas numazuensis* (WP_034843082.1), 5: *Photobacterium frigidiphilum*

(WP_107243853.1), 6: *Enterovibrio baiacu* (WP_129495674.1), 7: *Vibrio thalassae* (WP_096991926.1), 8: *Vibrio* (Multispecies) (WP_146490653.1), 9: *Vibrio harveyi* BSW5 (A0A0D0IA63), 10: *Gloeocapsa sp.* PCC 7428 (WP_015189061.1), 11: *Nodularia sp.* NIES-3585 (WP_089089827.1), 12: *Pandoraea terrae* (WP_150700210.1).

Fig. 3. De novo protein design using RFdiffusion.

(A–G) Cartoon representations of de novo designed protein structures predicted by AlphaFold2.

Fig. 4. Expression screening and structural analysis of de novo designed proteins.

(A) FSEC traces of unpurified de novo designed proteins ZZ1–ZZ7 mixed with EGFP-tagged NbALFA on a Superdex 200 Increase 10/300 GL column, detected via EGFP fluorescence. (B–C) Crystal structures of de novo designed proteins ZZ1 (B) and ZZ4 (C). The AlphaFold2-predicted structures are superimposed and shown in gray.

Fig. 5. Purification and preliminary cryo-EM of ZZ1-fused VhMgtE.

(A) AlphaFold-predicted structure (Construct 3) and domain composition of the ZZ1-fused VhMgtE construct. (B) FSEC traces of unpurified ZZ1-fused VhMgtE constructs mixed with EGFP-tagged NbALFA on a Superose 6 Increase 10/300 GL column, detected via EGFP fluorescence. (C) SDS-PAGE of purified ZZ1-fused VhMgtE construct 3. Lane 1: Flowthrough from Ni-NTA affinity chromatography. Lane 2: Elution fraction. (D) Size-exclusion chromatography of ZZ1-fused VhMgtE construct 3 after amphipol

reconstitution, detected via UV absorbance. (E) Cryo-EM image of ZZ1-fused VhMgtE construct 3.

Table 1. The number of MgtE homologs

Fig. S1. AlphaFold2 statistics of VhMgtE constructs

(A) pLDDT scores for each amino acid position. (B) PAE maps.

Table S1. Sequence and AlphaFold2 statistics of de novo designed proteins

Table S2. X-ray data collection and refinement statistics

References

1 Smith RL, Thompson LJ, Maguire ME. Cloning and characterization of MgtE, a putative new class of Mg2+ transporter from Bacillus firmus OF4. J Bacteriol 1995, 177: 1233-1238

2 Townsend DE, Esenwine AJ, George J, 3rd, Bross D, Maguire ME, Smith RL. Cloning of the mgtE Mg2+ transporter from Providencia stuartii and the distribution of mgtE in gram-negative and gram-positive bacteria. J Bacteriol 1995, 177: 5350-5354

3 Hattori M, Iwase N, Furuya N, Tanaka Y, Tsukazaki T, Ishitani R, Maguire ME, *et al.* Mg(2+)-dependent gating of bacterial MgtE channel underlies Mg(2+) homeostasis. The EMBO journal 2009, 28: 3602-3612

4 Dann CE, 3rd, Wakeman CA, Sieling CL, Baker SC, Irnov I, Winkler WC. Structure and mechanism of a metal-sensing regulatory RNA. Cell 2007, 130: 878-892

5 Lee DD, Galera-Laporta L, Bialecka-Fornal M, Moon EC, Shen Z, Briggs SP,

Garcia-Ojalvo J, *et al.* Magnesium Flux Modulates Ribosomes to Increase Bacterial Survival. Cell 2019, 177: 352-360.e313

6 Anderson GG, Yahr TL, Lovewell RR, O'Toole GA. The Pseudomonas aeruginosa magnesium transporter MgtE inhibits transcription of the type III secretion system. Infect Immun 2010, 78: 1239-1249

7 Chakravarty S, Melton CN, Bailin A, Yahr TL, Anderson GG. Pseudomonas aeruginosa Magnesium Transporter MgtE Inhibits Type III Secretion System Gene Expression by Stimulating rsmYZ Transcription. J Bacteriol 2017, 199

