

Short Communication

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

Zhixuan Zhao¹, Kimiho Omae^{2,3,*}, Wataru Iwasaki^{2,4}, Ziyi Zhang¹, Fazhi Pan¹, Eun-Jin Lee⁵, Koichi Ito⁶, and Motoyuki Hattori^{1,*}

¹State 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 200438, China, ²Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 2770882, Japan, ³RIKEN Pioneer Research Institute, RIKEN, Saitama 3510198, Japan, ⁴Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 1130033, Japan, ⁵Department of Life Sciences, Korea University, Seoul 02841, South Korea, and ⁶Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa City, Chiba 2778562, Japan

*Correspondence address. E-mail: omae@edu.k.u-tokyo.ac.jp (K.O.); hattori@fudan.edu.cn (M.H.)

Received 10 May 2025 Accepted 29 July 2025 Published 13 January 2026

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 focused predominantly on *Thermus thermophilus* MgtE, leaving the structural and mechanistic diversity of MgtE family proteins largely unexplored. In this study, via a genome mining approach, we identify 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 cannot be stably purified. To address this issue, we design a series of *de novo* proteins and determine their crystal structures. A selected *de novo* protein is fused to a mini-N-type MgtE, enabling successful purification and preliminary cryo-EM imaging. Our findings demonstrate that *de novo*-designed protein fusions serve as powerful tools for stabilizing and purifying otherwise unstable membrane proteins, opening new avenues for 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 the transcription of the type III secretion system, a critical virulence factor, by promoting the expression of the translation-inhibiting sRNAs RsmY and RsmZ [6,7]. In humans, MgtE homologs 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 potential therapeutic targets.

To date, extensive structural and mechanistic studies have been

conducted on MgtE [12–14]. These studies revealed that MgtE functions as a “ Mg^{2+} -gated Mg^{2+} channel” where the binding of Mg^{2+} 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], whereas the cryo-EM structure of MgtE in the absence of the Mg^{2+} domain has provided insights into the gating mechanism [20].

However, these studies have focused predominantly on *Thermus thermophilus* MgtE (TtMgtE), and structural and biochemical data for MgtE homologs from other species are extremely limited [21–23]. 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 via genome mining approaches [24,25].

One major challenge in studying membrane proteins such as MgtE lies in their generally low expression levels and poor stability in detergents [26], which significantly hinders structural and functional analyses. Recently, AI-based *de novo* protein design technologies, such as ProteinMPNN [27] and RF diffusion [28], have rapidly advanced [29] and are expected to facilitate research on membrane proteins [30]. *De novo*-designed proteins typically exhibit high stability, and fusing them to target membrane proteins may increase their 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 designed *de novo* proteins 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 via a *de novo*-designed protein fusion approach.

Materials and Methods

Bioinformatics analysis of MgtE proteins

The prokaryotic proteomes and taxonomic information were downloaded from the Genome Taxonomy Database Release 06-RS202 [31]. Only the high-quality genomes (completeness $\geq$ 95% and contamination $\leq$ 5%) identified via CheckM [32] were used for the analysis. The genome quality of phylum Patescibacteria was estimated by using a 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 [33,34]. The domain architecture of MgtE was determined via an InterProScan search [35]. The InterPro IDs IPR006667, IPR046342, and IPR006668 were used as proxies for the TM, CBS and N domains of MgtE, respectively.

Data analysis and visualization were performed via R 4.3.2 [36] with the following packages: tidyverse [37], plyranges [38], Biostrings [39] and patchwork [40]. The code for the analysis is provided in the GitHub repository https://github.com/0mae/mgte_short.

Screening of MgtE expression via FSEC-Nb

MgtE genes and their mutants were synthesized by Azenta (Suzhou, China) and subcloned and inserted into a pETNb-nALFA vector [41] containing an N-terminal 8× His tag, an ALFA tag [42], and an HRV 3C protease cleavage site. *Escherichia coli* Rosetta (DE3) cells were transformed with the expression vectors and cultured in 8 mL of LB medium containing 50 µg/mL kanamycin and 30 µg/mL chloramphenicol at 37°C with the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) when the OD₆₀₀ reached ~0.6. *E. coli* cells were then cultured at 18°C for 16 h and harvested via 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) n-dodecyl-β-D-maltoside (DDM) and

0.5 mM PMSF for 1 h, followed by ultracentrifugation (200,000 g, 20 min). One microgram of EGFP-tagged NbALFA was added to half the 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, Waltham, USA) equilibrated with buffer TBS containing 0.05% (w:v) DDM for FSEC analysis [43,44]. The fluorescence was monitored via an RF-20Axs fluorescence detector (Shimadzu, Kyoto, Japan) at an excitation wavelength of 480 nm and a mission wavelength of 512 nm.

Expression and purification of MgtE

Large-scale expression of the MgtE protein followed the same method as that used for small-scale expression screening, and 6 L of LB medium was used for expression. After harvesting by centrifugation, the *E. coli* cell pellets were resuspended and disrupted via a microfluidizer in TBS buffer. After centrifugation to remove cell debris, the supernatant was ultracentrifuged at 200,000 g for 1 h to isolate the membrane fraction. The membrane was solubilized in S buffer [50 mM Tris, pH 8.0, 150 mM NaCl, 1 mM PMSF, and 1 mM β-mercaptoethanol (β-ME)], and one of the following detergents: 2% DDM, 2% DDM + 0.2% cholesteryl hemisuccinate (CHS), or 1% lauryl maltose neopentyl glycol (LMNG)] for 1 h. After ultracentrifugation at 200,000 g for 1 h, the supernatant was loaded onto Ni-NTA agarose beads (QIAGEN, Hilden, 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, 1 mM β-ME, and one of the following detergents: 0.03% DDM, 0.03% DDM with 0.003% CHS, or 0.01% LMNG), and the proteins were eluted via buffer B (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 250 mM imidazole, 1 mM β-ME, and one of the following detergents: 0.03% DDM, 0.03% DDM with 0.003% CHS, or 0.01% LMNG) and concentrated via Amicon Ultra 50K and 100K centrifuge filters (Merck Millipore, Billerica, USA) for the wild-type construct and ZZ1-fused construct, respectively. For the wild-type construct, the concentrated elution was further analyzed by FSEC on the Superdex 200 Increase 10/300 GL column equilibrated with buffer TBS containing 0.03% DDM, 0.03% DDM + 0.003% CHS, or 0.01% LMNG, detected via intrinsic protein fluorescence (excitation: 280 nm, emission: 325 nm). For the ZZ1-fused construct, the eluted fractions were dialyzed against buffer C (20 mM HEPES, pH 7.0, 150 mM NaCl, 1 mM β-ME, and one of the following detergents: 0.03% DDM, 0.03% DDM with 0.003% CHS, or 0.01% LMNG) for 16 h at 4°C. The protein was injected onto the Superdex 200 Increase 10/300 GL size-exclusion chromatography column (Cytiva) and eluted with buffer D [20 mM HEPES, pH 7.0, 150 mM NaCl, 0.5 mM Tris (2-carboxyethyl) phosphine (TCEP), and one of the following detergents: 0.03% DDM, 0.03% DDM with 0.003% CHS, or 0.01% LMNG]. The main peak fractions in SEC were further analyzed by FSEC on the Superdex 200 Increase 10/300 GL column equilibrated with buffer TBS containing 0.03% DDM, 0.03% DDM + 0.003% CHS, or 0.01% LMNG, detected via intrinsic protein fluorescence. The main peak fractions in DDM were used for the subsequent amphipol reconstitution experiment.

