



Review

Engineered Chimera Protein Constructs to Facilitate the Production of Heterologous Transmembrane Proteins in *E. coli*

Adeyemi Ogunbowale and Elka R. Georgieva *

Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, TX 79409, USA; aogunbow@ttu.edu

* Correspondence: elgeorgi@ttu.edu

Abstract: To delve into the structure–function relationship of transmembrane proteins (TMPs), robust protocols are needed to produce them in a pure, stable, and functional state. Among all hosts that express heterologous TMPs, *E. coli* has the lowest cost and fastest turnover. However, many of the TMPs expressed in *E. coli* are misfolded. Several strategies have been developed to either direct the foreign TMPs to *E. coli*'s membrane or retain them in a cytosolic soluble form to overcome this deficiency. Here, we summarize protein engineering methods to produce chimera constructs of the desired TMPs fused to either a signal peptide or precursor maltose binding protein (pMBP) to direct the entire construct to the periplasm, therefore depositing the fused TMP in the plasma membrane. We further describe strategies to produce TMPs in soluble form by utilizing N-terminally fused MBP without a signal peptide. Depending on its N- or C-terminus location, a fusion to apolipoprotein AI can either direct the TMP to the membrane or shield the hydrophobic regions of the TMP, maintaining the soluble form. Strategies to produce G-protein-coupled receptors, TMPs of *Mycobacterium tuberculosis*, HIV-1 Vpu, and other TMPs are discussed. This knowledge could increase the scope of TMPs' expression in *E. coli*.

Keywords: *E. coli* expression host of heterologous transmembrane proteins; transmembrane protein fusion strategies; protein engineering; soluble transmembrane proteins

Citation: Ogunbowale, A.; Georgieva, E.R. Engineered Chimera Protein Constructs to Facilitate the Production of Heterologous Transmembrane Proteins in *E. coli*. *Int. J. Mol. Sci.* **2024**, *25*, 2354. <https://doi.org/10.3390/ijms25042354>

Academic Editor: Clemente Capasso

Received: 11 January 2024

Revised: 8 February 2024

Accepted: 14 February 2024

Published: 16 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Membrane proteins fulfill vital physiological functions in all living organisms and, in some cases, are involved in causing disease conditions due to, for example, mutations [1–4]. Our focus here is on transmembrane proteins (TMPs), which constitute about 60% of pharmacological targets [5–7]. Because of their essential roles, acquiring detailed knowledge of these proteins' molecular mechanisms is vital to gain control over their functions, e.g., designing inhibitors of TMPs encoded by pathogens or characterizing and possibly reversing protein malfunction. In this regard, in vitro studies are important, particularly obtaining the high-resolution structure of TMPs, studying their interactions with ligands, elucidating the conformational rearrangements taking place during their function, and conducting functional assays [8–10]. However, these in vitro studies typically require large quantities of highly pure TMPs. To obtain the highly pure TMPs, heterologous expression in a host organism, such as *E. coli*, yeast, insect, and mammalian cells, is typically required [11–14]. Among these, *E. coli* is the most extensively used protein expression host because of its relatively low cost, rapid expression rate, and easy genetic manipulations [15–18]. *E. coli* is particularly suitable for producing bacterial TMPs. However, some extraneous bacterial TMPs have been found in the insoluble inclusion bodies when expressed in *E. coli*, e.g., *Mycobacterium tuberculosis* (*Mtb*) TMPs [19]. Furthermore, when expressed in *E. coli*, eukaryotic and viral TMPs are almost inevitably aggregated since they require different than *E. coli*'s translocon and chaperone systems for delivery to the membrane and folding [20]. Refolding and purifying these aggregated proteins is

laborious and often inefficient [17,21]. To overcome these obstacles in TMPs' production, protein engineering has been advanced toward designing and expressing in *E. coli* chimeric protein constructs containing a specific protein tag fused to the target TMP. This allows directing the TMPs of interest to the *E. coli* plasma membrane or maintaining it in soluble form [20,22–26]. Such approaches provide new possibilities to study the structure and function of TMPs, which aid the understanding of physiological mechanisms and pharmacological developments.

This review focuses on some of the most successful TMPs' fusion strategies, which made it possible to express eukaryotic, viral, and prokaryotic TMPs in the easy-to-handle state in *E. coli*. The development and applications of fusion tags such as signal peptide and signal peptide with maltose binding protein (MBP), mystic protein, apolipoprotein AI, and mature MBP (without signal peptide) are summarized in Table 1 and discussed in greater detail in the following sections of this review paper.

Table 1. Fusion tags, the transmembrane proteins (TMPs) produced using them, the benefit of these fusion strategies for TMPs structural and/or functional studies, and the reported TMPs' yields are listed.

Fusion Tag	Produced TMP	Benefit for Structural and/or Functional Studies	Protein Yield
MBP signal peptide/entire MBP	Serotonin 5-HT1A [27], neurotensin receptor [28], NK-2 (neurokinin A) [29], M2 muscarinic acetyl choline receptor [30], and peripheral cannabinoid receptor [31]	It promotes the proper folding and insertion of the recombinant fusion protein into the plasma membrane. It supports the application of functional assays in the study of the activities of the transmembrane protein.	Yeliseev et al. (2005) showed that the small scale purification yielded 1–2 mg of recombinant receptor/ 1L of culture [31]
Mistic protein	aKv1.1 channel [20], and eukaryotic type I rhodopsin [22]	It promotes high expression yield of heterologous TMPs as well as facilitating the expression of functional proteins with both N-terminus inside or N-terminus outside.	The yield of Mistic-aKv1.1 was approximately 2 mg/L culture [20] The Mistic-ARI yield amounted to roughly 0.12 mg/L culture [22]
Apolipoprotein AI	Mtb-EfpA [32], EmrE transporter [26], human cyt b5 [26], HSD17β3 [26], GluA2 [26], DsbB [26], CLDN1 [26], CLDN3 [26], S5aR1 [26], S5aR2 [26], NRC-1bR [26], OmpX [26], and VDAC1 [26]	The tertiary conformation of the TMP–lipid–apoAI forms a discoidal nanoparticle stabilized by a double belt of apoAI. It increases the solubilization of TMPs with high levels of expression and supports the functional study of the protein (e.g., ligand binding and protein–protein interaction).	Mizrachi et al. (2015) showed that the diverse range of IMPs yielded approximately 5–10 mg/1 L of culture [26]
mMBP without signal peptide	Vpu [24], p18 [33], and Yqgp protease [34]	It is useful as a purification affinity tag when in combination with polyhistidine tag for Ni-affinity purification. It is a natural fusion tag that is a solubility enhancer.	The approximate yield of MBP-p18 reached around 20mg/L culture [33] The yield of MBP-Yqgp protease was 2 mg/1 L of culture [34]