8 Goytain A, Quamme GA. Functional characterization of ACDP2 (ancient conserved domain protein), a divalent metal transporter. Physiological genomics 2005, 22: 382-389

9 Sahni J, Scharenberg AM. The SLC41 family of MgtE-like magnesium transporters. Molecular aspects of medicine 2013, 34: 620-628

10 Schweigel-Röntgen M, Kolisek M. SLC41 transporters--molecular identification and functional role. Current topics in membranes 2014, 73: 383-410

11 Kolisek M, Sponder G, Mastrototaro L, Smorodchenko A, Launay P, Vormann J, Schweigel-Rontgen M. Substitution p.A350V in Na(+)/Mg(2)(+) exchanger SLC41A1, potentially associated with Parkinson's disease, is a gain-of-function mutation. PLoS One 2013, 8: e71096

12 Hattori M, Tanaka Y, Fukai S, Ishitani R, Nureki O. Crystal structure of the MgtE Mg2+ transporter. Nature 2007, 448: 1072-1075

13 Maruyama T, Imai S, Kusakizako T, Hattori M, Ishitani R, Nureki O, Ito K, *et al.* Functional roles of Mg(2+) binding sites in ion-dependent gating of a Mg(2+) channel, MgtE, revealed by solution NMR. Elife 2018, 7

14 Jin F, Huang Y, Hattori M. Recent Advances in the Structural Biology of Mg(2+) Channels and Transporters. J Mol Biol 2022, 434: 167729

15 Hattori M, Tanaka Y, Fukai S, Ishitani R, Nureki O. Crystallization and preliminary X-ray diffraction analysis of the full-length Mg2+ transporter MgtE. Acta crystallographica Section F, Structural biology and crystallization communications 2007, 63: 682-684

16 Ishitani R, Sugita Y, Dohmae N, Furuya N, Hattori M, Nureki O. Mg2+-sensing mechanism of Mg2+ transporter MgtE probed by molecular dynamics study. Proceedings

of the National Academy of Sciences of the United States of America 2008, 105: 15393-15398

17 Tomita A, Zhang M, Jin F, Zhuang W, Takeda H, Maruyama T, Osawa M, *et al.* ATP-dependent modulation of MgtE in Mg(2+) homeostasis. Nature communications 2017, 8: 148

18 Takeda H, Hattori M, Nishizawa T, Yamashita K, Shah ST, Caffrey M, Maturana AD, *et al.* Structural basis for ion selectivity revealed by high-resolution crystal structure of Mg2+ channel MgtE. Nature communications 2014, 5: 5374

19 Teng X, Sheng D, Wang J, Yu Y, Hattori M. Ion selectivity mechanism of the MgtE channel for Mg(2+) over Ca(2). iScience 2022, 25: 105565

20 Jin F, Sun M, Fujii T, Yamada Y, Wang J, Maturana AD, Wada M, *et al.* The structure of MgtE in the absence of magnesium provides new insights into channel gating. PLoS biology 2021, 19: e3001231

21 Brahma R, Raghuraman H. Characterization of a novel MgtE homolog and its structural dynamics in membrane mimetics. Biophys J 2024, 123: 1968-1983

22 Wang M, Zhao Y, Hayashi Y, Ito K, Hattori M. Novel Mg (2+) binding sites in the cytoplasmic domain of the MgtE Mg (2+) channels revealed by X-ray crystal structures. Acta Biochim Biophys Sin (Shanghai) 2023, 55: 683-690

23 Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes - a review. Nat Prod Rep 2016, 33: 988-1005

24 Medema MH, de Rond T, Moore BS. Mining genomes to illuminate the specialized chemistry of life. Nat Rev Genet 2021, 22: 553-571

25 Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. Curr Opin Struct Biol 2008, 18: 581-586

26 Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF, Wicky BIM, *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. Science 2022, 378: 49-56

27 Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, Ahern W, *et al.* De novo design of protein structure and function with RFdiffusion. Nature 2023, 620: 1089-1100