De novo protein design

All *de novo* proteins in this study were designed via the ColabDesign framework (<https://github.com/sokrypton/ColabDesign>). Briefly, the backbone structures of the ZZ1-ZZ3 proteins were

generated via RFdiffusion [28] with a specified length of 200 amino acid residues and an unconditional folding mode. The corresponding sequences were then generated via ProteinMPNN [27], and the designed protein sequences were subsequently validated via AlphaFold [45]. The ZZ4–ZZ7 proteins were designed similarly but in the conditional folding mode.

Expression screening of *de novo*-designed proteins

The DNA sequence encoding each *de novo*-designed protein was synthesized by Azenta and subcloned and inserted into a pETNb-cALFA vector [41] 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 supplemented with 50 µg/mL ampicillin and 30 µg/mL chloramphenicol at 37°C. When the OD₆₀₀ 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 h and harvested via 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 the Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated with TBS buffer for the FSEC assay.

Expression and purification of *de novo*-designed proteins

Large-scale expression of *de novo*-designed proteins was performed via the same method as that used for small-scale expression screening, with 6 L of LB medium used 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). The proteins were then eluted with 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 the HRV 3C protease, and 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 the Superdex 75 Increase 10/300 GL size-exclusion chromatography column (Cytiva) 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), whereas 1 µL of ZZ4 protein was mixed with 1 µL of crystallization reservoir buffer I (0.2 M MgCl₂, 0.1 M KCl, 0.025 M sodium citrate, pH 4.0, 24.5% PEG 400) for crystallization via 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 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, Shanghai, China) [46]. Data processing was carried out via XDS [47]. The initial phase was determined by molecular replacement via Phaser [48], which uses AlphaFold [45] models as templates. Manual model building and iterative refinement were performed with Coot [49] and PHENIX [50], respectively. Structural figures were prepared via PyMOL (<https://pymol.org>).

Construction design of ZZ1-fused VhMgtE

ZZ1-fused VhMgtE harbors a 3- or 4-residue truncation at its C-terminus and is fused via one or two EK2R1 (AEEERKRK) helices [51]. The construct was validated by AlphaFold [45] before the wet experiments.

Cryo-EM image of ZZ1-fused MgtE

The purified ZZ1-fused MgtE protein was mixed with Amphipol A8-35 (Anatrace, Ohio, USA) at a mass ratio of 1:10 and incubated for 16 h. Bio-Beads SM-2 (Bio-Rad, Hercules, USA) were then added to the mixture and incubated at 4°C for 4 h to remove detergent. 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, 0.5 mM TCEP). The peak fractions were concentrated to 2.0 mg/mL. Freshly prepared samples were incubated with 2 mM MgCl₂ on ice for 30 min before vitrification. Cryo-grid preparation was performed at 4°C and 100% humidity using a Vitrobot Mark IV (Thermo Fisher Scientific, Waltham, USA). A 2.5 µL sample was applied to each freshly glow-discharged Au 300-mesh grid (R1.2/1.3) (Quantifoil, Thüringen, Germany). The grids were then plunge-frozen in liquid ethane. Cryo-grids were screened via a 200 kV Glacios 2 microscope equipped with a Falcon 4i camera (Thermo Fisher Scientific).

Complementation of Mg²⁺ auxotrophy in *E. coli*

The *E. coli* strain 3KO (BW25113 Δ mgtA Δ corA Δ yhiD DE3) [3], which requires a high concentration of Mg²⁺ for growth, was transformed with a control vector or plasmids designed to express VhMgtE or TtMgtE on LB agar plates containing 50 µg/mL kanamycin and 100 mM MgSO₄. The resulting transformants were spotted and streaked on LB agar plates containing 50 µg/mL kanamycin under one of the following conditions: 100 mM MgSO₄; no supplementation; or 7.5 µM IPTG. The cells were grown overnight at 37°C. Notably, in this strain, a basal level of T7 RNA polymerase is constitutively produced from the chromosomal lacUV5 promoter even in the absence of IPTG. This low but sufficient amount of polymerase initiates transcription from the T7 promoter, resulting in enough MgtE expression to sustain growth, which is particularly noticeable with TtMgtE.

Results

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

We conducted an extensive search against prokaryotic genome datasets 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., TtMgtE) contains all three domains: the N domain, the CBS domain, and the TM domain (N-CBS-TM; Figure 1A). While the binding of Mg²⁺ to the cytoplasmic domains of MgtE inactivates the channel, the Mg²⁺-binding sites are located at two interfaces: between the N domain

and the CBS domain and between the CBS domain and the TM domain [3]. The binding of Mg²⁺ to these sites controls the orientation of the plug helix that links the CBS and TM domains, thereby modulating channel activity. Consequently, the CBS domain, which is directly connected to the plug helix, serves as the central module for Mg²⁺-dependent gating of MgtE, whereas the N domain provides additional structural support. In contrast, the TM domain forms an ion-conducting pore and provides a selectivity filter for Mg²⁺.

However, in our genome mining, some MgtE proteins lack the N domain (CBS-TM), whereas others lack both the N and CBS domains, retaining only the TM domain (TM; Figure 1B). These proteins are not negligible in number (~2600), and they are conserved in both bacteria and archaea (Supplementary Table S1). Notably, some proteins lacking both the N and CBS domains possess tandem-repeat TM (TM-TM) domains, 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 (Figures 1C and Supplementary Figure S1). We designated this class of MgtE the “mini-N” type (Figure 1D). The discovery of mini-N-type MgtE was surprising, as in canonical MgtE, the CBS domain is known to be essential for channel regulation, whereas the N domain itself does not appear to play a direct role in gating regulation [3]. Accordingly, mini-N-type MgtE represents a highly intriguing target for further investigation.