2. Fusion Strategies to Produce Heterologous Transmembrane Proteins in *E. coli*

2.1. Fusion Proteins Aid the Insertion and Folding of Heterologous TMPs in the *E. coli* Plasma Membrane

2.1.1. Signal Peptides and Precursor Maltose Binding Protein Fusion Strategies

The fusion of a signal peptide to the N-terminus of eukaryotic TMPs was among the earliest strategies to produce these proteins in *E. coli*. Typically, upon synthesis, eukaryotic, viral, and some bacterial TMPs are not recognized for membrane insertion and end up in a misfolded inclusion body state [27,35]. However, the addition of short 20–30 amino acids signal peptides to the target protein's N-terminus makes the protein recognizable by the *E. coli* machinery for trafficking to the plasma membrane [12]. Therefore, for expression in *E. coli*, the periplasmic leader sequences derived from ompT, ompA, pelB, phoA, malE, lamB, β -lactamase and PelB can generally be used to direct eukaryotic TMPs to *E. coli*'s plasma membrane [12,36]. In this case, the signal peptide–TMP polypeptides are translocated post-translationally via the Sec-dependent pathway. Conversely, the native to *E. coli* TMPs have highly hydrophobic signal peptides and are translocated via the SRP-dependent pathway utilizing a co-translational mechanism. These hydrophobic signal peptides (e.g., the peptide derived from the DsbA protein) can also be used as an N-terminal tag to express heterologous TMPs [36].

The application of the malE (maltose binding protein, MBP) signal peptide has been successful in the production of several members of the G protein-coupled receptors (GPCRs) family. In these studies, the peptide containing the signal sequence for periplasmic localization of the *E. coli*-encoded MBP, or even the entire MBP with the signal peptide included (the precursor MBP, pMBP), was fused to the N-terminus of GPCRs (Figure 1). This chimeric construct was directed to the plasma membrane, where it adopted a natively folded and functional state [28,30,37]. Initially, this method was used to express in *E. coli* serotonin 5-HT_{1A} and neurotensin receptors in a membrane-bound state [28,38]. Later, the strategy was applied to several other GPCRs, such as the rat NK-2 (neurokinin A) receptor [29], rat neurotensin receptor [37], M2 muscarinic acetylcholine receptor [30], peripheral cannabinoid receptor [31], and others. The success of these studies was partly due to the extracellular localization of the GPCRs' N-terminus, which allowed the MBP signal sequence to direct this protein region to the *E. coli* periplasmic space and ensure proper orientation of the first TM helices of the receptors [28,39]. These advancements were instrumental in progressing GPCRs' structural and functional studies, aiding pharmacological developments. In their original work, Henderson and colleagues and later studies [28,40] found that the expressed in *E. coli* membranes neurotensin receptor with an N-terminus fusion signal sequence with and without the entire MBP could bind the ligand neurotensin. However, the presence of pMBP significantly increased the receptor–ligand affinity. After that, the high-resolution structures of GPCRs produced in *E. coli* were solved, thus further enhancing the understanding of these proteins' structure–function relationship. As a result, multiple X-ray structures of neurotensin receptor one was solved at high resolution [40]. Furthermore, the high-level functional GPCRs' expression in *E. coli* have greatly facilitated NMR studies of these proteins as well, providing structural and dynamic insights underlying the interaction with agonist and antagonist molecules [41,42].

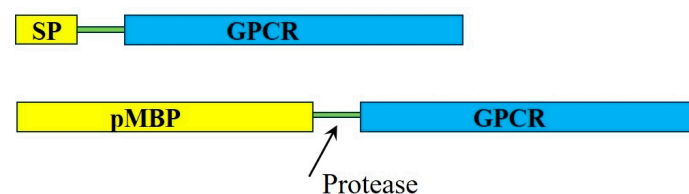


Figure 1. Schematic representation of the pMBP signal peptide (SP)-GPCR and pMBP-GPCR chimeric constructs used to produce functional GPCRs in *E. coli* plasma membranes. In some cases, a

protease site between pMBP and GPCR was introduced to remove the tag after purifying the protein.

All these studies were based on a similar construct design and cloning in the *E. coli* expression vectors pRG/II-pMBP or pRG/III-hs-pMBP created in the original studies of neurotensin under the control of *lac* promoter and IPTG induction [28,29,37]. The original vector containing the Thrombin (Thr) cleavage site to remove the tag was further replaced by a more selective HRV 3C protease site because Thr was found to aggregate the GPCR [43]. In addition to protein engineering to incorporate a signal peptide, the high-yield production of functional GPCRs in *E. coli* was improved through the optimization of protein expression temperature (typically at 22 °C or lower) and concentration of IPTG (typically low concentration of 0.1–0.3 mM was used) [30,31].

2.1.2. Mystic Protein Fusion Strategies

In other studies, the mistic protein fused to the N-termini of eukaryotic TMPs for expression in *E. coli* was utilized [20,44]. Mystic (an acronym for “membrane-integrating sequence for the translation of integral membrane protein constructs”) is encoded by *Bacillus* species and was originally found in *Bacillus subtilis* [45,46]. The protein folds into a four-helix bundle with a hydrophobic core and a significant fraction of polar and charged amino acids (Figure 2A) [45]. Mystic is found in both cytoplasmic and membrane-bound states [20,47]. The mistic protein of *Bacillus subtilis* (M110) comprises 110 amino acid residues and has a net charge of −12.0 at pH 7. It has been suggested that its acidic nature enables the tight association with the lipid bilayer alone or as a fusion tag when expressed in *E. coli* [20]. However, the shorter than M110 mistic constructs or orthologs found in other species with also highly acidic nature, e.g., the 84 amino acids C-terminal truncated version of M110 (referred to as M1) as well as mistic from *B. leicheniformis* (referred to as M2) and from *B. mojavensis* (referred to as M3), are highly soluble with almost exclusive cytoplasmic localization.

In contrast, the mistic from *B. atrophaeus* (M4) is comparable to M110 membrane affinity [20]. Interestingly, outside the membrane, soluble mistic forms fibrils with a protomer’s structure that is largely different from those determined by NMR for non-fibrinous mistic [45,46]. The fibrous structures possibly shield the charged regions of mistic and facilitate its interaction with hydrophobic membranes [46]. To determine the membrane association regions, Marino et al. analyzed truncated mistic constructs containing individual or combined helices. They found that helices 1, 2, and 4 interact with lipid membranes, whereas helix 3 is primarily soluble [47]. It was found in the same study that the single helices 1, 2, and 4 fused to the N-terminus of Y4 GPCR can direct the protein to the *E. coli* membrane [47], similarly to full-length (FL) mistic [48]. However, only the fusion of Y4 GPCR to helix 2 yielded an expression level comparable to those when FL mistic was used and a segment of amino acids “GLDAFIQLY” in helix 2 was identified as the minimal sequence for mistic and its fusion protein to interact with the membrane [47].