28 Kortemme T. De novo protein design-From new structures to programmable functions. Cell 2024, 187: 526-544

29 Zhu J, Lu P. Computational design of transmembrane proteins. Curr Opin Struct Biol 2022, 74: 102381

30 Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 2018, 36: 996-1004

31 Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 2015, 25: 1043-1055

32 Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, *et al.* Pfam: the protein families database. Nucleic Acids Res 2014, 42: D222-230

33 Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol 2011, 7: e1002195

34 Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, *et al.* InterProScan 5: genome-scale protein function classification. Bioinformatics 2014, 30: 1236-1240

35 Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://wwwR-projectorg/ 2023,

36 Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, *et al.* Welcome to the Tidyverse. Journal of Open Source Software 2019, 4: 1686

37 Lee S, Cook D, Lawrence M. plyranges: a grammar of genomic data transformation. Genome Biol 2019, 20: 4

38 Pagès H AP, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings. R package version 2.74.1. https://bioconductororg/packages/Biostrings 2024,

39 T P. patchwork: The Composer of Plots_. R package version 1.2.0. https://CRANR-projectorg/package=patchwork 2024,

40 Jin F, Shen C, Wang Y, Wang M, Sun M, Hattori M. Fluorescence-detection size-exclusion chromatography utilizing nanobody technology for expression screening of membrane proteins. Commun Biol 2021, 4: 366

41 Gotzke H, Kilisch M, Martinez-Carranza M, Sograte-Idrissi S, Rajavel A, Schlichthaerle T, Engels N, *et al.* The ALFA-tag is a highly versatile tool for nanobody-

based bioscience applications. Nat Commun 2019, 10: 4403
42 Kawate T, Gouaux E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure 2006, 14: 673-681
43 Hattori M, Hibbs RE, Gouaux E. A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. Structure 2012, 20: 1293-1299
44 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K*, et al.* Highly accurate protein structure prediction with AlphaFold. Nature 2021, 596: 583-589
45 Liu K, Zhou H, Xu Q, Kong H-T, Zhang K-H, Wang W-W, Li M-J*, et al.* BL02U1: the relocated macromolecular crystallography beamline at the Shanghai Synchrotron Radiation Facility. Nuclear Science and Techniques 2023, 34: 193
46 Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr 2010, 66: 125-132
47 Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B*, et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol 2019, 75: 861-877
48 Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 2010, 66: 486-501
49 Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B*, et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallographica Section D 2019, 75: 861-877
50 Wolny M, Batchelor M, Bartlett GJ, Baker EG, Kurzawa M, Knight PJ, Dougan L*, et al.* Characterization of long and stable de novo single alpha-helix domains provides novel insight into their stability. Sci Rep 2017, 7: 44341
51 Huang Y, Zhang Z, Hattori M. Recent Advances in Expression Screening and Sample Evaluation for Structural Studies of Membrane Proteins. J Mol Biol 2024, 436: 168809
52 Errasti-Murugarren E, Bartoccioni P, Palacin M. Membrane Protein Stabilization Strategies for Structural and Functional Studies. Membranes (Basel) 2021, 11
53 Cui W, Niu Y, Chen L. The Protein Fusion Strategy Facilitates the Structure

Determination of Small Membrane Proteins by Cryo-EM. Biochemistry 2023, 62: 196-200

54 Yang F, Moss LG, Phillips GN, Jr. The molecular structure of green fluorescent protein. Nat Biotechnol 1996, 14: 1246-1251

55 van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Soding J*, et al.* Fast and accurate protein structure search with Foldseek. Nat Biotechnol 2024, 42: 243-246

56 Huang PS, Boyken SE, Baker D. The coming of age of de novo protein design. Nature 2016, 537: 320-327

57 Gupte TM, Ritt M, Sivaramakrishnan S. ER/K-link-Leveraging a native protein linker to probe dynamic cellular interactions. Methods Enzymol 2021, 647: 173-208

58 Pidathala S, Liao S, Dai Y, Li X, Long C, Chang CL, Zhang Z*, et al.* Mechanisms of neurotransmitter transport and drug inhibition in human VMAT2. Nature 2023, 623: 1086-1092