Notably, the MgtE subclasses identified in our genome mining, although lacking part or all of the cytoplasmic domains, still retain the conserved TM domain responsible for ion conduction (Figure 1B). Accordingly, we consider these diverse subclasses to be members of the MgtE Mg²⁺ channel family, similar to canonical MgtE.

Expression screening of the mini-N-type MgtE proteins

To characterize these various types of MgtE proteins structurally

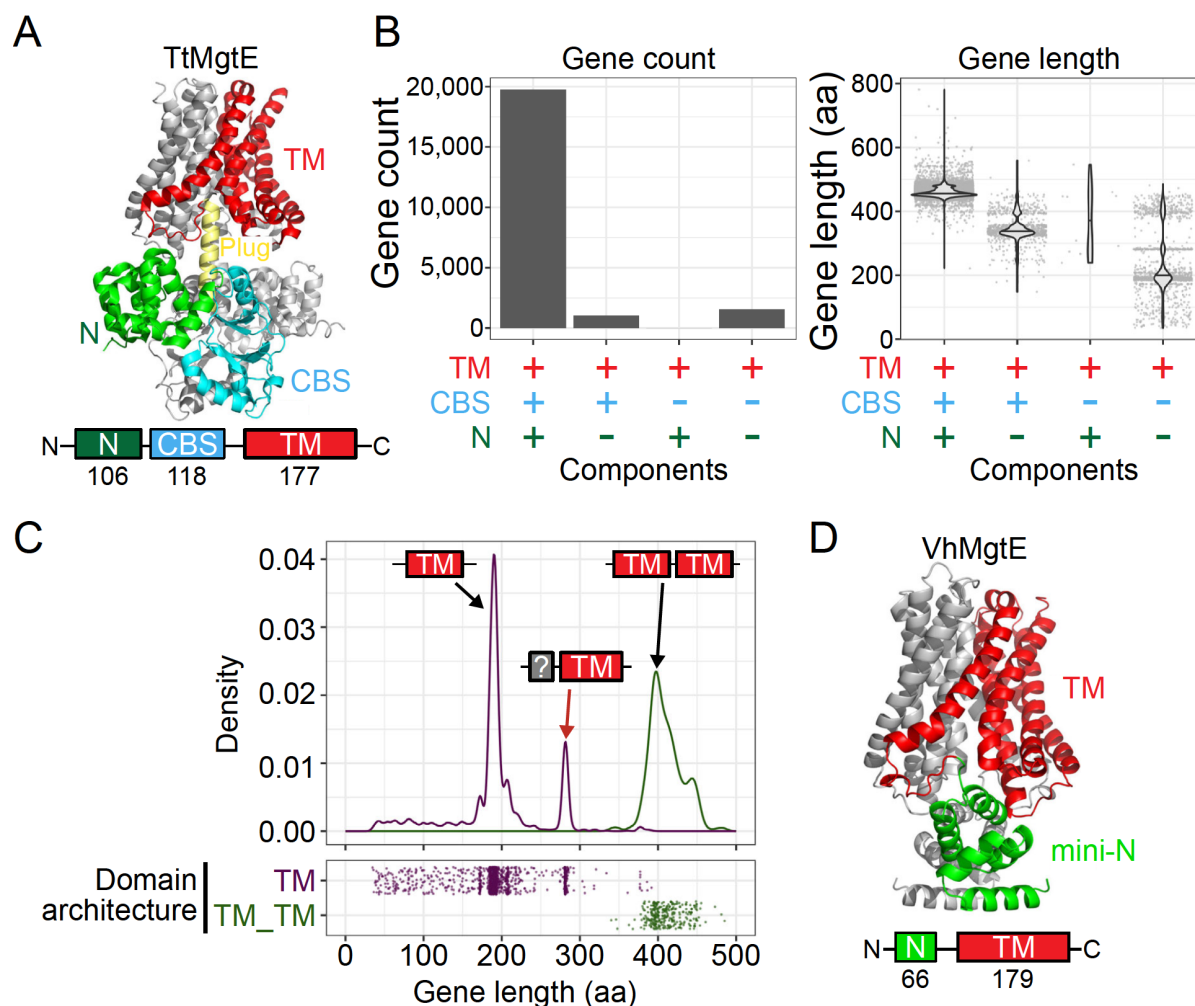


Figure 1. Bioinformatics analysis of MgtE proteins revealed various types of MgtE proteins (A) Canonical MgtE structure containing “N”, “CBS” and transmembrane (“TM”) domains. The structure of *Thermus thermophilus* MgtE (PDB ID: 2ZY9) is shown. (B) Bar plot (left) and violin plot (right) showing the number and length of MgtE proteins identified via bioinformatics analysis. Proteins are categorized on the basis of the presence (+) or absence (-) of the three domain components. (C) Length distribution of the MgtE gene, which lacks both the N and CBS domains. MgtE proteins are classified on the basis of 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 is shown.

and functionally, we initially selected mini-N-type MgtE proteins for expression screening before performing 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 via fluorescence-detection size-exclusion chromatography (FSEC) [43,51], specifically through nanobody (Nb)-based FSEC (FSEC-Nb) [41] (Figure 2). In the FSEC-Nb method, targeted membrane proteins are fused to an ALFA peptide tag [42], and solubilized whole-cell extracts are 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 revealed that multiple homologs (Nos. 5–9 in Figure 2) presented sharp, high-intensity peaks, indicating high levels of expression and monodispersity; thus, these homologs might be suitable for further large-scale expression and purification. While the ALFA tag has picomolar affinity for its nanobody, making it ideal for FSEC-Nb screening of membrane proteins [41], the His tag allows cost-effective and rapid purification by IMAC, and our expression plasmid carries both a His tag and an ALFA tag in tandem. Accordingly, we employed the ALFA tag during small-scale FSEC expression screening and the His tag during large-scale expression and purification. However, in further expression and purification trials, these proteins tended to aggregate or precipitate during purification. In particular, because its peak in the FSEC-Nb analysis was favorable (Figure 2C), we purified the homolog from *Vibrio harveyi* (No9, VhMgtE), a well-characterized marine pathogenic bacterium [52], under three detergent conditions (DDM, DDM-CHS, and LMNG; Figures 2D,E). However, the yield was very low in all the cases (DDM, 11.1 µg; DDM-CHS, 50.0 µg; and LMNG, 27.8 µg per liter of *E. coli* culture).