It has been proposed that the absence of a detectable signal sequence, which is a unique feature of mistic, enables this protein to avoid the Sec translocon’s pathway of *E. coli*; due to this, mistic’s and mistic-tagged TMPs’ expression does not overload the protein translation machinery [20,45]. Therefore, high expression yields of heterologous TMPs in mistic-tagged TMP chimeras can be achieved [20,49]. It has also been reported that mistic facilitates the expression of functional proteins with both the N-terminus inside or N-terminus outside the cell [44,50], suggesting its adaptive membrane-bound topology to accommodate the expression and folding of the target protein.

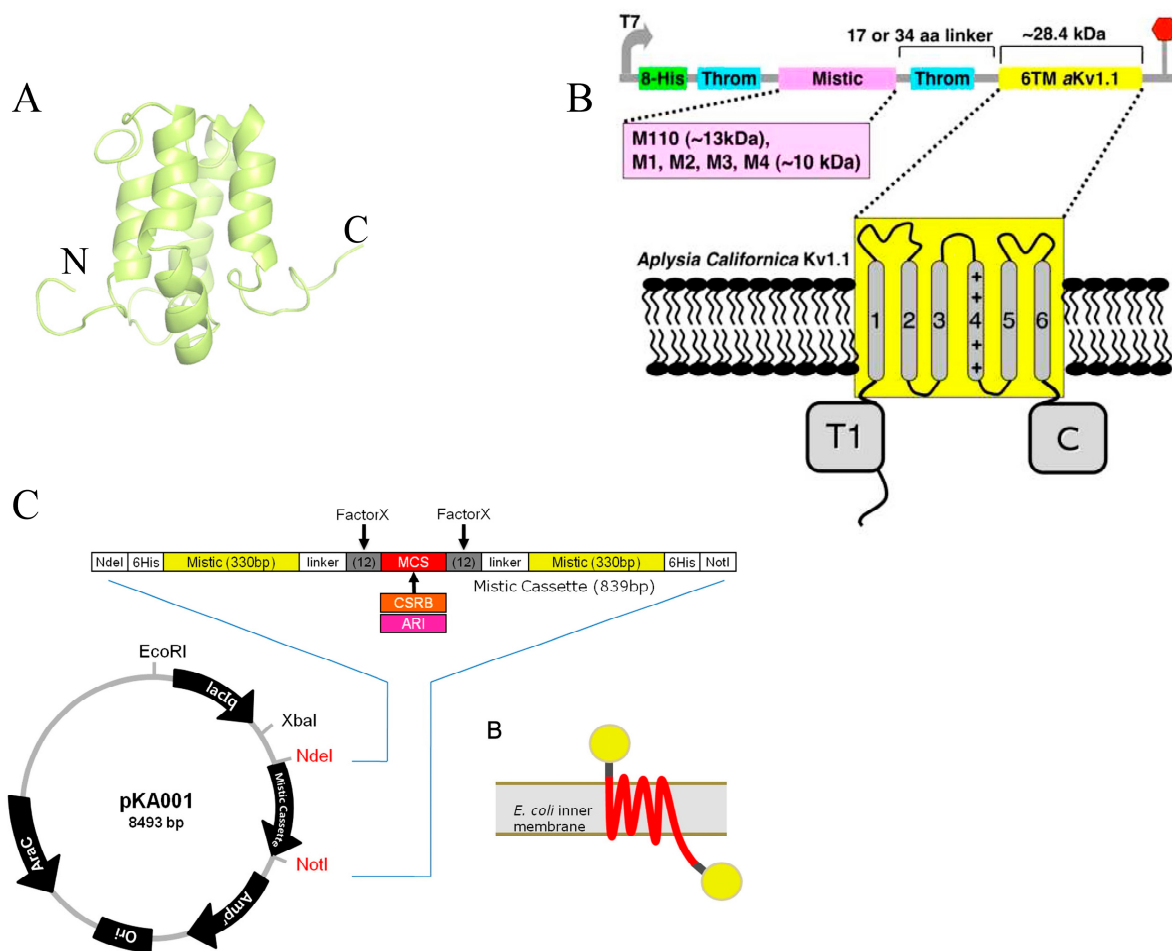


Figure 2. Mystic fusion strategy to produce heterologous TMPs in *E. coli*. (A) NMR structure of mistic, PDB code 1YGM. The N- and C-termini are indicated. The protein folds into a 4-helix bundle. (B) Fusion strategy used to produce the aKv1.1 channel and aKv1.1 6TM. The figure was adopted from Ref. [20] with permission from Elsevier (License number 5665041200130). Only one mistic copy was fused to the N-terminal of aKv1.1. (C) Fusion strategy to produce eukaryotic rhodopsin variants. The figure adopted from Ref. [22] with permission from Elsevier (license number 5665050773775). Two copies of mistic were fused to the N- and C-termini of the TMPs.

In addition to GPCR [47,48], the aKv1.1 channel and its six-transmembrane helix (6TM) domain have also been successfully produced in *E. coli* as mistic–fusion constructs (Figure 2B) [20]. It was found that the expression of the aKv1.1 6TM and shortening of the mistic–aKv1.1 6TM linker had a positive effect on the target protein expression levels due to possibly better interaction between mistic’s C-terminus and aKv1.1 6TM as well as reduced proteolysis in the linker region [20]. It was further established that the fusion of aKv1.1 6TM to the C-termini of mistic M110 and mistic M4 resulted in comparable expression levels, which was possibly because both M110 and M4 aided the membrane insertion of aKv1.1 6TM similarly [20].

The mistic fusion strategy has facilitated the studies of the eukaryotic type I rhodopsin as well because it enabled the economical production of the functional form of this protein in *E. coli*. [22]. Interestingly, this study found that two mistics copies fused to the N- and C-termini of the target proteins were needed to direct them into the *E. coli* membranes; the study was conducted on several eukaryotic rhodopsin variants, including ARI and CSRB, as well as other eukaryotic TMPs (Figure 2C) [22]. It was also found in this study that the mistic moieties of the fusion construct do not severely affect the proton transport function of the ARI [22], which might be advantageous as the expression level

of some heterologous TMPs is relatively low, and the removal of fusion tag typically leads to a further reduction in protein quantities.

Expressed in *E. coli* and purified mistic-tagged eukaryotic proteins were also used as antigens for raising polyclonal antibodies, and it was found that the mistic-TMPs antibodies recognized the corresponding TMPs in native membranes more efficiently than the antibodies raised against just the soluble domains of these TMPs [49]. As the study's authors suggest, this could be because the soluble domains of the studied TMPs might adopt a distinct conformation when included in the membrane-bound FL protein vs. truncated soluble versions [49].

Generally, the use of mistic for the expression of different membrane proteins depends on the target protein's proteolytic susceptibility, the protein expression induction conditions, and the number of amino acids that connect the mistic to the recombinant membrane protein [45]. Mistic's structure and membrane affinity are critical for their ability to facilitate the production of heterologous TMPs [45]. This was confirmed by the work of Tarmo and colleagues, who used three mutant variants of mistic protein (W13A, Q36E, and M75A) with amino acid substitutions in different helical regions. The expression level of the mutants in the cytoplasm and membrane were tested when alone and fused to aKv1.1. It was seen that the mutation of methionine-75 to alanine destabilized the structure of mistic protein due to its substantial partitioning between the membrane and cytoplasm. Also, when the mutant was fused to aKv1.1, there was no expression of this protein in the membrane [45].