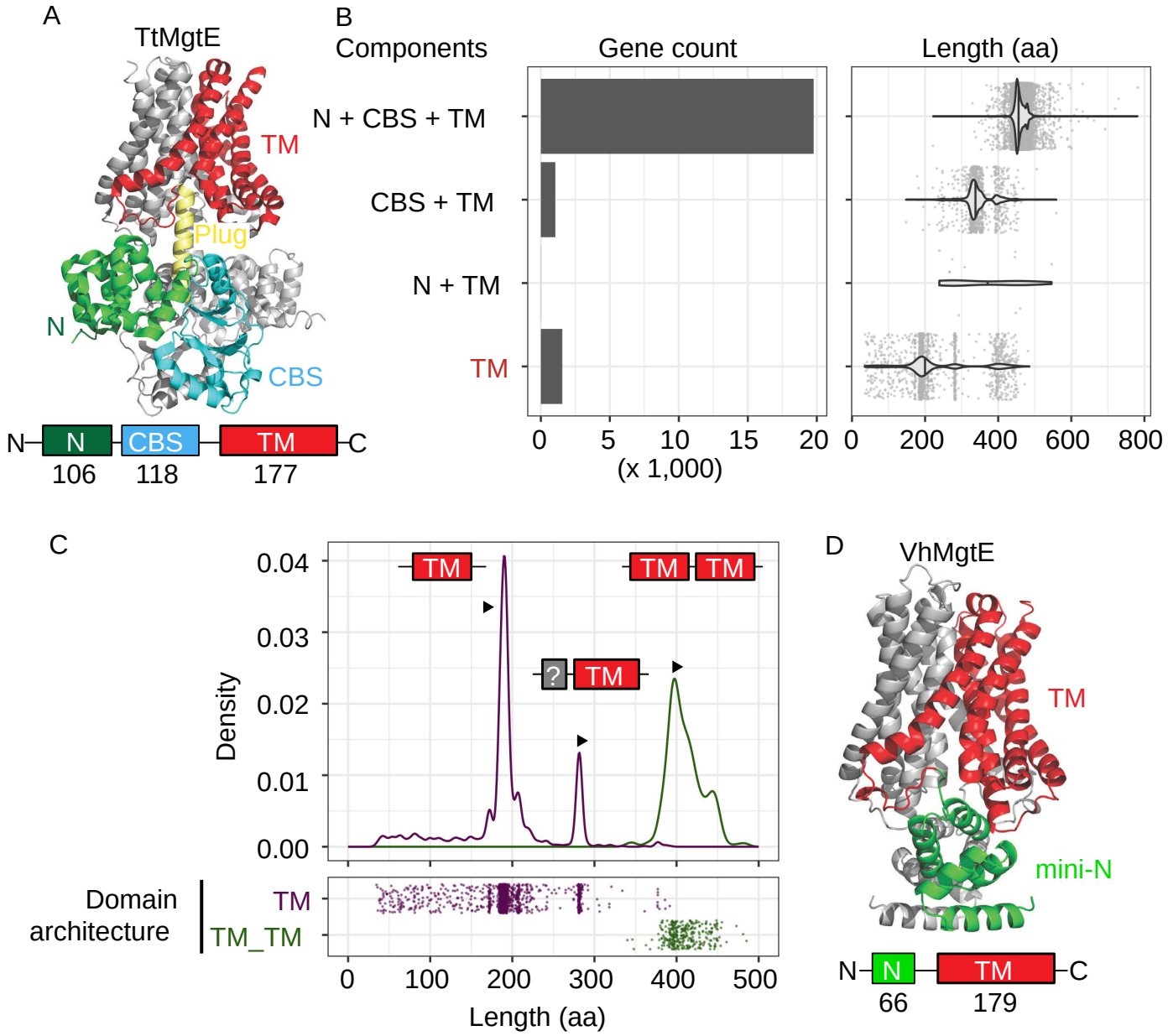
A
TtMgtE
TM
Plug
N
CBS
N
N
CBS
TM
C
106
118
177
B
Components
Gene count
Length (aa)
N + CBS + TM
CBS + TM
N + TM
TM
0
5
10
15
20
(x 1,000)
0
200
400
600
800
C
Density
0.04
0.03
0.02
0.01
0.00
TM
TM
TM
?
TM
Domain architecture
TM
TM_TM
0
100
200
300
400
500
Length (aa)
D
VhMgtE
TM
mini-N
N
N
TM
C
66
179