Notably, while we encountered difficulty in purification, we verified the functionality of VhMgtE in a growth-complementation assay (Figure 3). An *E. coli* triple knockout devoid of Mg²⁺ transporters ($\Delta mgtA \Delta corA \Delta yhiD$) normally grows only when LB medium is supplemented with excess Mg²⁺ [3]. When this strain was transformed with VhMgtE, it regained the ability to proliferate in Mg²⁺-deficient LB, matching the rescue observed with the positive control construct TtMgtE (Figure 3). While VhMgtE lacks the CBS domain, canonical MgtE dimerization is not mediated solely by the CBS domain. As shown in Figure 1A, the TM region, which consists of five helices per subunit, forms a tight dimer interface that is essential for creating ion-conducting pores [3]. This interface is also conserved in the VhMgtE structure, as predicted by AlphaFold2 (Figure 1D). Given that VhMgtE also showed Mg²⁺ uptake activity (Figure 3), it is most plausible that VhMgtE also still assembles as a homodimer. We believe that a detailed comparison of the functional or regulatory differences between mini-N-type MgtE and other MgtE homologs is beyond the scope of the present work and should be addressed in future studies.

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 4 and Supplementary Table S2) 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 increase 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 [28,29], potentially making them particularly suitable as fusion partners for membrane proteins. Using RFdiffusion [28], we designed seven *de novo* proteins: three (ZZ1–ZZ3) in unconditional folding mode with a fixed length of 200 amino acid residues and four (ZZ4–ZZ7) in 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 (Supplementary Table S2). In unconditional folding mode, we generated all- α (ZZ1), α/β (ZZ2), and $\alpha+\beta$ (ZZ3) structures (Figure 4A–C). In the conditional folding mode, in addition to 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* via the FSEC-Nb method [41] prior to structural analysis (Figure 5A). Although FSEC techniques are primarily utilized for screening membrane protein expression, they can also be effectively applied to soluble proteins [51]. All seven designed proteins presented high and sharp peaks corresponding to the complexes of the ALFA-tagged proteins and NbALFA fused to GFP in FSEC (Figure 5A) 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 Å and 2.03 Å, respectively, and their structures were successfully determined by molecular replacement using AlphaFold2 models as templates (Figure 5B and Supplementary Table S3). Structural superposition revealed RMSD values of 0.65 Å and 0.83 Å for the Ca atoms of ZZ1 and ZZ4, respectively, between the AlphaFold-predicted models and crystal structures, demonstrating the high accuracy of our designs (Figure 5B).

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

To fuse our *de novo*-designed proteins to the mini-N-type MgtE, we selected ZZ1 on the basis of its α -helix bundle architecture, which is highly compatible with fusion to membrane proteins in various configurations, as well as its demonstrated stability, as evidenced by successful X-ray crystallography (Figure 5). For the mini-N-type MgtE, we continued to choose VhMgtE.

To fuse ZZ1 to VhMgtE, we employed an EK2R1 (AEEKRRK) helical linker (Figure 6A), which is known for its high stability [57] and utility in membrane protein structural determination via 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