The mistic protein can also be combined with another fusion protein to increase the expression rate of some TMPs. Ananda et al. discovered that the *CB2* gene can be expressed only when mistic and TarCf are fused to its N- and C-terminus, respectively, indicating a synergistic effect of the two tags on the expression [44].

2.1.3. Apolipoprotein A-I Fusion Strategy

Apolipoprotein A-I (apoAI) belongs to the spherical high-density lipoproteins (HDLs), which are abundant in human plasma. ApoAI is a highly α -helical protein of 28 kDa, which in vivo serves as a "glue" to hold HDL particles together [51]. The protein is easily produced in *E. coli*. It has been widely utilized in structural and functional studies of membrane-reconstituted TMPs as the tertiary complex of target TMP-lipid-apoAI form discoidal nanoparticles stabilized by a double belt of apoAI [8,52–54].

Recently, motivated by a study on soluble TMPs fused to the N-terminus of apoAI (discussed in more detail below) [26], our research group designed and expressed in *E. coli* a chimera construct of apoAI with the *Mycobacterium tuberculosis* EfpA (*Mtb*-EfpA) drug exporter [32]. By doing so, we produced, for the first time to the best of our knowledge, highly pure FL *Mtb*-EfpA in quantities sufficient for downstream in vitro characterization. Remarkably, when reconstituted in lipid, because of the presence of apoAI in the apoAI-EfpA fusion construct, we observed by electron microscopy the formation of protein-lipid nanoparticles [32], which are similar to previously described nanodiscs [52–54]. This suggests that we can carry out future studies on EfpA's properties (e.g., drug binding, structure determination, assessing the conformational dynamics) using these two-component (apoAI-EfpA protein and lipid) nanoparticles. Moreover, the methodology could also be adopted in studies on other TMPs.

Interestingly, apoAI is typically expressed as a soluble protein in *E. coli* [55]. We also found that the untagged EfpA is deposited in inclusion bodies upon expression [32]. Therefore, there is a question of how and why the apoAI-EfpA is directed to the membrane. One explanation could be that the additional sequence at EfpA's N-terminus prevents the protein from misfolding at the stage of protein translation as was previously proposed for mistic's mechanism to prevent TMPs' aggregation [46]. Similar effects on protein expression were also observed when TMPs were tagged at their N-termini with glutathione S-transferase (GST) [56] or YbeL and YnaI [57].

2.2. Fusion Proteins Aid the Production of Heterologous TMPs in Soluble Form in *E. coli*

In addition to membrane-bound heterologous TMPs, some TMPs or their transmembrane portions have been produced in *E. coli* in soluble form. Some of these developments are discussed below.

2.2.1. Mature (without Signal Peptide) Maltose-Binding Protein Fusion Strategies

The mature MBP (mMBP) lacking the signal peptide to direct it to the periplasm has also been used to produce a range of TMPs in *E. coli*, which remain inside the cell due to the absence of the MBP signal sequence. These TMPs were commonly small and obtained in soluble form, but some proteins or protein fractions were also found in the inclusion bodies [19,24,33,34,58,59]. In addition to serving as a solubilization tag, MBP is also very useful as a purification affinity tag, which, in combination with a polyhistidine tag for Ni-affinity purification, makes it a powerful tool to isolate proteins of high purity [24,60].

Some studies suggest that the solubility-enhancing property of mMBP is mediated by its open conformation of the ligand-binding cleft state [23,61–63]. However, the exact mechanism by which mMBP increases the solubility of recombinant proteins needs to be better understood. Lebendiker and colleagues suggested that mMBP can act either as a magnet that attracts chaperones toward the environment of the recombinant protein or by forming micelle-like aggregates that hold incompletely folded proteins [62]. It was also proposed that mMBP acts as an electrostatic shield by reducing the electrostatic repulsion between highly charged soluble polypeptide extensions, thereby reducing the chance of repulsion [62,64]. One plausible explanation for the increased solubility of the mMBP-tagged TMPs is that the located at the N-terminus MBP moiety is translated first from the ribosome and becomes fully folded before the TMP is translated [24,65], but to whether the TMP is natively folded or just held by mMBP in the solution could depend on the particular protein and needs further characterization.

mMBP fusion was instrumental in producing several TMPs of *Mtb* for structural studies by NMR [19,58]. As discussed above, when expressed in *E. coli*, *Mtb*'s TMPs are typically deposited in the insoluble fraction. Therefore, using a mMBP solubilization tag fused to the N-termini of the TMPs of interest proved helpful in producing these proteins. It is tempting to mention that mMBP-tagged TMPs of *Mtb* were not found in the *E. coli* plasma membrane [19,58], which is the opposite to what we observed in the case of apoAI-EfpA [32]. These differences might be because of the TMPs size, i.e., single-pass small TMPs [66] vs. large multi-pass TMP [67], but future examinations might be needed to understand this better.

A study of a truncated form (p18) of the Bax apoptotic protein found that the highly hydrophobic and membrane-residing p18 can be expressed and handled in soluble form when fused to the C-terminus of mMBP [33]. This soluble form was competent in interacting with the membrane of isolated liver mitochondria and triggering cytochrome c release in a dependent manner [33].

An interesting study was reported about catalytically active soluble oligomers of mMBP-tagged YqgP protease, which could cleave a substrate within the transmembrane domain [34]. YqgP is a membrane-residing rhomboid protease homolog believed to have an active site in the membrane interior [68,69]. It may well be that the YqgP's oligomerization outside the membrane protects the hydrophobic protein regions from the aqueous environment and provides conditions similar to the membrane for assembling the enzymatic site. Although no structural information about these soluble oligomers was delivered, one would expect that the protein monomers should have similar conformations in soluble oligomers and membranes to maintain the activity.

Similarly to YqgP protease, our lab recently found that the HIV-1 Vpu protein also forms soluble oligomers when expressed as a fusion construct with mMBP (Figure 3) [24,25]. Previously, Vpu was exclusively considered an TMP. Although the possible physiological role of these soluble Vpu oligomers is currently unknown, the existence of such

a role cannot be ruled out. Significantly, the soluble Vpu could interact with membranes undergoing conformational changes.

