



Article

Generalized Approach towards Secretion-Based Protein Production via Neutralization of Secretion-Preventing Cationic Substrate Residues

Hyunjong Byun ^{1,†} , Jiyeon Park ², Benedict U. Fabia ², Joshua Bingwa ² , Mihn Hieu Nguyen ², Haeshin Lee ^{3,*} and Jung Hoon Ahn ^{1,2,*}

¹ Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Korea; hyunjongbyun@kaist.ac.kr

² Department of Chemistry and Biology, Korea Science Academy of Korea Advanced Institute of Science and Technology, Busan 47162, Korea; jyp131@kaist.ac.kr (J.P.); fabiabenedict@gmail.com (B.U.F.); jbmakhanu10@gmail.com (J.B.); hieuthao1203@gmail.com (M.H.N.)

³ Department of Chemistry, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Korea

* Correspondence: haeshin@kaist.ac.kr (H.L.); hoony@kaist.ac.kr (J.H.A.); Tel.: +82-10-8974-5141 (H.L.); +82-10-8826-6109 (J.H.A.)

† Current Address: School of Medicine, CHA University, Pocheon 11160, Korea.

Abstract: Many heterologous proteins can be secreted by bacterial ATP-binding cassette (ABC) transporters, provided that they are fused with the C-terminal signal sequence, but some proteins are not secretable even though they carry the right signal sequence. The invention of a method to secrete these non-secretable proteins would be valuable both for understanding the secretory physiology of ABC transporters and for industrial applications. Herein, we postulate that cationic “supercharged” regions within the target substrate protein block the secretion by ABC transporters. We also suggest that the secretion of such substrate proteins can be rescued by neutralizing those cationic supercharged regions via structure-preserving point mutageneses. Surface-protruding, non-structural cationic amino acids within the cationic supercharged regions were replaced by anionic or neutral hydrophilic amino acids, reducing the cationic charge density. The examples of rescued secretions we provide include the spike protein of SARS-CoV-2, glutathione-S-transferase, streptavidin, lipase, tyrosinase, cutinase, growth factors, etc. In summary, our study provides a method to predict the secretability and a tool to rescue the secretion by correcting the secretion-blocking regions, making a significant step in understanding the physiological properties of ABC transporter-dependent protein secretion and laying the foundation for the development of a secretion-based protein-producing platform.

Keywords: ABC transporter; type I secretion system (T1SS); protein secretion; recombinant protein expression; protein mutagenesis



Citation: Byun, H.; Park, J.; Fabia, B.U.; Bingwa, J.; Nguyen, M.H.; Lee, H.; Ahn, J.H. Generalized Approach towards Secretion-Based Protein Production via Neutralization of Secretion-Preventing Cationic Substrate Residues. *Int. J. Mol. Sci.* **2022**, *23*, 6700. <https://doi.org/10.3390/ijms23126700>

Academic Editor:
Giuliano Ciarimboli

Received: 12 May 2022

Accepted: 13 June 2022

Published: 15 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Most laboratories in academia and biotechnology companies utilize non-secretory *Escherichia coli* systems for protein preparations [1]. In such methods, *E. coli* transformed with an expression vector is cultured and lysed, and during this lysis a protein of interest is unavoidably mixed with other contaminant proteins (see Figure 1a). Thus, in these conventional systems, a target protein can only be prepared via multi-steps of column chromatography, salting out, and dialysis [2]. Broad ranges of proteins are prepared by this *E. coli* strategy, with examples including most laboratory enzymes (e.g., lipases), cell-signaling proteins (e.g., fibroblast growth factor, bone morphogenic proteins), fluorescent proteins (e.g., green fluorescent protein), tag proteins (e.g., streptavidin and thrombin), and countless others.

secretion-based protein preparation methods. The centrifuged culture supernatant is relatively devoid of contaminants, reducing the complexity and cost of the purification process. (c) An illustration of ABC transporter-mediated protein secretion. Some proteins of interest were not compatible with the secretion-based preparation methods, possibly because of the “cationic supercharged” region within their sequences. The inside-negative membrane potential makes it energetically unfavorable to secrete cationic residues. IM: inner membrane. (d) The cationic supercharged regions can be identified by analyzing the local charge density of linearized protein, or “linear charge density”. (e) Linear charge density (LCD) of a protein at a certain residue (say, residue i) is calculated by summing the average charge at pH 7 of the side chains from residue i to residue $i + 19$. This block of 20 residues is labeled “window # i ”. We repeated this from $i = 1$ to $i = (n - 19)$, where n was the length of the polypeptide. (f,g) LCD graphs of two typical non-secretable proteins, NKC-TliA, CTP-TliA, M37 lipase, and SARS-CoV-2 spike protein. Note that all of them have high peaks, which represent regions crowded with positively charged amino acids (“cationic supercharged” regions).

The *E. coli* protein preparation system almost guarantees to produce functional proteins of interest, unless they require a large degree of post-translational modifications. This preparation fidelity is the primary reason for widespread usages of *E. coli* systems in most laboratories [3]. However, multiple, labor-intensive, and time-consuming purification steps accompanying cell lysis is a main disadvantage.

One solution to overcome this long-time, un-addressed challenging task is secretion of a target protein into extracellular space (see Figure 1b). In these bacterial secretion-based production schemes, the host microorganism secretes the target protein to the medium so that the other cellular proteins remain in the interior of the cell. In this case, intensive purification steps are no longer necessary, greatly saving time, materials, and human resources [4,5]. *Pseudomonas fluorescens* is a good microorganism for this aim [3]. The host is a generally non-pathogenic bacterium [6] with a good tolerance to high-density cell culture [7,8]. It carries an ATP-binding cassette (ABC) transporter [9], which is a type I secretion system (T1SS) capable of secreting many heterologous proteins, such as epidermal growth factor [10], green fluorescent protein (GFP) [11], alkaline phosphatase [12], phospholipase A1 [13], and maltose-binding protein [14].

However, current bacterial secretion systems, including the ABC transporters, are protein-dependent, which means that not all proteins are able to be secreted even though the right signal is attached [5]. Due to this reason, most researchers (even at this time) still use the reliable *E. coli* system [3]. If the criteria separating the secretable and non-secretable cargo protein are identified, then researchers and industry workers can predict whether a specific cargo protein can be prepared via secretion-based methods, even before experimentally testing it. This would greatly reduce the time and personnel costs.