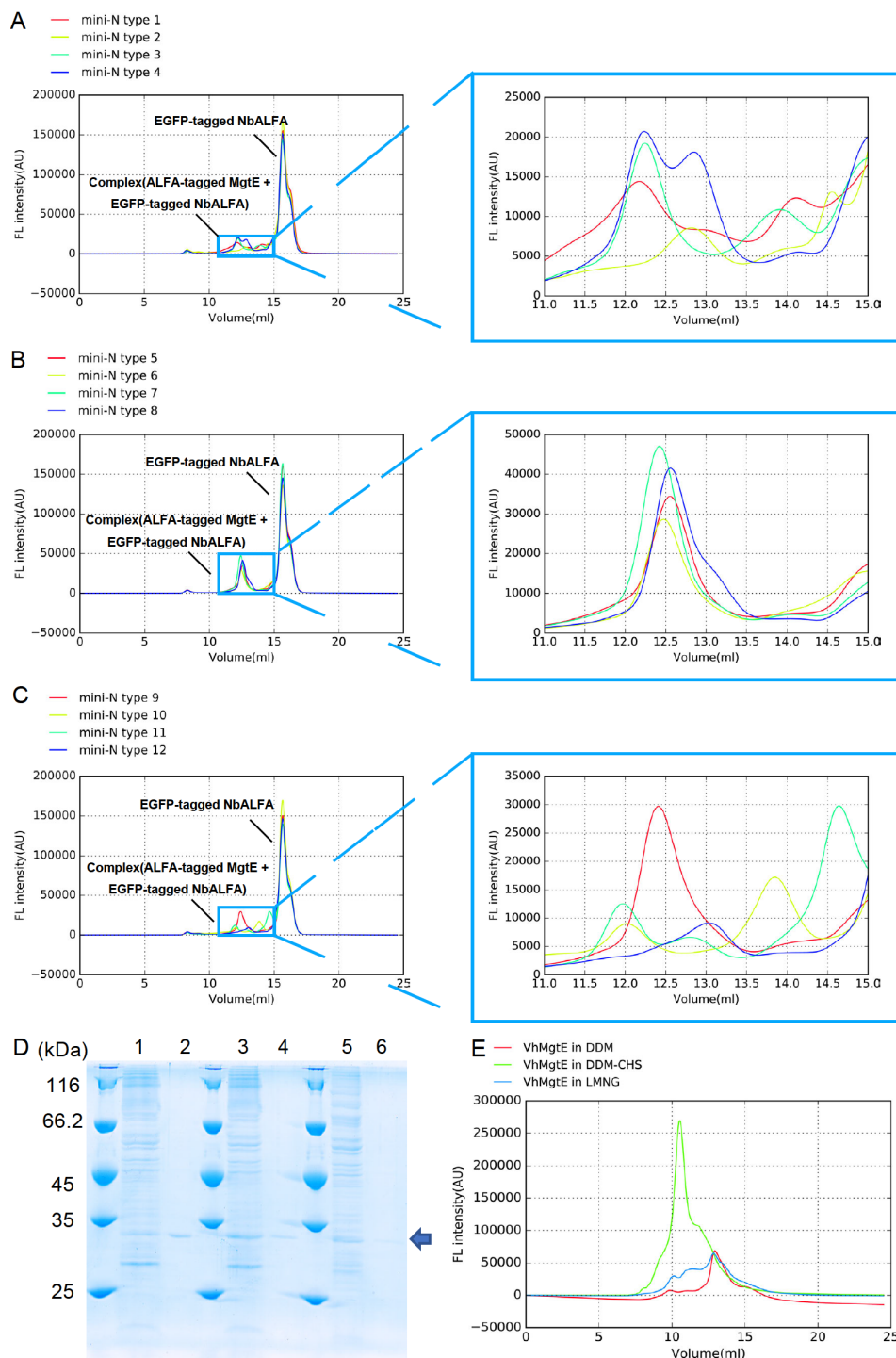


Figure 2. Expression screening of mini-N-type MgtE proteins via FSEC-Nb (A-C) FSEC traces of unpurified ALFA-tagged mini-N-type MgtE proteins mixed with EGFP-tagged NbALFA on a Superdex 200 Increase 10/300 GL column, detected via EGFP fluorescence. The expression of mini-N-type MgtE proteins from the 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 frigidophilum* (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). (D) SDS-PAGE for the purification of VhMgtE (No. 9) under different conditions. Lane 1: 5 μ L flow-through fraction purified with DDM; Lane 2: 5 μ L elution fraction purified with DDM; Lane 3: 5 μ L flow-through fraction purified with DDM-CHS; Lane 4: 5 μ L elution fraction purified with DDM-CHS; Lane 5: 5 μ L flow-through fraction purified with LMNG; Lane 6: 5 μ L elution fraction purified with LMNG. The total volumes of the flow-through and elution fractions were 14 mL and 4 mL, respectively, for each condition. (E) FSEC analysis of the elution fractions on a Superdex 200 Increase 10/300 GL column, detected via intrinsic protein fluorescence. Five micrograms of protein were applied for each condition.

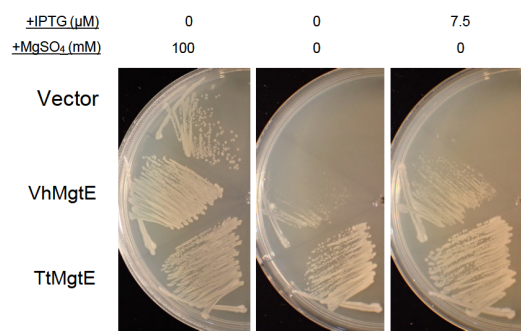


Figure 3. Growth complementation of the Mg²⁺-auxotrophic *E. coli* strain The Mg²⁺-auxotrophic *E. coli* strain (BW25113 Δ mgtA Δ corA Δ yhiD DE3), transformed with the indicated plasmids (Vector, VhMgtE, TtMgtE), was streaked on LB plates containing 50 μg/mL kanamycin. MgSO₄ and IPTG were added as indicated above the panels. The plates were incubated overnight at 37°C.

with two EK2R1 linkers), as validated by AlphaFold2 structure prediction (Supplementary Figure S1). These constructs were then screened for expression via FSEC-Nb (Figure 6B).

All three constructs presented expression levels comparable to those of the wild-type (WT) VhMgtE (Figure 6B). Although the constructs displayed similar expression profiles and AlphaFold2 prediction scores, construct 3 carried a shorter MgtE segment, a four-residue truncation at the C-terminus. We therefore expected a tighter linkage between ZZ1 and VhMgtE and selected construct 3 for purification, which we performed under three detergent conditions: DDM, DDM-CHS, and LMNG (Figure 6C,D, and Supplementary Figure S2). The IMAC elution yields for construct 3 were markedly greater than those of VhMgtE WT (DDM, 3.4 mg; DDM-CHS, 1.7 mg; and LMNG, 0.77 mg per liter of *E. coli* culture). Although the SEC profiles showed void peaks, we collected the main non-void peak fractions (Supplementary Figure S2, 0.16 mg from B3–B5 in DDM; 0.18 mg from D11–E1 in DDM-CHS; and