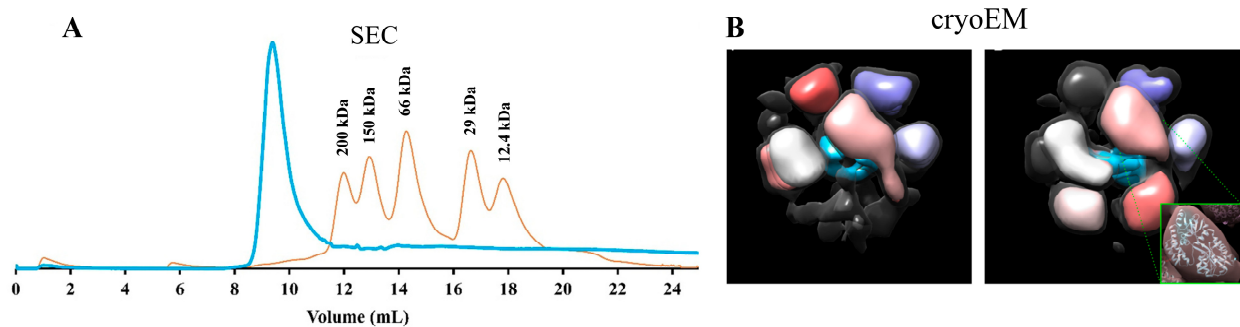


Figure 3. MBP-Vpu chimera construct forms soluble oligomers. **(A)** Size exclusion chromatography (SEC) of purified 50 μ M MBP-Vpu (blue) shows that the protein forms oligomers with molecular weight greater than 250 kDa. The SEC of a mixture of protein molecular weight standards is in white brown, and the peaks corresponding to proteins with different molecular weights are indicated. The figure was adopted from Ref. [24] with permission from Elsevier (license number 5665051248242; accessed on 9 November 2023). **(B)** cryoEM analysis of soluble MBP-Vpu oligomers revealed predominantly hexamers and hexamer-to-heptamer equilibrium. The MBP moieties of each MBP-Vpu monomer are colored in pink, gray, shades of purple, and red. The electron density, which most likely represents the Vpu oligomerization core, is colored in blue. The figure was adopted from Ref. [25] under the conditions of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>; accessed on 9 November 2023).

All these examples show that the fusion strategy with mMBP successfully aids the production in *E. coli*, where it is otherwise difficult to obtain heterologous TMPs to facilitate downstream investigations on these proteins. The discussed studies also demonstrated that the TMPs produced in soluble form retain activities in terms of being able to interact with membranes and fulfill their catalytic functions [24,33,34]. Therefore, mMBP can be used as a powerful protein engineering tool to manipulate the TMPs' production in *E. coli*, but it could be utilized in enzyme immobilization for biotechnological applications [70].

2.2.2. Apolipoprotein A-I Strategies to Produce Soluble TMPs

A study by Mizrachi and colleagues has described a method to obtain soluble TMPs by fusing them to the N-terminus of truncated apoAI [26]. The authors found that apoAI forms a scaffold around the hydrophobic TMP's regions to shield them from the water environment and stabilize the TMP-apoAI complex. Multiple prokaryotic and eukaryotic TMPs, including single and multi-pass and oligomeric TMPs with C-terminal apoAI tags, were produced in substantial quantities in *E. coli* in soluble form outside the membrane. Furthermore, these apoAI-scaffolded TMPs were stable and uniform, and they were characterized by small angle X-ray scattering (Figure 4). Remarkably, the authors found that the solubilized EmrE transporter retains ligand-binding activity. Due to the fact that a large number of TMPs were studied, this methodology promises that many other TMPs can be expressed in *E. coli* in soluble form, and further functional (e.g., ligand binding and protein-protein interaction) and structural studies can be conducted on them. It is worth mentioning that these apoAI-tagged proteins were not directed to the membrane, which was possibly because the apoAI was expressed after the corresponding TMP. Also, care was taken that the prokaryotic TMPs were genetically modified and/or mMBP was added to their N-termini to keep them in the cytoplasm [26].

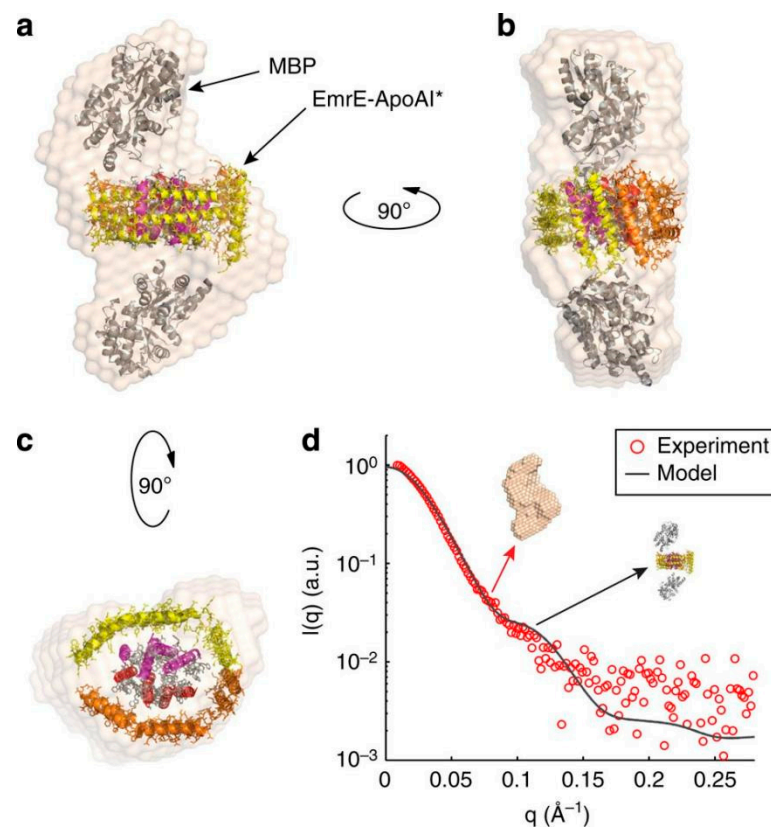


Figure 4. Soluble apoAI–EmrE studies by SAXS. (a–c) Multiple views of the reconstructed particle envelope calculated ab initio from the dimer SAXS data (red circles in (d)) using DAMMIF [71]. The asterisk (*) in (a) denotes ApoAI lacking its 43-residue globular N-terminal domain. (d) Comparison between the experimental scattering profile of the dimer (red circles) and the theoretical profile calculated for the proposed model using CRYSOLOG software (solid line). The figure was reproduced from Ref. [26] under the conditions of Creative Commons Attribution 4.0 International License. Note, in addition to the C-terminal apoAI tag, the EmrE protein had an N-terminally fused mMBP tag as well. The dimeric state of EmrE was preserved in the soluble form.

2.2.3. Other Protein Design Strategies to Produce and Stabilize Soluble TMPs

In addition to fusion tags, other protein engineering strategies have helped to overcome the challenges imposed by the low expression and instability of TMPs produced in *E. coli*. In the last decades, multiple approaches with different levels of applicability have been reported. Here, we highlight just two examples.