During brief examinations, we noticed that the non-secretable proteins almost always had high-density cationic supercharged regions, or “cationic supercharged regions”. We suspected that this might be the cause of the non-secretability of these proteins, as illustrated in Figure 1c. We assumed that we could identify those cationic supercharged regions by examining the charge density distribution of a linearized polypeptide (see Figure 1d,e), and drawing them as graphs. With these informative graphs (such as Figure 1f,g), we noticed that all non-secretable proteins indeed contained at least one cationic supercharged region.

Herein, we present a step forward to developing a reliable secretion-based protein production method, by suggesting a method to predict whether a secretion-based production is possible before actual experiments, and by providing tools to rescue the non-secretable proteins. We experimentally demonstrate the secretion-blocking property of cationic supercharged regions by adding artificial cationic supercharged regions to secretable proteins. We also identify the secretion-impeding cationic supercharged regions within a few known non-secretable proteins. In detail, we scan for these cationic supercharged regions by examining local charge densities of a linearized amino acid sequence of a target protein, or “linear charge density (LCD)”. In this LCD analysis, a cationic supercharged window is de-

defined as when summation of overall charges in a 20-amino acid box (a “window”) exceeds +2. We could repeat this calculation for all possible windows throughout the entire amino acid sequence of a protein (see Figure 1e again). This process is automated by a computer code created by us. We demonstrate that regions where the LCD values are higher than +2 (i.e., cationic supercharged regions) inhibit the secretion of a protein, exemplifying this in several proteins, including the SARS-CoV-2 receptor binding domain. After the identification of cationic supercharged regions, the corresponding cationic amino acids are mutated into negatively charged or neutral amino acids to reduce the windows’ overall charge to $\leq +2$. Finally, we demonstrate secretion-enhancing strategies by rescuing secretions of non-secretable, yet valuable proteins, such as glutathione S-transferase, streptavidin, cutinase, MelC2 tyrosinase, M37 lipase, TNF β , and the spike protein of SARS-CoV-2, etc.

2. Results

2.1. Conversion of SARS-CoV-2 Spike Protein into Secretable Variants

The spike protein of SARS-CoV-2 is a vaccine candidate, which underlines the importance of its preparation considering the current COVID-19 pandemic and potential future coronavirus outbreaks. As previously shown in Figure 1g, nine cationic supercharged regions which can potentially inhibit its secretion were identified by LCD analysis. Two cationic supercharged regions were identified in both the N-terminal domain (NTD) and the receptor binding domain (RBD) (see Figure 2a). Due to the presence of the supercharged regions (see Figure 2d,e, top row), the unmutated proteins of NTD and RBD were not secreted (see Figure 2f,g, 1st and 4th lanes). Thus, we designed mutants of these proteins to eliminate the aforementioned cationic supercharged regions, by replacing cationic lysine or arginine into anionic aspartate or glutamate. In this way, overall charges of each window (20 amino acid) were suppressed below +2 (see Figure 2b,c, green lines) or +1 (see Figure 2b,c, red lines). Western blot results clearly showed that each of the $\leq +2$ mutants of NTD and RBD were secreted into extracellular medium, and the secretions were further enhanced in the $\leq +1$ mutants (see Figure 2f,g).

2.2. Introducing Artificial “Cationic Supercharged Regions” to Substrate Proteins

The aforementioned hypothesis, supported by a clear representative demonstration using SARS-CoV-2 spike protein, suggested that the cationic supercharge (over +2) can be a determinant factor for protein secretion with the ABC transporter. If the hypothesis were to be true, insertion of oligopeptide sequences consisting of cationic amino acids in the middle of a well-secreted substrate protein would block its secretion. To test this, we added the following four oligopeptides at C-terminal sides of the cargo proteins, just upstream of the LARD3 signal sequence: Arg05 (five consecutive arginine residues), Arg10, Lys05, and Lys10 (see Figure 3a–e for the structures). We used green fluorescent protein (GFP) as a model substrate. In Figure 3f,g, note that GFP was well-secreted before the insertion of these amino acid patches. However, after the attachment of Arg05, Arg10, and Lys10 sequences, their secretions stopped. The linear charge density “peaks” that these oligopeptides introduced are illustrated in Appendix A Figure A1a–e. Meanwhile, Lys05 sequence seemed to have little effect on GFP secretion. This suggests that lysine exhibits a much weaker secretion-blocking effect than that of arginine. In addition, we performed identical experiments with two different model cargo proteins with better secretion than GFP, namely the *P. fluorescens* ABC transporter’s native substrate TliA or the negatively charged mutant of GFP named GFP(–30) [14]. The results were similar to that of GFP: Arg05, Arg10, and Lys10 blocked the secretion, while Lys05 had little effect (see Appendix A Figure A2a–d).