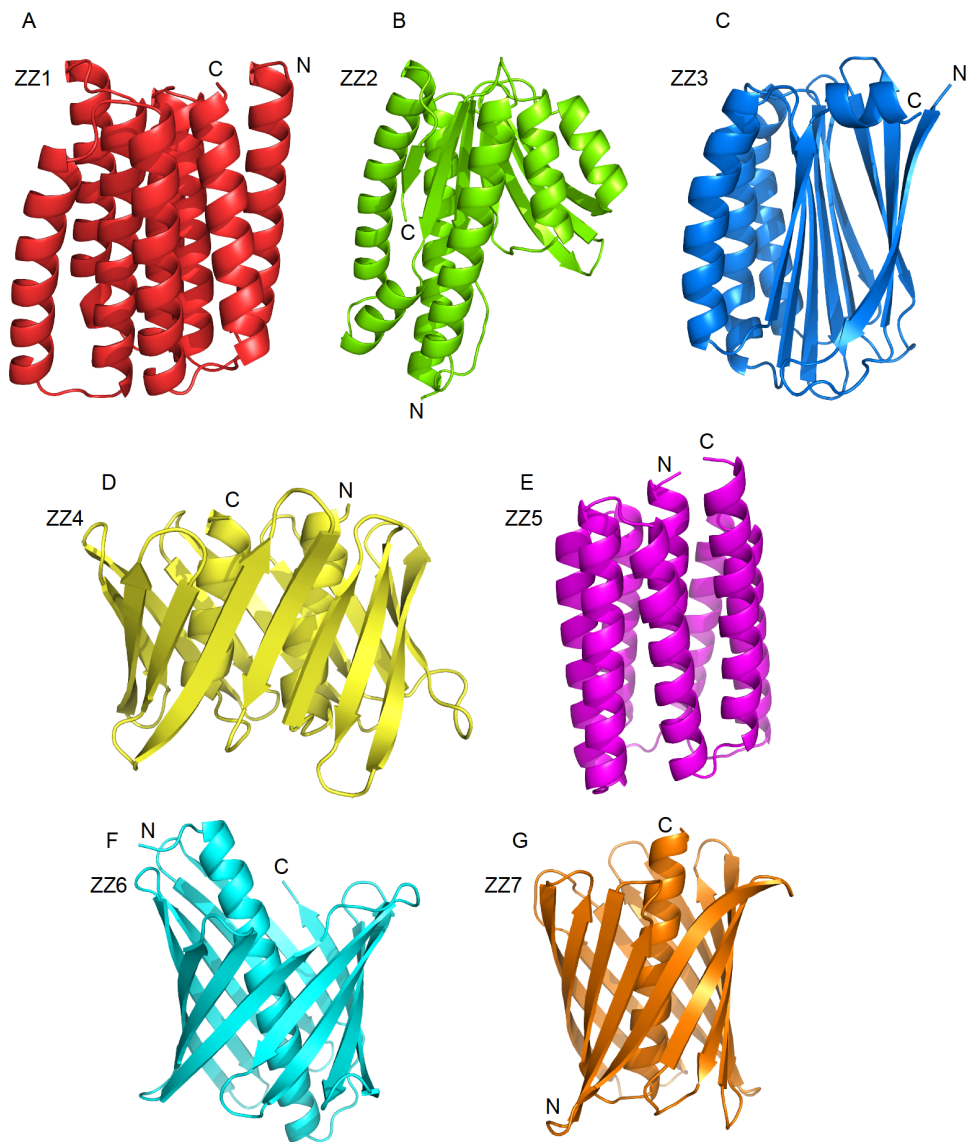


Figure 4. *De novo* protein design via RF diffusion (A–G) Cartoon representations of *de novo*-designed protein structures predicted by AlphaFold2.

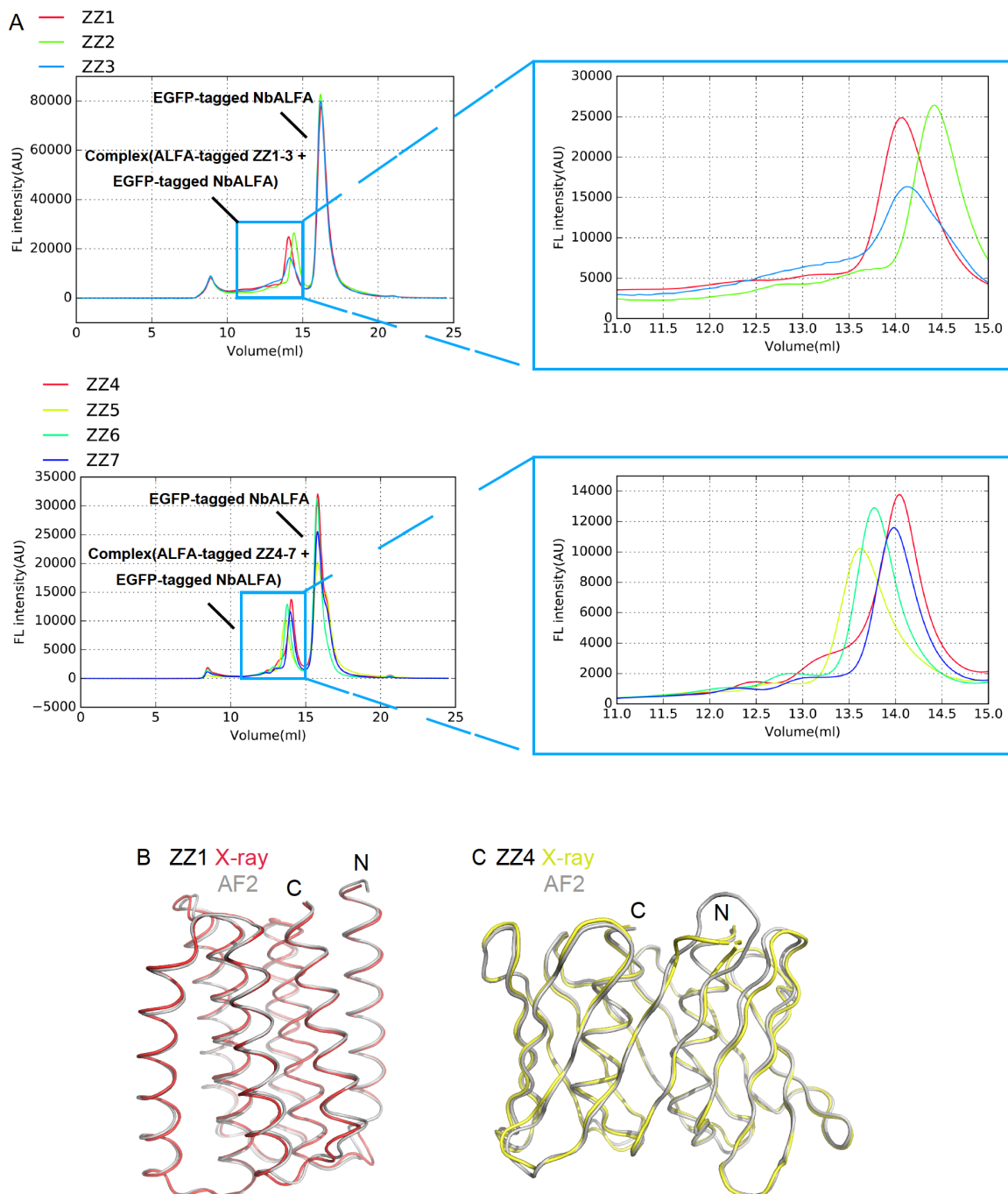


Figure 5. Expression screening and structural analysis of *de novo*-designed proteins (A) FSEC traces of unpurified *de novo*-designed proteins ZZ1–ZZ7 mixed with EGFP-tagged NbALFA on a Superdex 200 Increase 10/300 GL column, detected via EGFP fluorescence. (B,C) Crystal structures of the *de novo*-designed proteins ZZ1 (B) and ZZ4 (C). The AlphaFold2-predicted structures are superimposed and shown in gray.

0.07 mg from F11–F12 in LMNG). Subsequent FSEC analysis demonstrated that these main SEC fractions, particularly those purified in DDM, retained a well-defined peak (Figure 6D). Accordingly, the ZZ1-fused VhMgtE construct purified in DDM was reconstituted into amphipol A8-35 and again displayed a favorable peak in the SEC analysis (Figure 6E). Furthermore, preliminary cryo-EM imaging of amphipol-reconstituted ZZ1-fused VhMgtE via a 200 kV Glacios 2 microscope revealed little aggregation of particles (Figure 6F).