Figure 1

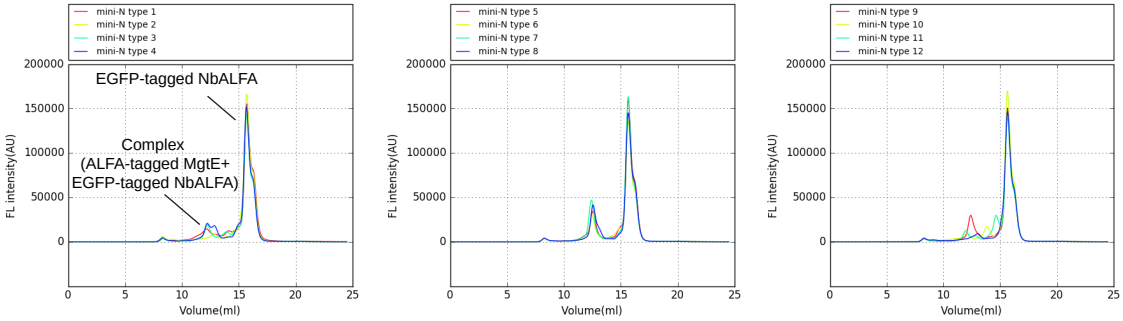
mini-N type 1
mini-N type 2
mini-N type 3
mini-N type 4
EGFP-tagged NbALFA
Complex
(ALFA-tagged MgtE+
EGFP-tagged NbALFA)
FL intensity(AU)
Volume(ml)
mini-N type 5
mini-N type 6
mini-N type 7
mini-N type 8
mini-N type 9
mini-N type 10
mini-N type 11
mini-N type 12