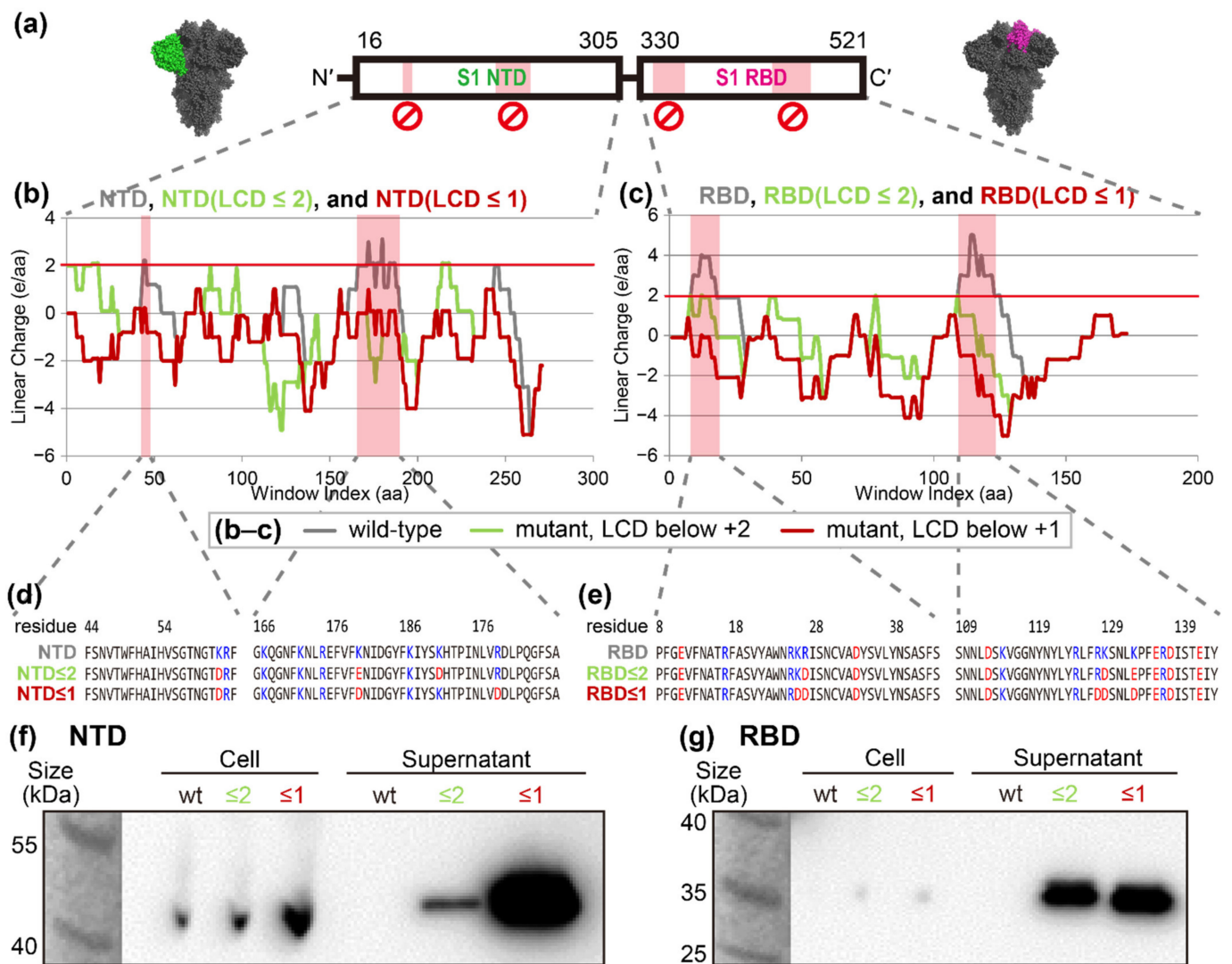


Figure 2. The secretion of SARS-CoV-2 spike proteins and its mutants without cationic supercharged regions. (a) The SARS-CoV-2 spike protein homotrimer (PDB ID 6zp0) S1 has two domains: N-terminal domain (NTD, dark green) and receptor binding domain (RBD, pink). (b,c) Linear charge density (LCD) graphs of native and mutant forms of NTD and RBD (gray lines). The cationic supercharged windows of NTD and RBD are indicated by the red backgrounds. Two mutants, NTD (LCD ≤ 2) and RBD (LCD ≤ 2), were mutated to keep the heights of LCD graphs lower than +2 (green lines). NTD (LCD ≤ 1) and RBD (LCD ≤ 1) were mutated to keep the heights of LCD graphs lower than +1 (red lines). (d,e) The amino acid sequences of cationic supercharged regions marked in panels (b,c). (f,g) The NTD, RBD, and their derivatives were expressed in *P. fluorescens* harboring the ABC transporter. Both cell pellet and supernatant samples were taken and analyzed via Western blot. Darker signal in the supernatant samples implies more secretion.

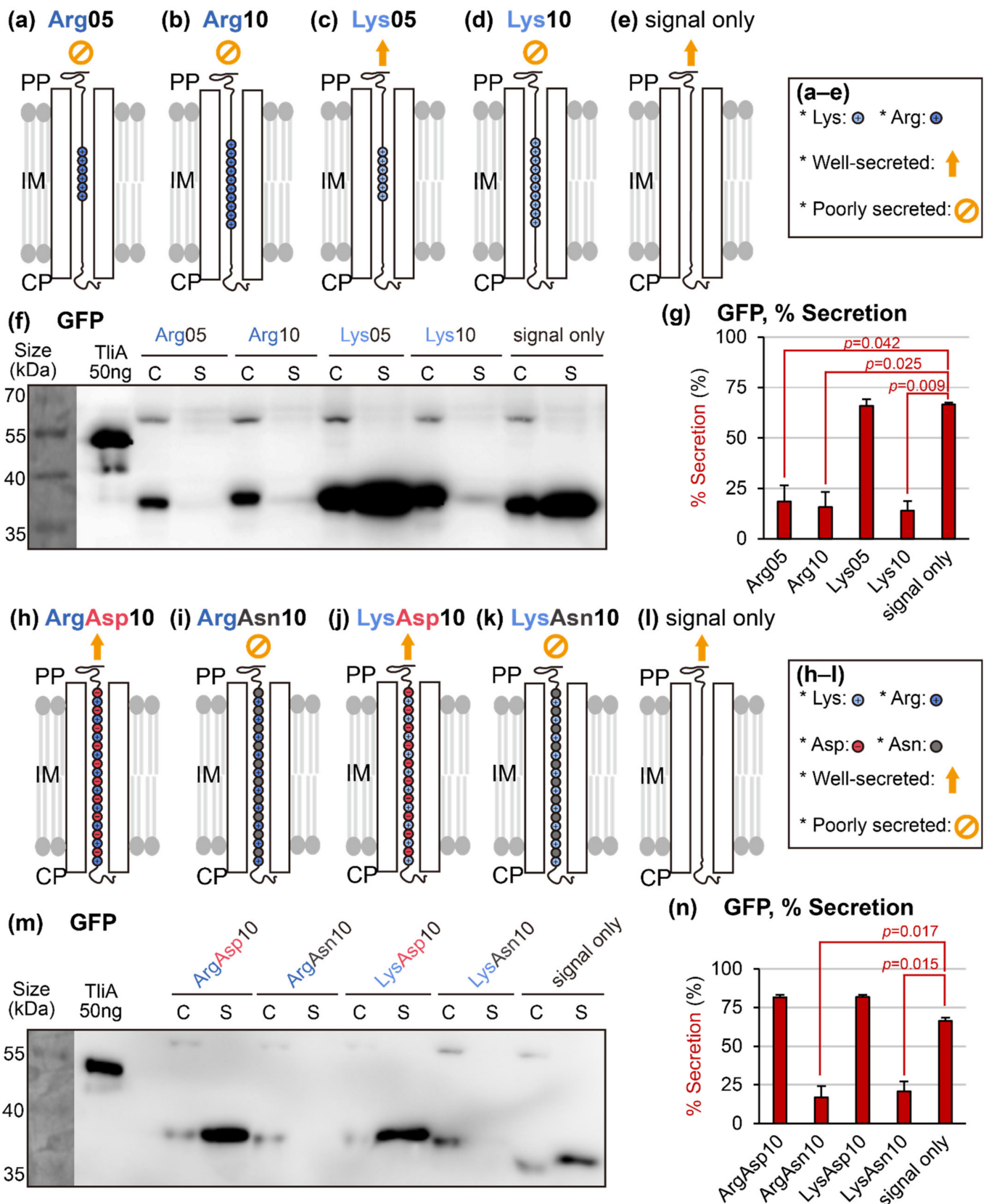


Figure 3. Investigation of secretable conditions by ABC transporters. (a–e) Structures of the constructed artificial cationic supercharged regions consisting of consecutive cationic amino acids, Arg05, Arg10, Lys05, and Lys10. Yellow arrows indicate that the protein was secreted, while yellow