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

Discussion

Our bioinformatics pipeline identified a novel “mini-N” subclass of MgtE Mg²⁺ channels with atypical domain architectures, suggesting previously unrecognized structural diversity within the MgtE family. While we encountered challenges in the expression and

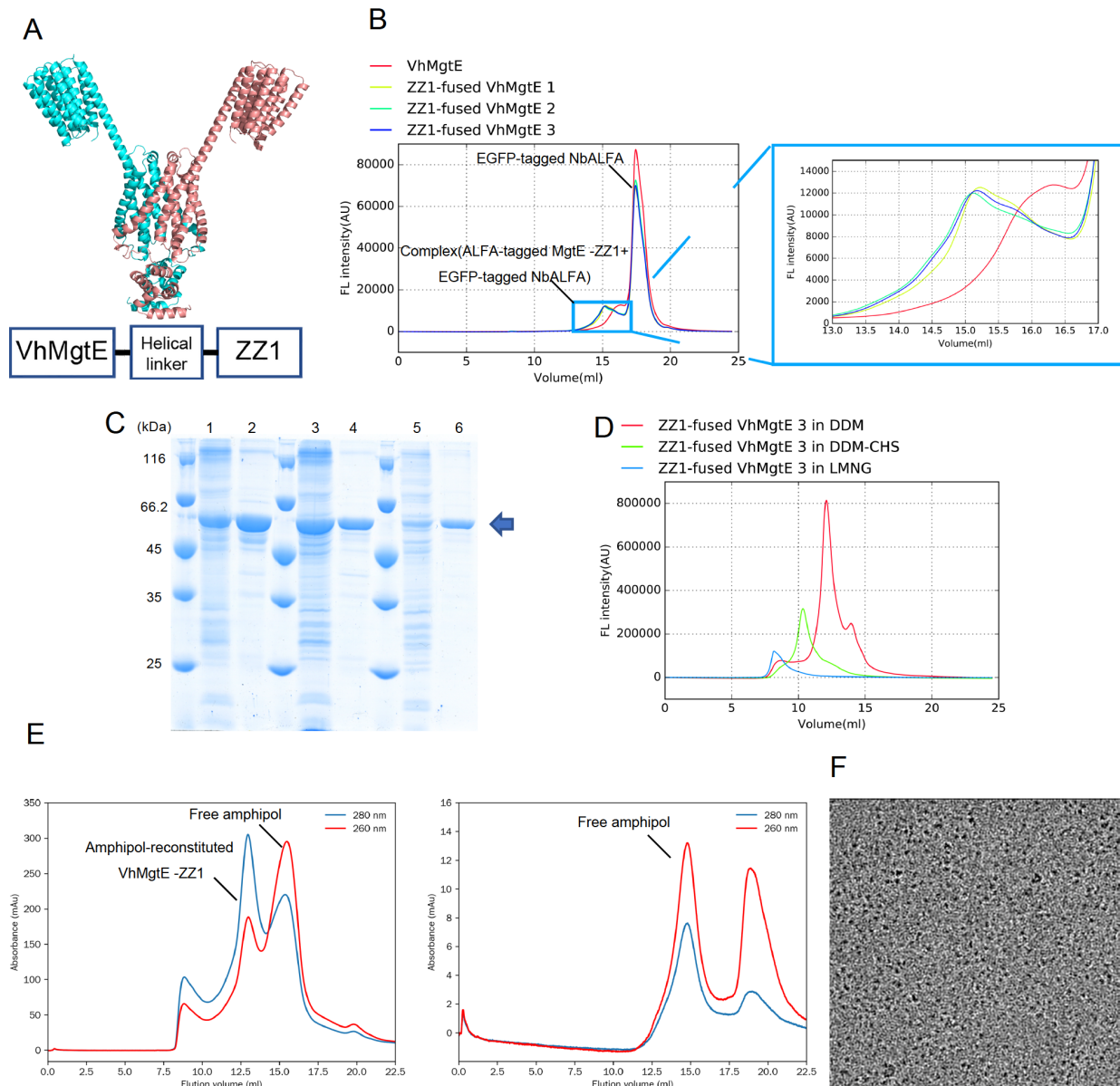


Figure 6. Purification and preliminary cryo-EM of ZZ1-fused VhMgtE (A) AlphaFold-predicted structure (Construct 3) and domain composition of the ZZ1-fused VhMgtE construct. (B) FSEC traces of unpurified ZZ1-fused VhMgtE constructs mixed with EGFP-tagged NbALFA on a Superose 6 Increase 10/300 GL column, detected via EGFP fluorescence. (C) SDS-PAGE for the purification of ZZ1-fused VhMgtE construct 3 under different conditions. Lane 1: 5 μ L flow-through fraction purified with DDM; Lane 2: 5 μ L elution fraction purified with DDM; Lane 3: 5 μ L flow-through fraction purified with DDM-CHS; Lane 4: 5 μ L elution fraction purified with DDM-CHS; Lane 5: 5 μ L flow-through fraction purified with LMNG; Lane 6: 5 μ L elution fraction purified with LMNG. The total volumes of the flow-through and elution fractions were 14 mL and 4 mL, respectively, for each condition. (D) FSEC analysis of the SEC fractions for each purification condition on a Superdex 200 Increase 10/300 GL column, detected via intrinsic protein fluorescence. The SEC profiles and SDS-PAGE results are shown in [Supplementary Figure S2](#). Five micrograms of protein were applied for each condition. (E) SEC profile of ZZ1-fused VhMgtE construct 3 after amphipol reconstitution, as detected via UV absorbance. The SEC profile of free amphipol A8-35 is also shown as a reference. (F) Cryo-EM image of ZZ1-fused VhMgtE construct 3.

purification of mini-N-type MgtE, we addressed these technical issues via AI-driven *de novo* protein design and protein engineering. We designed stable *de novo* proteins via state-of-the-art deep learning tools and determined the crystal structures of selected constructs, validating the design approach. The 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. Our current findings demonstrate that the ZZ1-fusion strategy improves purification, but whether it facilitates particle alignment and enhances cryo-EM reconstruction remains to be determined in future work.

Supplementary Data

Supplementary data are available at *Acta Biochimica et Biophysica Sinica* online.

Data Availability

The atomic coordinates of the *de novo* designed protein structures were deposited in the Protein Data Bank under the accession codes 9LNI (ZZ1) and 9LX4 (ZZ4). All relevant data are included in the paper, its supplemental information or deposited in Mendeley Data (<https://data.mendeley.com/datasets/kdns5hnmw/2>) and GitHub (https://github.com/0mae/mgte_short).

Acknowledgement

We thank the staff from BL02U1 at the 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 the Hattori laboratory for providing technical support regarding the FSEC screening of *de novo*-designed proteins. The computation time was provided by the SuperComputer System, Institute for Chemical Research, Kyoto University.