Protocols for enhancing the expression level of TMPs in *E. coli* and improving the stability of the purified protein for structure determinations by termini restraining tags (such as variants of green fluorescence protein, thioredoxin or T4 lysozyme) were also developed [72]. In this recent study, along with TMPs expressed in *Pichia pastoris* and mammalian cells, prokaryotic proteins, such as DsbB cloned in the pET28b vector, were expressed in *E. coli*, and their structure successfully was determined [72].

“Solubilization by design” was used to produce in *E. coli* soluble dimers of motility protein B (MotB) [73], which is a component of bacterial flagellum [74]. In this study, the two transmembrane helices of the MotB dimer were replaced by a leucine zipper; the dimers were stable and monodisperse, and they were composed of adequately folded subunits; these engineered MotB dimers were of high quality for crystallization and structure determination [73].

A 24-amino acids peptide was designed to form an amphipathic helix with a “flat” hydrophobic surface that would interact with a transmembrane protein as a detergent [75]. Alone, the peptide forms a homo-oligomeric 4-helix bundle with a helix length of 30 \AA , which is sufficient to traverse the membrane; when the peptide was mixed with

bacteriorhodopsin and rhodopsin, a large percentage of these TMPs (>60%) remained in solution even without detergent [75].

3. Conclusions

Unlike soluble proteins, the expression, purification, and characterization of TMPs is notoriously difficult and expensive. To minimize the cost and time for heterologous TMPs' production, *E. coli* has become a host of choice for producing these proteins. However, due to physiological differences in translocon systems among TMPs' original organisms and *E. coli*, many of the proteins of interest end up in a misfolded and challenging-to-handle state because *E. coli*'s TMP membrane-targeting machinery does not recognize them. Furthermore, the refolded from this aggregated state TMPs often have low or no activity. Several strategies to produce heterologous TMPs in *E. coli* in either membrane-bound or soluble form have been developed to tackle these problems. Most of these strategies require thoughtful protein engineering to select a fusion tag with particular properties and link this tag to either the N- or C-terminus of the TMP of interest.

This review summarized the progress made in producing several eukaryotic, viral, and prokaryotic TMPs tagged with either MBP, apoA1, or other proteins. Along with providing rationales for how the fusion tag can affect the expression of TMPs, we described several examples of produced and studied TMPs. These successful scientific stories provide unambiguous evidence that the developed methodologies have been instrumental in studies of key physiologically and pharmacologically important TMPs. The progress made so far lays a solid foundation for further advancements in membrane biology to explore the fascinating mechanisms of TMPs from diverse organisms.

Author Contributions: A.O.: data collection, writing the manuscript; E.R.G.: conception, data collection, writing the manuscript, figures, funding. All authors have read and agreed to the published version of the manuscript.

Funding: This manuscript was supported by start-up funds from Department of Chemistry and Biochemistry at TTU (to ERG).

Data Availability: Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflict of interests.

References

1. Cournia, Z.; Allen, T.W.; Andricioaei, I.; Antony, B.; Baum, D.; Brannigan, G.; Buchete, N.-V.; Deckman, J.T.; Delemotte, L.; Del Val, C.; et al. Membrane Protein Structure, Function, and Dynamics: A Perspective from Experiments and Theory. *J. Membr. Biol.* **2015**, *248*, 611–640.
2. Kim, S.; Jeon, T.J.; Oberai, A.; Yang, D.; Schmidt, J.J.; Bowie, J.U. Transmembrane glycine zippers: Physiological and pathological roles in membrane proteins. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14278–14283.
3. Robertson, J.L. The lipid bilayer membrane and its protein constituents. *J. Gen. Physiol.* **2018**, *150*, 1472–1483.
4. Marinko, J.T.; Huang, H.; Penn, W.D.; Capra, J.A.; Schleich, J.P.; Sanders, C.R. Folding and Misfolding of Human Membrane Proteins in Health and Disease: From Single Molecules to Cellular Proteostasis. *Chem. Rev.* **2019**, *119*, 5537–5606.
5. Wallin, E.; von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **1998**, *7*, 1029–1038.
6. Overington, J.P.; Al-Lazikani, B.; Hopkins, A.L. How many drug targets are there? *Nat. Rev. Drug Discov.* **2006**, *5*, 993–996.
7. Yin, H.; Flynn, A.D. Drugging Membrane Protein Interactions. *Annu. Rev. Biomed. Eng.* **2016**, *18*, 51–76.
8. Majeed, S.; Ahmad, A.B.; Sehar, U.; Georgieva, E.R. Lipid Membrane Mimetics in Functional and Structural Studies of Integral Membrane Proteins. *Membranes* **2021**, *11*, 685.
9. Errasti-Murugarren, E.; Bartocioni, P.; Palacin, M. Membrane Protein Stabilization Strategies for Structural and Functional Studies. *Membranes* **2021**, *11*, 155.
10. Hu, J.; Qin, H.; Li, C.; Sharma, M.; Cross, T.A.; Gao, F.P. Structural biology of transmembrane domains: Efficient production and characterization of transmembrane peptides by NMR. *Protein Sci.* **2007**, *16*, 2153–2165.
11. Gulezian, E.; Crivello, C.; Bednenko, J.; Zafra, C.; Zhang, Y.; Colussi, P.; Hussain, S. Membrane protein production and formulation for drug discovery. *Trends Pharmacol. Sci.* **2021**, *42*, 657–674.