cancellation signs indicate the opposite. (f) Western blot analysis of the artificial cationic supercharged regions. The four artificial cationic supercharged regions were inserted into green fluorescent protein (GFP), which is intrinsically secretable. The cell pellet samples are labeled “C”, and the supernatant samples are labeled “S”. Signal in an S sample indicates that the protein was secreted. (g) Quantification of experiment presented in panel (f). Error bars represent sample standard errors. Performed on biological duplicates, $n = 3$. (h–l) Structures of artificial peptide sequences used for the “rescue of secretion” experiment, where peptides consisting of alternating cationic–anionic or cationic–neutral amino acids were compared to each other. (m) Western blot analysis of the secretion-rescuing experiment using adjacent negative charges. (n) Quantification of experiment presented in panel (m). Compared to the “signal only” control group. Error bars represent sample standard errors. Performed on biological duplicates, $n = 3$.

2.3. Rescuing Secretion by Adding Negative Charges near the Positive Charges

We immediately questioned if the secretion-blocking effect of the arginine and lysine oligopeptides could be rescued by proximal anionic charges. We attached the following four sequences to the C-terminus of the model protein GFP: ArgAsp10 (ten repeats of arginine–aspartic acid pairs), ArgAsn10 (ten repeats of arginine–asparagine pairs), LysAsp10 (ten repeats of lysine–aspartic acid pairs), and LysAsn10 (ten repeats of lysine–asparagine pairs) (see Figure 3h–l for the structures, and Appendix A Figure A1f–i for the LCD graphs). Indeed, ArgAsn10 and LysAsn10 blocked the secretion, just like the Arg10 and Lys10 sequences tested in Figure 3a. On the other hand, both ArgAsp10 and LysAsp10 did not harm the secretion, indicating that the adjacent aspartic acid residues were able to neutralize the secretion-blocking effect of the arginine and lysine residues. Furthermore, we performed identical experiments with two additional substrate proteins, TliA and GFP(–30). Both of these additional models yielded identical results (see Appendix A Figure A2e–h).

2.4. “Negatively Supercharging” Poorly Secreted Proteins

So far, we have demonstrated that the existence of a cationic supercharged region within the protein sequence blocks the secretion, which was rescued by adding negative charges nearby. We tested if this secretion-rescuing strategy can be generally applied for other various substrates. Our first plan was to mutate the target proteins by providing a significant amount of negative charges, because there is already an existing algorithm to perform this mass addition of negatively charged residues [15]. Using this method, surface-exposed cationic residues were selected as target residues for replacement with anionic amino acids. We first demonstrated that this strategy indeed works for glutathione S-transferase (GST) and streptavidin (SAv), which are intrinsically non-secretable (see Figure 4a, black lanes) due to the existence of cationic supercharged regions. Conveniently, “negatively supercharged” mutant proteins of these two were already previously reported as GST(–20) and SAv(–10) [15]. Thus, we directly proceeded to compare the secretion of the wild-types and these pre-made variants. Indeed, the negatively supercharged mutants exhibited superior secretion compared to the native ones (see Figure 4a, red lanes). The surface charges reversed by negative supercharging in the previous report lowered the peaks in the LCD graphs of GST and SAv (see Appendix A Figure A3c,d).