Funding

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

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- 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
- Townsend DE, Esenwine AJ, George 3rd J, 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
- Hattori M, Iwase N, Furuya N, Tanaka Y, Tsukazaki T, Ishitani R, Maguire ME, *et al.* Mg²⁺-dependent gating of bacterial MgtE channel underlies Mg²⁺ homeostasis. *EMBO J* 2009, 28: 3602–3612
- Dann III CE, Wakeman CA, Sieling CL, Baker SC, Irnov I, Winkler WC. Structure and mechanism of a metal-sensing regulatory RNA. *Cell* 2007, 130: 878–892
- 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.e13
- 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
- Chakravarty S, Melton CN, Bailin A, Yahr TL, Anderson GG, O'Toole G. *Pseudomonas aeruginosa* magnesium transporter MgtE inhibits Type III secretion system gene expression by stimulating *rsmYZ* transcription. *J Bacteriol* 2017, 199
- Goytain A, Quamme GA. Functional characterization of ACDP2 (ancient conserved domain protein), a divalent metal transporter. *Physiol Genomics* 2005, 22: 382–389
- Sahni J, Scharenberg AM. The SLC41 family of MgtE-like magnesium transporters. *Mol Aspects Med* 2013, 34: 620–628
- Schweigel-Röntgen M, Kolisek M. SLC41 transporters—molecular identification and functional role. *Current topics in membranes* 2014, 73: 383–410
- Kolisek M, Sponder G, Mastrototaro L, Smorodchenko A, Launay P, Vormann J, Schweigel-Röntgen M, *et al.* Substitution p.A350V in Na⁺/Mg²⁺ exchanger SLC41A1, potentially associated with Parkinson's disease, is a gain-of-function mutation. *PLoS One* 2013, 8: e71096
- Hattori M, Tanaka Y, Fukai S, Ishitani R, Nureki O. Crystal structure of the MgtE Mg²⁺ transporter. *Nature* 2007, 448: 1072–1075
- Maruyama T, Imai S, Kusakizako T, Hattori M, Ishitani R, Nureki O, Ito K, *et al.* Functional roles of Mg²⁺ binding sites in ion-dependent gating of a Mg²⁺ channel, MgtE, revealed by solution NMR. *eLife* 2018, 7: e31596
- Jin F, Huang Y, Hattori M. Recent advances in the structural biology of Mg²⁺ channels and transporters. *J Mol Biol* 2022, 434: 167729
- 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 Crystallogr F Struct Biol Cryst Commun* 2007, 63: 682–684
- Ishitani R, Sugita Y, Dohmae N, Furuya N, Hattori M, Nureki O. Mg²⁺-sensing mechanism of Mg²⁺ transporter MgtE probed by molecular dynamics study. *Proc Natl Acad Sci USA* 2008, 105: 15393–15398
- Tomita A, Zhang M, Jin F, Zhuang W, Takeda H, Maruyama T, Osawa M, *et al.* ATP-dependent modulation of MgtE in Mg²⁺ homeostasis. *Nat Commun* 2017, 8: 148
- Takeda H, Hattori M, Nishizawa T, Yamashita K, Shah STA, Caffrey M, Maturana AD, *et al.* Structural basis for ion selectivity revealed by high-resolution crystal structure of Mg²⁺ channel MgtE. *Nat Commun* 2014, 5: 5374
- Teng X, Sheng D, Wang J, Yu Y, Hattori M. Ion selectivity mechanism of the MgtE channel for Mg²⁺ over Ca²⁺. *iScience* 2022, 25: 105565
- 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 Biol* 2021, 19: e3001231
- Brahma R, Raghuraman H. Characterization of a novel MgtE homolog and its structural dynamics in membrane mimetics. *Biophys J* 2024, 123: 1968–1983
- Wang M, Zhao Y, Hayashi Y, Ito K, Hattori M. Novel Mg²⁺ binding sites in the cytoplasmic domain of the MgtE Mg²⁺ channels revealed by X-ray crystal structures. *Acta Biochim Biophys Sin* 2023, 55: 683–690
- Chatterjee S, Brahma R, Raghuraman H. Gating-related structural dynamics of the MgtE magnesium channel in membrane-mimetics utilizing site-directed tryptophan fluorescence. *J Mol Biol* 2021, 433: 166691
- Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes—a review. *Nat Prod Rep* 2016, 33: 988–1005
- Medema MH, de Rond T, Moore BS. Mining genomes to illuminate the specialized chemistry of life. *Nat Rev Genet* 2021, 22: 553–571
- Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol* 2008, 18: 581–586
- 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
- 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
- Kortemme T. *De novo* protein design—From new structures to programmable functions. *Cell* 2024, 187: 526–544
- Zhu J, Lu P. Computational design of transmembrane proteins. *Curr Opin Struct Biol* 2022, 74: 102381
- 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
- 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
33. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, Heger A, *et al.* Pfam: the protein families database. *Nucleic Acids Res* 2014, 42: D222–D230
 34. Eddy SR, Pearson WR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011, 7: e1002195
 35. 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
 36. R Core Team. R: a language and environment for statistical computing, vol. *R Foundation for Statistical Computing* 2015
 37. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, *et al.* Welcome to the tidyverse. *JOSS* 2019, 4: 1686
 38. Lee S, Cook D, Lawrence M. plyranges: a grammar of genomic data transformation. *Genome Biol* 2019, 20: 4
 39. Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: efficient manipulation of biological strings. *R package version* 2019, 2: 10.18129
 40. Pedersen T. Patchwork: the composer of plots. *CRAN: Contributed Packages* 2019
 41. 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
 42. Götzke H, Kilisch M, Martínez-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
 43. Kawate T, Gouaux E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* 2006, 14: 673–681
 44. 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
 45. 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
 46. Liu K, Zhou H, Xu Q, Kong HT, Zhang KH, Wang WW, Li MJ, *et al.* BL02U1: the relocated macromolecular crystallography beamline at the Shanghai Synchrotron Radiation Facility. *Nuclear Science and Techniques* 2023, 34: 193
 47. Kabsch W. Xds. *Acta Crystallogr D Biol Crystallogr* 2010, 66: 125–132
 48. Liebschner D, Afonine PV, Baker ML, Bunkóczi 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
 49. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010, 66: 486–501
 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, Palacín M. Membrane protein stabilization strategies for structural and functional studies. *Membranes* 2021, 11: 155
 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 Jr. GN. 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, Söding 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