Figure 2

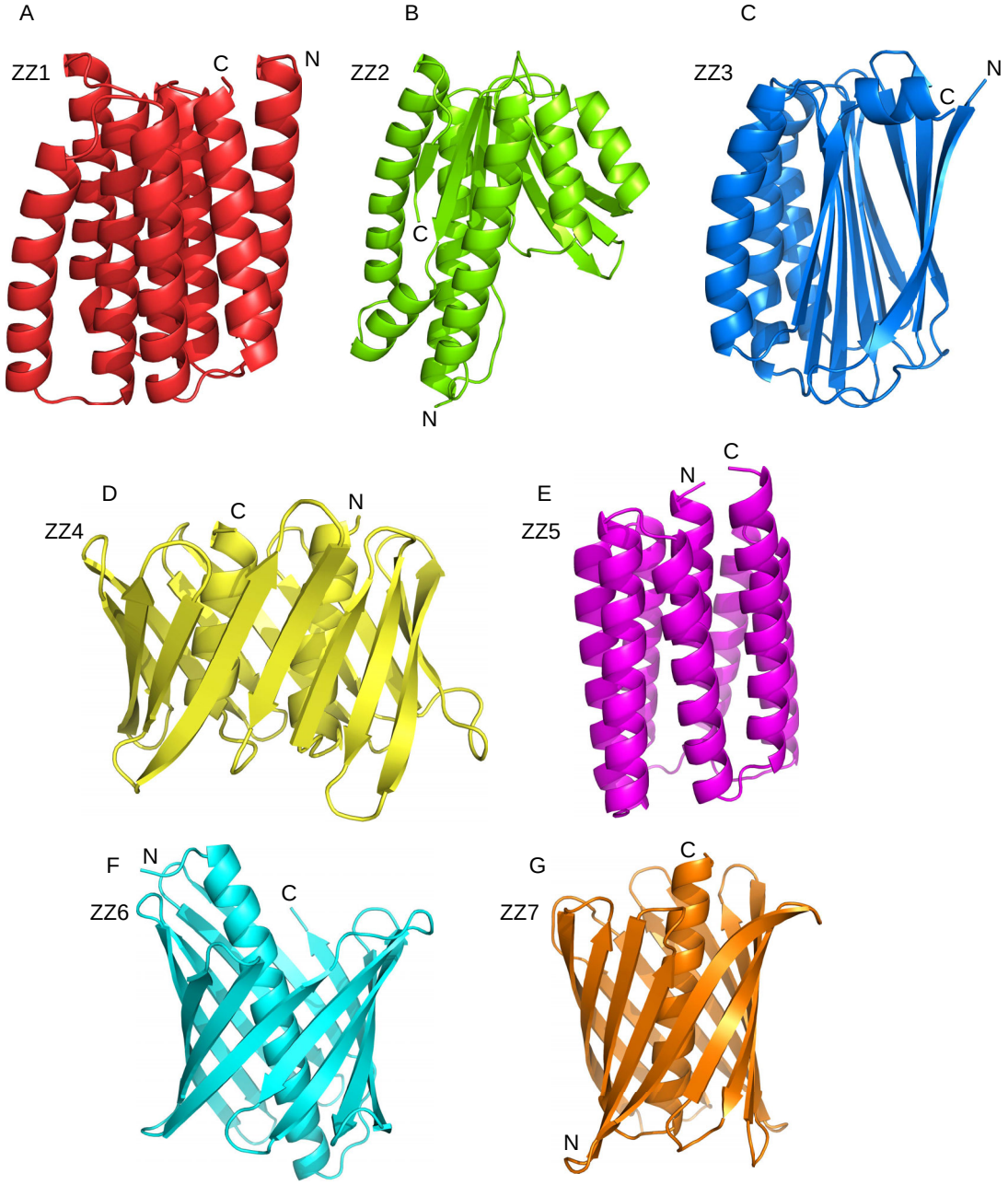

Figure 3

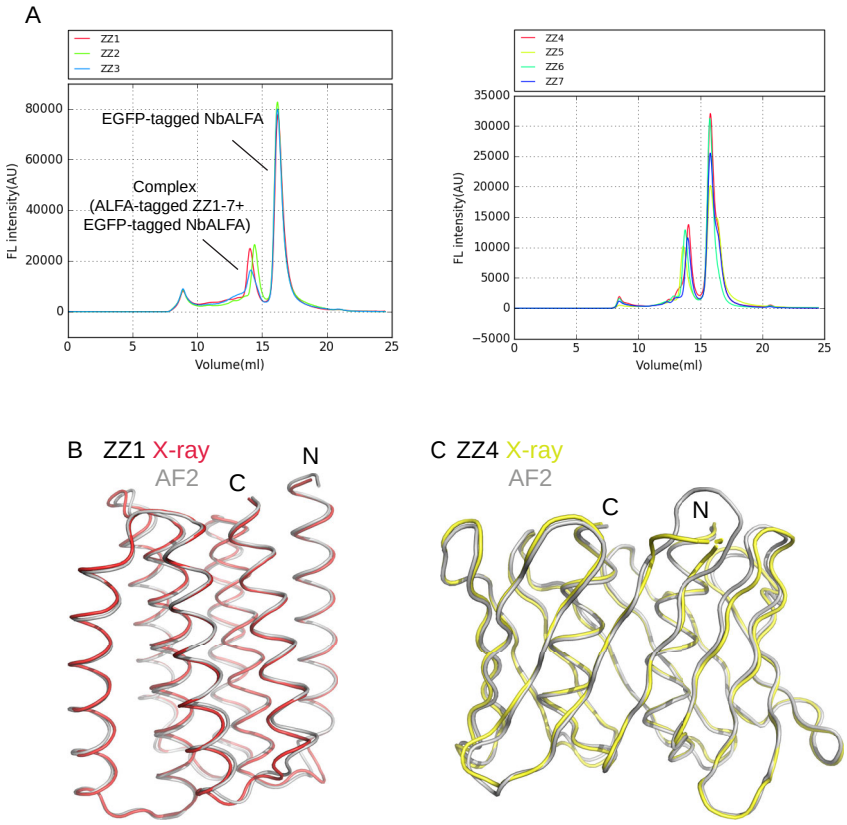
A
ZZ1
ZZ2
ZZ3
EGFP-tagged NbALFA
Complex
(ALFA-tagged ZZ1-7+
EGFP-tagged NbALFA)
FL intensity(AU)
Volume(ml)
ZZ4
ZZ5
ZZ6
ZZ7
FL intensity(AU)
Volume(ml)
B ZZ1 X-ray
AF2
C
N
C ZZ4 X-ray
AF2
C
N