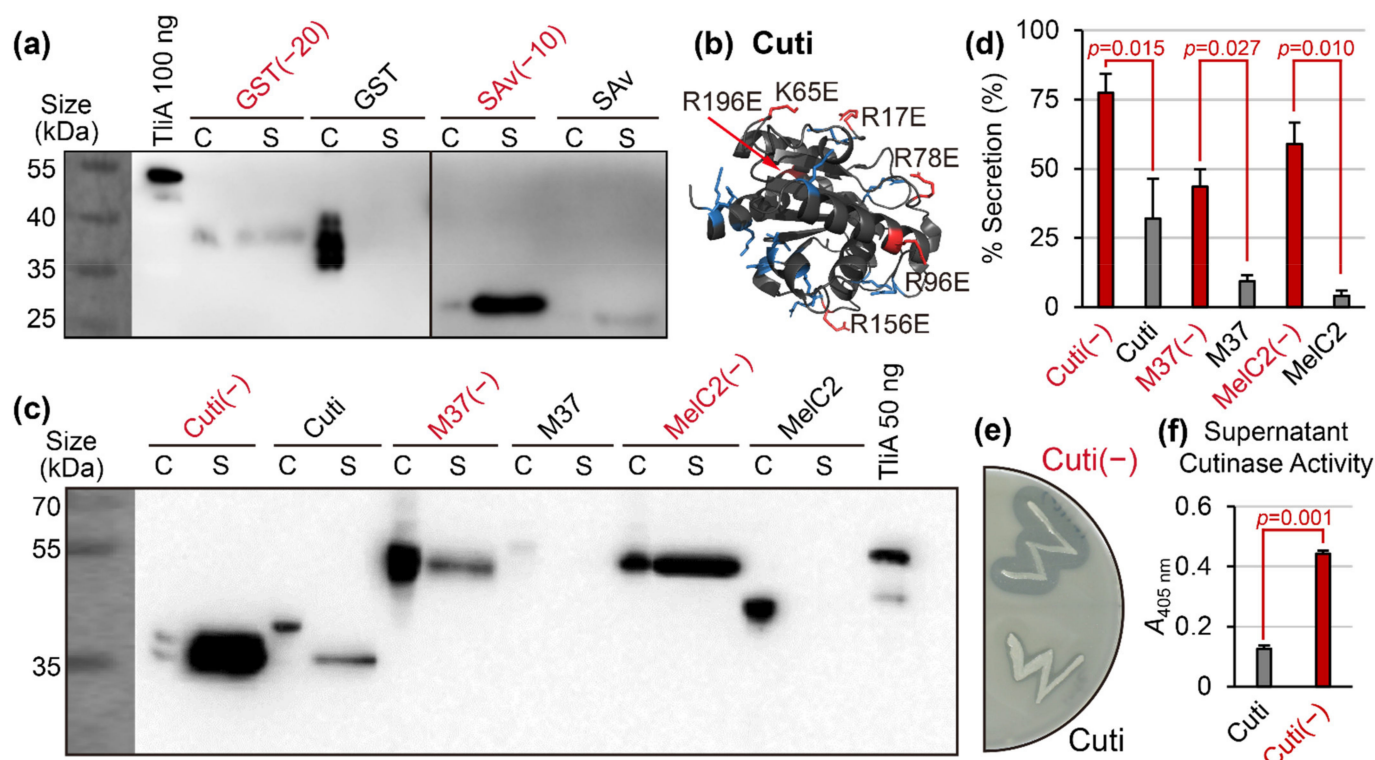


Figure 4. Secretion of negatively supercharged proteins. The liquid cultures of *P. fluorescens* were grown in TB for two days at 25 °C. The cell pellet samples are labeled “C” while the supernatant samples are labeled “S”. A signal in “S” indicates that the sample was secreted. (a) Secretion enhancement in negatively supercharged versions of glutathione S-transferase (GST) and streptavidin (SAv). The negatively supercharged mutants were secreted while the wild-types were not. (b) The structure of *N. haematococca* cutinase (Cuti) (PDB ID 1cex), used for the supercharging mutation. Red color indicates the surface-exposed cationic residues with AvNAPSA value ≤ 100 , which were replaced by anionic glutamates. Blue color indicates the rest of the cationic residues. (c) Secretion enhancement by negatively supercharged enzymes. Negatively supercharged mutants of Cuti(-), M37(-), and MelC2(-) were compared with wild-type cutinase, M37 lipase, and MelC2 tyrosinase for secretion using Western blot. Note that Cuti(-) C lane has two bands, probably corresponding to oxidized and reduced forms. (d) Quantification of Western blot experiment presented in panel (c). All three of the negatively supercharged mutants (red colors) exhibited a significant increase in secretion. Error bars represent sample standard errors. Performed on biological duplicates, $n = 3$. (e) Plate enzymatic activity assay of Cuti(-) and Cuti. The negatively supercharged mutant Cuti(-) colony exhibits a much larger halo, due to the increased amount of secreted enzyme. (f) Quantification of results in panel (e). Cell culture supernatant cutinase activity was measured with the pNPB assay.

Using the same algorithm used to create GST(-20) and SAv(-10), we were able to create a negatively supercharged variant of *Nectria haematococca* cutinase (Cuti), *Photobacterium lipolyticum* M37 lipase (M37), and *Streptomyces antibioticus* MelC2 tyrosinase (MelC2). The method was to replace solvent-exposed cationic residues into anionic amino acids, as exemplified in Figure 4b. The results were promising: even though the native Cuti, M37, and MelC2 were poorly secreted, the mutated Cuti(-), M37(-), and MelC2(-) became well-secreted (see Figure 4c,d). Notably, the mutant Cuti(-) retained its intrinsic enzymatic activity, and with the increased secretion, Cuti(-) culture’s extracellular cutinase activity increased (see Figure 4e,f). The LCD graph analysis confirmed that the charge-flooding supercharging mutations completely removed the cationic supercharged regions from cutinase, MelC2 tyrosinase, and M37 lipase (see Appendix A Figure A3e–g).

Overall, these results indicated that it is possible to rescue the secretion by flooding the target protein with anionic amino acids, i.e., negatively supercharging them.

2.5. “Super-Neutralization”: An Alternative Secretion-Rescuing with Less Structural Disturbance

Although the negative supercharging approach for target proteins made them secretible, reversals of the surface cationic charges into anionic charges in multiple residues can potentially cause elimination of intrinsic protein functions. The intrinsic enzymatic activity was retained for the previous case of cutinase (see Figure 4e), but M37 lipase lost the enzymatic activity by negative supercharging (as detailed below). Supercharging processes often involve a reversal of the charge from cationic to anionic, which can lead to a loss of protein function. In contrast, changing cationic residues into neutral hydrophilic ones would likely cause a decreased level of structural perturbation. To implement this idea, we replaced solvent-exposed cationic residues of M37 lipase into neutral hydrophilic glutamine (single letter code, Q). This mutant was named M37(Q). Likewise, MelC2(Q) was prepared from *S. antibioticus* MelC2 tyrosinase. We termed this approach “super-neutralization”. The result showed successful secretion of M37(Q) and MelC2(Q) (see Figure 5a,b, orange lanes), and importantly, the mutated M37(Q) retained the enzymatic activity (see Figure 5c). The LCD analyses showed that the super-neutralization processes removed the cationic supercharged regions in both M37(Q) (see Figure 5d) and MelC2(Q) (see Figure 5e). The super-neutralized mutants demonstrated that replacing cationic residues with the neutral hydrophilic amino acid, glutamine (Q), can also rescue the secretion and, in addition, has less effects on enzymatic activity than the negative supercharging approach.