12. Kesidis, A.; Depping, P.; Lodé, A.; Vaitsoyopoulos, A.; Bill, R.M.; Goddard, A.D.; Rothnie, A.J. Expression of Eukaryotic Membrane Proteins in Eukaryotic and Prokaryotic Hosts. Available online: https://publications.aston.ac.uk/id/eprint/41666/1/Kesidis_et_al_2020_accepted_version.pdf (accessed on).
13. Bernaudat, F.; Frelet-Barrand, A.; Pochon, N.; Dementin, S.; Hivin, P.; Boutigny, S.; Rioux, J.B.; Salvi, D.; Seigneurin-Berny, D.; Richaud, P.; et al. Heterologous expression of membrane proteins: Choosing the appropriate host. *PLoS ONE* **2011**, *6*, e29191.
14. Rosano, G.L.; Ceccarelli, E.A. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* **2014**, *5*, 172.
15. Hattab, G.; Warschawski, D.E.; Moncoq, K.; Miroux, B. *Escherichia coli* as host for membrane protein structure determination: A global analysis. *Sci. Rep.* **2015**, *5*, 12097.
16. Mathieu, K.; Javed, W.; Vallet, S.; Lesterlin, C.; Candusso, M.P.; Ding, F.; Xu, X.N.; Ebel, C.; Jault, J.-M.; Orelle, C. Functionality of membrane proteins overexpressed and purified from *E. coli* is highly dependent upon the strain. *Sci. Rep.* **2019**, *9*, 2654.
17. Kleiner-Grote, G.R.; Risse, J.M.; Friehs, K. Secretion of recombinant proteins from *E. coli*. *Eng. Life Sci.* **2018**, *18*, 532–550.
18. Francis, D.M.; Page, R. Strategies to optimize protein expression in *E. coli*. *Curr. Protoc. Protein Sci.* **2010**, *61*, 5.24.1–5.24.29.
19. Korepanova, A.; Moore, J.D.; Nguyen, H.B.; Hua, Y.; Cross, T.A.; Gao, F. Expression of membrane proteins from *Mycobacterium tuberculosis* in *Escherichia coli* as fusions with maltose binding protein. *Protein Expr. Purif.* **2007**, *53*, 24–30.
20. Dvir, H.; Choe, S. Bacterial expression of a eukaryotic membrane protein in fusion to various Mistic orthologs. *Protein Expr. Purif.* **2009**, *68*, 28–33.
21. Rogl, H.; Kosemund, K.; Kuhlbrandt, W.; Collinson, I. Refolding of *Escherichia coli* produced membrane protein inclusion bodies immobilised by nickel chelating chromatography. *FEBS Lett.* **1998**, *432*, 21–26.
22. Lee, K.A.; Lee, S.S.; Kim, S.Y.; Choi, A.R.; Lee, J.H.; Jung, K.H. Mistic-fused expression of algal rhodopsins in *Escherichia coli* and its photochemical properties. *Biochim. Biophys. Acta* **2015**, *1850*, 1694–1703.
23. Reuten, R.; Nikodemus, D.; Oliveira, M.B.; Patel, T.R.; Brachvogel, B.; Breloy, I.; Stetefeld, J.; Koch, M. Maltose-binding protein (MBP), a secretion-enhancing tag for mammalian protein expression systems. *PLoS ONE* **2016**, *11*, e0152386.
24. Majeed, S.; Adetuyi, O.; Borbat, P.P.; Islam, M.M.; Ishola, O.; Zhao, B.; Georgieva, E.R. Insights into the oligomeric structure of the HIV-1 Vpu protein. *J. Struct. Biol.* **2023**, *215*, 107943.
25. Majeed, S.; Dang, L.; Islam, M.M.; Ishola, O.; Borbat, P.P.; Ludtke, S.J.; Georgieva, E.R. HIV-1 Vpu protein forms stable oligomers in aqueous solution via its transmembrane domain self-association. *Sci. Rep.* **2023**, *13*, 14691.
26. Mizrahi, D.; Chen, Y.; Liu, J.; Peng, H.M.; Ke, A.; Pollack, L.; Turner, R.J.; Auchus, R.J.; DeLisa, M.P. Making water-soluble integral membrane proteins in vivo using an amphipathic protein fusion strategy. *Nat. Commun.* **2015**, *6*, 6826.
27. Bertin, B.; Freissmuth, M.; Breyer, R.M.; Schütz, W.; Strosberg, A.D.; Marullo, S. Functional expression of the human serotonin 5-HT1A receptor in *Escherichia coli*. Ligand binding properties and interaction with recombinant G protein alpha-subunits. *J. Biol. Chem.* **1992**, *267*, 8200–8206.
28. Grisshammer, R.; Duckworth, R.; Henderson, R. Expression of a rat neurotensin receptor in *Escherichia coli*. *Biochem. J.* **1993**, *295*, 571–576.
29. Grisshammer, R.; Little, J.; Aharony, D. Expression of rat NK-2 (neurokinin A) receptor in *E. coli*. *Recept. Channels* **1994**, *2*, 295–302.
30. Furukawa, H.; Haga, T. Expression of functional M2 muscarinic acetylcholine receptor in *Escherichia coli*. *J. Biochem.* **2000**, *127*, 151–161.
31. Yeliseev, A.A.; Wong, K.K.; Soubias, O.; Gawrisch, K. Expression of human peripheral cannabinoid receptor for structural studies. *Protein Sci.* **2005**, *14*, 2638–2653.
32. Ishola, O.; Ogunbowale, A.; Islam, M.M.; Hadadianpour, E.; Majeed, S.; Adetuyi, O.; Georgieva, E.R. Protein engineering, production, reconstitution in lipid nanoparticles, and initial characterization of the *Mycobacterium tuberculosis* EfpA drug exporter. *bioRxiv* **2023**, 2023.06.26.546575.
33. Eliseev, R.; Alexandrov, A.; Gunter, T. High-yield expression and purification of p18 form of Bax as an MBP-fusion protein. *Protein Expr. Purif.* **2004**, *35*, 206–209.
34. Lei, X.; Ahn, K.; Zhu, L.; Ubarretxena-Belandia, I.; Li, Y.-M. Soluble oligomers of the intramembrane serine protease YqgP are catalytically active in the absence of detergents. *Biochemistry* **2008**, *47*, 11920–11929.
35. Fiermonte, G.; Walker, J.E.; Palmieri, F. Abundant bacterial expression and reconstitution of an intrinsic membrane-transport protein from bovine mitochondria. *Biochem. J.* **1993**, *294 Pt 1*, 293–299.
36. Jappelli, R.; Perrin, M.H.; Lewis, K.A.; Vaughan, J.M.; Tzitzilonis, C.; Rivier, J.E.; Vale, W.W.; Riek, R. Expression and functional characterization of membrane-integrated mammalian corticotropin releasing factor receptors 1 and 2 in *Escherichia coli*. *PLoS ONE* **2014**, *9*, e84013.
37. Tucker, J.; Grisshammer, R. Purification of a rat neurotensin receptor expressed in *Escherichia coli*. *Biochem. J.* **1996**, *317 Pt 3*, 891–899.
38. 333
39. Abiko, L.A.; Rogowski, M.; Gautier, A.; Schertler, G.; Grzesiek, S. Efficient production of a functional G protein-coupled receptor in *E. coli* for structural studies. *J. Biomol. NMR* **2021**, *75*, 25–38.