Figure 4

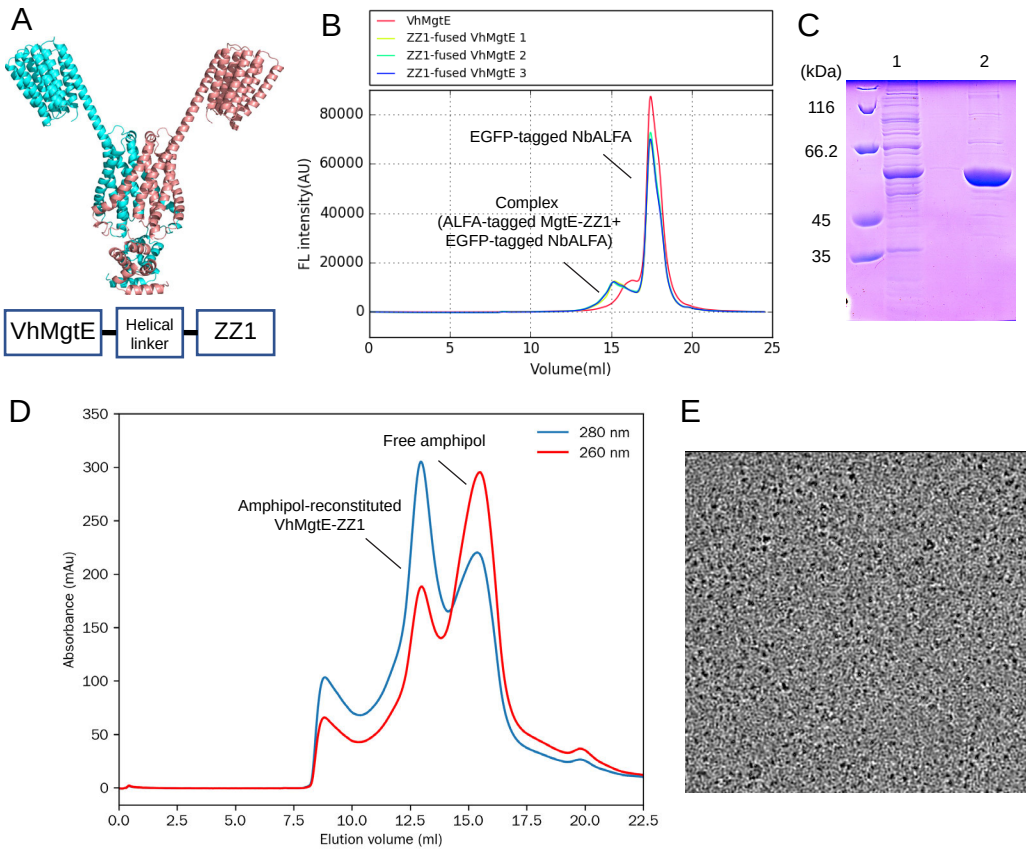
A
VhMgtE
Helical linker
ZZ1
B
VhMgtE
ZZ1-fused VhMgtE 1
ZZ1-fused VhMgtE 2
ZZ1-fused VhMgtE 3
EGFP-tagged NbALFA
Complex
(ALFA-tagged MgtE-ZZ1+
EGFP-tagged NbALFA)
FL intensity(AU)
Volume(ml)
C
(kDa)
1
2
116
66.2
45
35
D
Free amphipol
Amphipol-reconstituted
VhMgtE-ZZ1
280 nm
260 nm
Absorbance (mAu)
Elution volume (ml)
E

Figure 5

Table 1. The number of MgtE homologs

Domain components	Archaea	Bacteria	Total
N + CBS + TM	82	19,690	19,772
CBS + TM	71	979	1,050
N + TM	0	11	11
TM	964	591	1,555