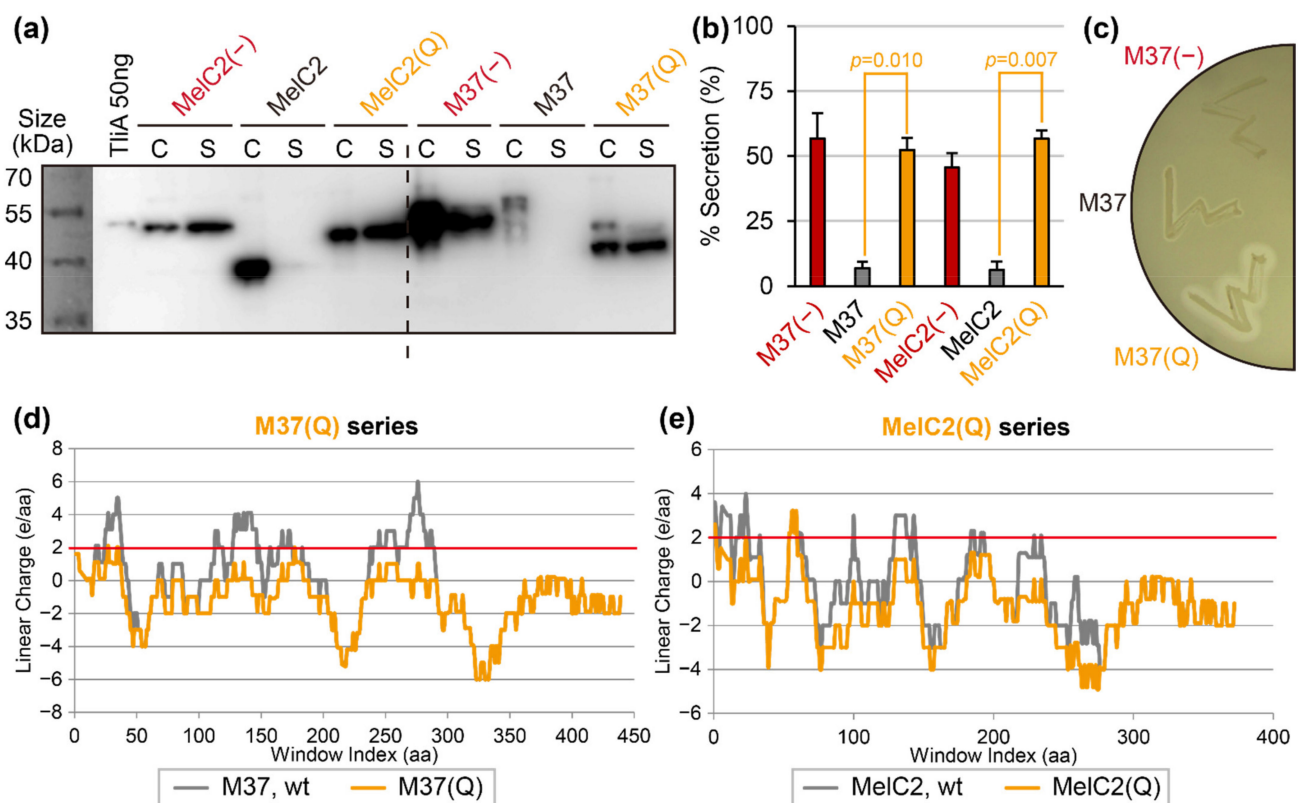


Figure 5. Secretion of proteins which were mildly mutated via super-neutralization. (a) The “super-neutralized” mutants, M37(Q) and MelC2(Q), were tested for secretion via Western blot. A signal in “S” implies secretion. (b) Quantification of results presented in panel (a). Error bars represent sample standard errors. (c) Enzymatic activity of M37 lipase and its derivatives developed in this study. A halo indicates that the protein had both been secreted and had enzymatic activity. (d) LCD graphs of M37(Q) and native M37 lipase. (e) LCD graphs of MelC2(Q) compared to native MelC2.

2.6. Random Mutation and Activity Screening: Avoiding Structure-Disrupting Mutations

From the above experiments, we verified that mutating all the solvent-exposed cationic residues can rescue the secretion of non-secretable proteins. However, this approach unavoidably results in excessive, unnecessary mutations of a given poorly secreted protein. We postulated that we can selectively collect secretion-promoting mutations by ruling out any activity-disrupting mutations. This approach might be similar to a screening process in a directed evolution. We used the mixed-based gene synthesis technique to insert either lysine or glutamate randomly at the solvent-exposed lysine residues. Then, we screened for the clone that exhibited the most prominent halo in the lipase activity assay. The resulting clone, M37(var)—“var” from “variable mutation”, was then characterized via DNA sequencing. M37(var) had both excellent secretion (see Figure 6a, orange lanes) and retained enzymatic activity (see Figure 6b). In fact, the culture supernatant enzymatic activity in Figure 6b actually increased compared to the wild-type, possibly due to the increased amount of enzyme secreted to the culture supernatant. The LCD graph of the M37(var) protein is presented in Appendix A Figure A3h. The SDS-PAGE gel staining experiment in Figure 6c shows that the culture supernatant of the M37(var) mutant mostly consists of M37 lipase enzyme, clearly demonstrating the concept of relatively contaminant protein-free supernatant in secretion-based protein production.

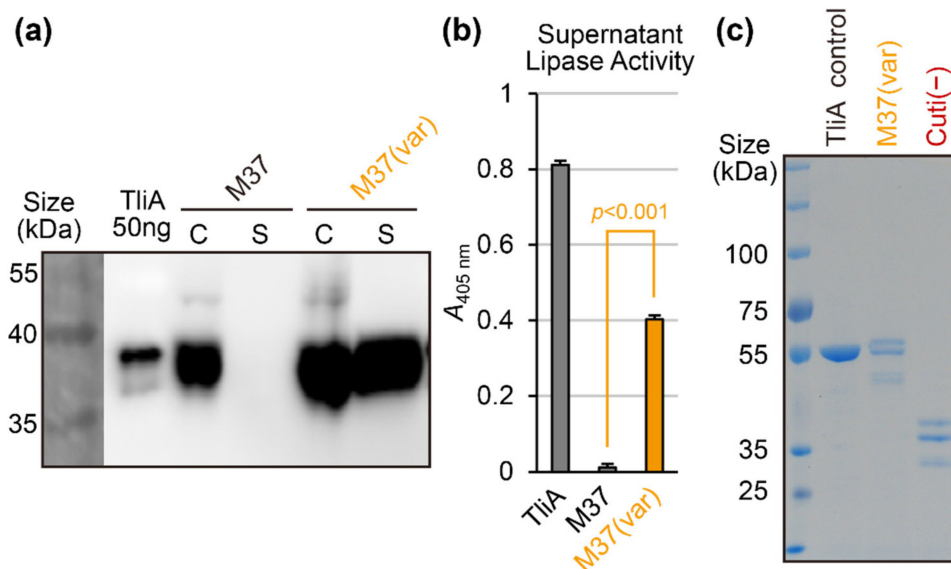


Figure 6. The secretion and supernatant activity of M37(var) lipase and Cuti(–). (a) A mutant clone of M37 obtained from random mutation and activity-screening strategy, M37(var), was tested for secretion. (b) The activity of the culture supernatant was measured for lipase/cutinase activity of three different enzymes in the supernatant. The activity of TliA and M37 lipase was measured using pNPP and the activity of cutinase using pNPB. (c) SDS-PAGE of culture supernatants. 16 μL of culture supernatant was loaded on SDS-PAGE and stained with Coomassie blue. The culture supernatant consisted only of the respective target proteins: M37(var) and Cuti(–). Note that each of these proteins were separated into several bands corresponding to oxidized and reduced forms of LARD3-conjugated proteins (upper two bands) and lower LARD3-cleaved proteins.