40. Egloff, P.; Hillenbrand, M.; Klenk, C.; Batyuk, A.; Heine, P.; Balada, S.; Schlinkmann, K.M.; Scott, D.J.; Schutz, M.; Pluckthun, A. Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E655–E662.
41. Wu, F.J.; Williams, L.M.; Abdul-Ridha, A.; Gunatilaka, A.; Vaid, T.M.; Kocan, M.; Whitehead, A.R.; Griffin, M.D.W.; Bathgate, R.A.D.; Scott, D.J.; et al. Probing the correlation between ligand efficacy and conformational diversity at the alpha(1A)-adrenoceptor reveals allosteric coupling of its microswitches. *J. Biol. Chem.* **2020**, *295*, 7404–7417.
42. Schuster, M.; Deluigi, M.; Pantić, M.; Vacca, S.; Baumann, C.; Scott, D.J.; Pluckthun, A.; Zerbe, O. Optimizing the alpha(1B)-adrenergic receptor for solution NMR studies. *Biochim. Biophys. Acta Biomembr.* **2020**, *1862*, 183354.
43. Waugh, D.S. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr. Purif.* **2011**, *80*, 283–293.
44. Chowdhury, A.; Feng, R.; Tong, Q.; Zhang, Y.; Xie, X.-Q. Mistic and TarCF as fusion protein partners for functional expression of the cannabinoid receptor 2 in *Escherichia coli*. *Protein Expr. Purif.* **2012**, *83*, 128–134.
45. Roosild, T.P.; Greenwald, J.; Vega, M.; Castronovo, S.; Riek, R.; Choe, S. NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science* **2005**, *307*, 1317–1321.
46. Dvir, H.; Lundberg, M.E.; Maji, S.K.; Riek, R.; Choe, S. Mistic: Cellular localization, solution behavior, polymerization, and fibril formation. *Protein Sci.* **2009**, *18*, 1564–1570.
47. Marino, J.; Bordag, N.; Keller, S.; Zerbe, O. Mistic's membrane association and its assistance in overexpression of a human GPCR are independent processes. *Protein Sci.* **2015**, *24*, 38–48.
48. Marino, J.; Geertsma, E.R.; Zerbe, O. Topogenesis of heterologously expressed fragments of the human Y4 GPCR. *Biochim. Biophys. Acta* **2012**, *1818*, 3055–3063.
49. Alves, N.S.; Astrinidis, S.A.; Eisenhardt, N.; Sieverding, C.; Redolfi, J.; Lorenz, M.; weberruss, M.; Moreno-Andres, D.; Antonin, W. MISTIC-fusion proteins as antigens for high quality membrane protein antibodies. *Sci. Rep.* **2017**, *7*, 41519.
50. Blain, K.Y.; Kwiatkowski, W.; Choe, S. The functionally active Mistic-fused histidine kinase receptor, EnvZ. *Biochemistry* **2010**, *49*, 9089–9095.
51. Curtiss, L.K.; Valenta, D.T.; Hime, N.J.; Rye, K.A. What is so special about apolipoprotein AI in reverse cholesterol transport? *Arterioscler. Thromb. Vasc. Biol.* **2006**, *26*, 12–19.
52. Denisov, I.G.; Grinkova, Y.V.; Lazarides, A.A.; Sligar, S.G. Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J. Am. Chem. Soc.* **2004**, *126*, 3477–3487.
53. Denisov, I.G.; Sligar, S.G. Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol.* **2016**, *23*, 481–486.
54. Denisov, I.G.; Sligar, S.G. Nanodiscs in Membrane Biochemistry and Biophysics. *Chem. Rev.* **2017**, *117*, 4669–4713.
55. Ryan, R.O.; Forte, T.M.; Oda, M.N. Optimized bacterial expression of human apolipoprotein A-I. *Protein Expr. Purif.* **2003**, *27*, 98–103.
56. Kiefer, H.; Krieger, J.; Olszewski, J.D.; von Heijne, G.; Prestwich, G.D.; Breer, H. Expression of an olfactory receptor in *Escherichia coli*: Purification, reconstitution, and ligand binding. *Biochemistry* **1996**, *35*, 16077–16084.
57. Leviatan, S.; Sawada, K.; Moriyama, Y.; Nelson, N. Combinatorial method for overexpression of membrane proteins in *Escherichia coli*. *J. Biol. Chem.* **2010**, *285*, 23548–23556.
58. Hu, J.; Qin, H.; Gao, F.P.; Cross, T.A. A systematic assessment of mature MBP in membrane protein production: Overexpression, membrane targeting and purification. *Protein Expr. Purif.* **2011**, *80*, 34–40.
59. Chen, G.Q.; Gouaux, J.E. Overexpression of bacterio-opsin in *Escherichia coli* as a water-soluble fusion to maltose binding protein: Efficient regeneration of the fusion protein and selective cleavage with trypsin. *Protein Sci.* **1996**, *5*, 456–467.
60. Duong-Ly, K.C.; Gabelli, S.B. Affinity Purification of a Recombinant Protein Expressed as a Fusion with the Maltose-Binding Protein (MBP) Tag. *Methods Enzymol.* **2015**, *559*, 17–26.
61. Sun, P.; Tropea, J.E.; Waugh, D.S. Enhancing the solubility of recombinant proteins in *Escherichia coli* by using hexahistidine-tagged maltose-binding protein as a fusion partner. In *Heterologous Gene Expression in E. coli: Methods and Protocols*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 259–274.
62. Lebediker, M.; Danieli, T. Purification of proteins fused to maltose-binding protein. In *Protein Chromatography: Methods and Protocols*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 281–293.
63. Nallamsetty, S.; Waugh, D.S. Mutations that alter the equilibrium between open and closed conformations of *Escherichia coli* maltose-binding protein impede its ability to enhance the solubility of passenger proteins. *Biochem. Biophys. Res. Commun.* **2007**, *364*, 639–644.
64. Raran-Kurussi, S.; Waugh, D.S. The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. *PLoS ONE* **2012**, *7*, e49589.
65. Raran-Kurussi, S.; Keefe, K.; Waugh, D.S. Positional effects of fusion partners on the yield and solubility of MBP fusion proteins. *Protein Expr. Purif.* **2015**, *110*, 159–164.
66. Luckey, M. *Membrane Structural Biology: With Biochemical and Biophysical Foundations*; Cambridge University Press: Cambridge, UK, 2014.
67. Smalinskaitė, L.; Hegde, R.S. The biogenesis of multipass membrane proteins. *Cold Spring Harb. Perspect. Biol.* **2023**, *15*, a041251.

68. Brown, M.S.; Ye, J.; Rawson, R.B.; Goldstein, J.L. Regulated intramembrane proteolysis: A control mechanism conserved from bacteria to humans. *Cell* **2000**, *100*, 391–398.
69. Wolfe, M.S.; Kopan, R. Intramembrane proteolysis: Theme and variations. *Science* **2004**, *305*, 1119–1123.
70. Georgieva, E.R. Protein Conformational Dynamics upon Association with the Surfaces of Lipid Membranes and Engineered Nanoparticles: Insights from Electron Paramagnetic Resonance Spectroscopy. *Molecules* **2020**, *25*, 5393.
71. Franke, D.; Svergun, D.I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **2009**, *42*, 342–346.
72. Liu, S.; Li, S.; Krezel, A.M.; Li, W. Stabilization and structure determination of integral membrane proteins by termini restraining. *Nat. Protoc.* **2022**, *17*, 540–565.
73. Andrews, D.A.; Xie, M.; Hughes, V.; Wilce, M.C.; Roujeinikova, A. Design, purification and characterization of a soluble variant of the integral membrane protein MotB for structural studies. *J. R. Soc. Interface* **2013**, *10*, 20120717.
74. Minamino, T.; Imada, K.; Namba, K. Molecular motors of the bacterial flagella. *Curr. Opin. Struct. Biol.* **2008**, *18*, 693–701.
75. Schafmeister, C.E.; Miercke, L.J.; Stroud, R.M. Structure at 2.5 Å of a designed peptide that maintains solubility of membrane proteins. *Science* **1993**, *262*, 734–738.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.