2.7. Linear Charge Density-Focused Rational Mutant Design

All the above experiments targeted highly solvent-exposed residues in the protein, but not necessarily all those residues were in cationic supercharged regions. This means that the above methods might involve unnecessary mutations at positions that were already secreted well. Applying the novel charge density hypothesis, we aimed to pin-point cationic supercharged regions within the target proteins and only modify the solvent-exposed cationic residues within those regions. This way, any unnecessary mutations could be avoided. As can be seen from our proof-of-concept case of SARS-CoV-2 spike

protein (see Figure 2), we arbitrarily defined cationic supercharged regions as peaks in the LCD graph with height above +2. Relatively smaller-sized, non-secretable *Homo sapiens* signaling proteins, namely transforming growth factor beta (TGF β), tumor necrosis factor beta (TNF β), and fibroblast growth factor I (FGF1), were chosen as model proteins. The LCD graph before and after the mutations revealed that we precisely removed the cationic supercharged regions and left the rest of the protein unmodified (see Figure 7a–c). The secretion patterns of the mutants showed that all of the supercharged mutants, TGF β (–), TNF β (–), and FGF1(–), were secreted well (see Figure 7d,e, red lanes), while the native versions were not (see Figure 7d,e, black lanes).

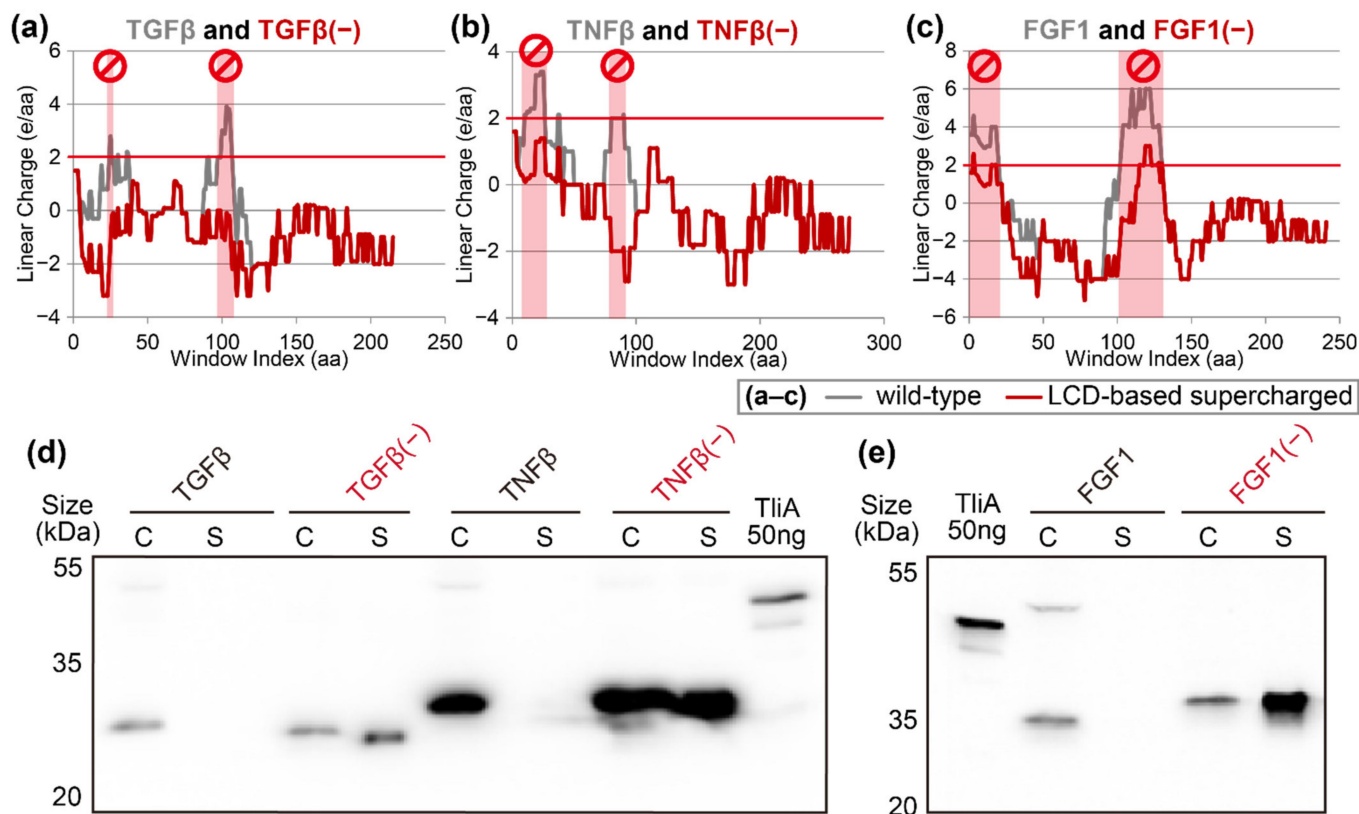


Figure 7. Linear charge density-focused supercharging significantly improves the secretion with minimal mutations. (a–c) LCD graphs of TGF β , TNF β , and FGF1 and their respective mutants. These mutants were designed by identifying >+2 peaks (red background) in the LCD graphs of native proteins (solid gray lines) and mutating positively charged solvent-exposed residues within those peaks. We only mutated residues with low ConSurf Bayesian conservation scores (<4). We also looked upon the homologs of the query protein and saw which amino acids were frequently occurring in that position, to select the replacement among anionic amino acids and neutral hydrophilic amino acids. Highly conserved amino acids (conservation score > 6) remained unmutated in all situations. In addition, the amino acids located in the ligand-receptor binding sites were not mutated. (d,e) The resulting mutant proteins from panels (a–c) were introduced into *P. fluorescens* cells and were tested via Western blot.

2.8. Other ABC Transporters Exhibited Similar Behaviors

To verify that the substrate charge density dependency was not limited to the *P. fluorescens* ABC transporter, we tested the secretion of negatively supercharged cargo proteins through a few different polypeptide-secreting ABC transporters. The complexes we tested were as follows: *Pseudomonas aeruginosa* AprDEF, *Dickeya dadantii* (also known as *Erwinia chrysanthemi*) PrtDEF, and *E. coli* HlyBD + TolC. We had already verified that the LARD3 signal sequence we used in this experiment was compatible with these transporters (see

Appendix A Figure A4). The sequence identities of these ABC transporters to *PfTliD* (the one we used in the other experiments of this study) were as follows: *PaAprD* 57.47%, *DdPrtD* 55.77%, and *EcHlyB* 23.64%. We chose the cutinase (Cuti) as the model protein because we could easily detect its secretion in the plate activity assay. We prepared double-transformed *E. coli* cells harboring the genes for the target proteins as well as the ABC transporter genes. Colonies of these cells were tested for secretion via the plate enzymatic activity assay (see Figure 8a,b) and Western blot (see Figure 8c,d). We also tested the supercharged protein M37(−) and wild-type M37 in Western blot and showed similar results (data not shown). The results showed that negatively supercharging the cargo proteins also improved the secretion in all three of the other bacterial ABC transporters.

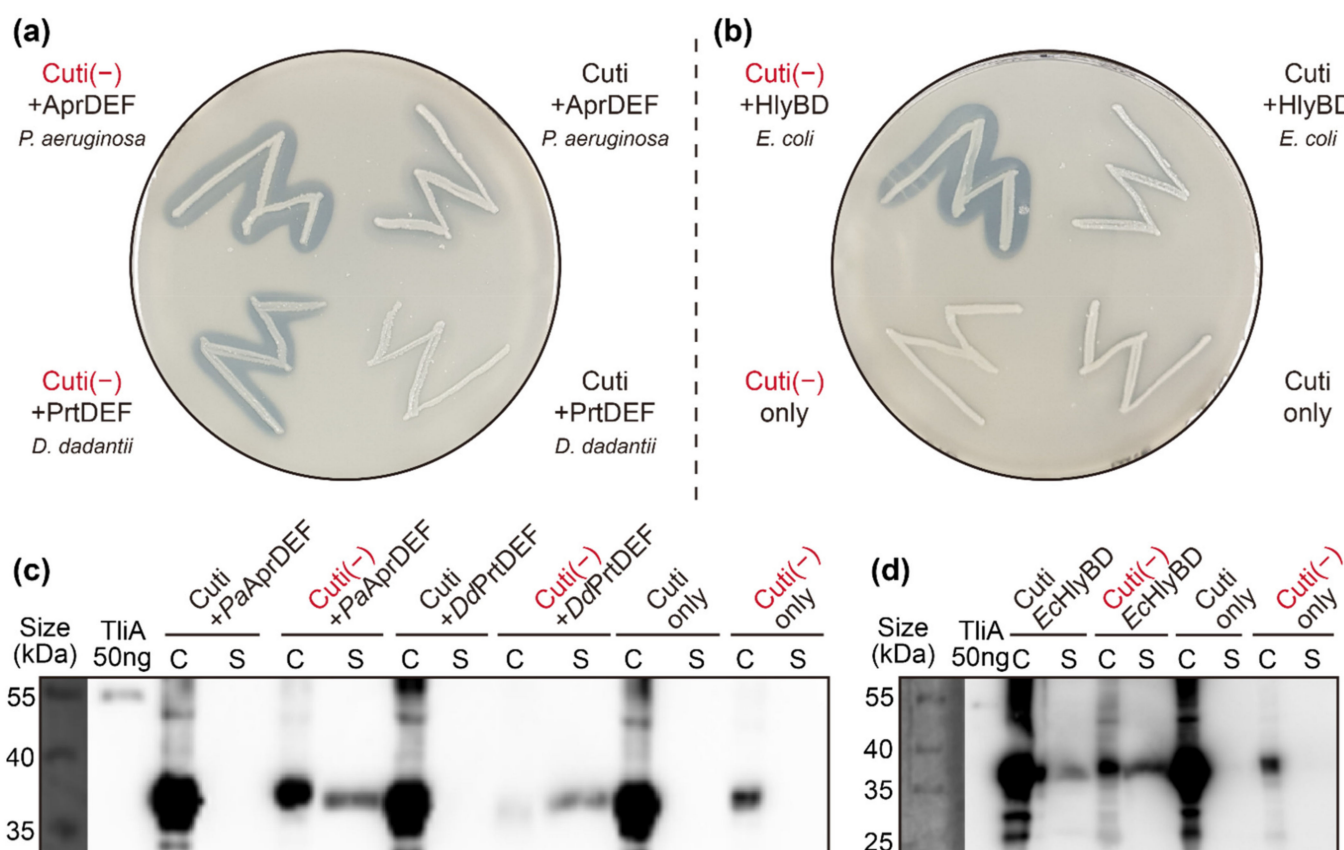


Figure 8. The effect of cargo protein supercharging in the other ABC transporters. *E. coli* XL1-Blue was double-transformed with two plasmids, one expressing cutinase variants and the other expressing the polypeptide-secreting ABC transporter. Empty vector without ABC transporter was used as the negative control. The colonies were screened from LB agar plates containing 50 µg/mL of carbenicillin and 100 µg/mL of chloramphenicol, then streaked on LB agar plates containing 0.5% colloidal tributyrin. (a,b) The three different ABC transporters, *PaAprDEF*, *DdPrtDEF*, and *EcHlyBD* + TolC complexes, successfully secreted the negatively supercharged cutinase (Cuti(−)). They secreted much less wild-type cutinase (Cuti). This could be clearly seen from the halo size comparisons. (c,d) The supercharged and wild-type versions of cutinase were tested for secretion via the three different polypeptide-secreting ABC transporters. In these Western blots, both the cell pellet (labeled “C”) and the culture supernatant (labeled “S”) samples were analyzed in parallel.

3. Discussion

In this study, we aimed to identify the exact conditions of heterologous cargo proteins that could be secreted by bacterial ABC transporters. We believe that our results disclosed that the “absence of a cationic supercharged region” is the determinant factor that converts poorly secreted proteins to well-secreted ones. Our theory was able to explain all results

from previous studies as well. For example, eliminating cationic supercharged regions along the amino acid sequences results in shifting the proteins' overall charges down to anionic states, which might explain why previously known natural substrates of ABC transporters (which are secreted) tend to be negatively charged [16].

Previous statements to predict the secretability of a cargo protein have been the following:

Hypothesis 1 (H1). *Proteins that are highly negatively charged overall are secreted, otherwise, the secretion is not allowed [14,16].*

However, our study proposes the following revised general statement that can predict secretion properties of a given protein:

Hypothesis 2 (H2). *Any region with a high cationic local charge density within a cargo protein inhibits the secretion. Otherwise, the secretion is allowed.*

According to Hypothesis 1, CTP-TliA and NKC-TliA were categorized as secretable proteins, but they were in fact non-secretable ones [14]. However, our new Hypothesis 2 can properly categorize CTP-TliA and NKC-TliA as non-secretable proteins as both of these proteins carry cationic supercharged regions (see Figure 1f). The novel "absence of a cationic supercharged region" hypothesis is also consistent with the fact that non-secretable proteins cannot be rescued by simply fusing highly anionic proteins at the polypeptide termini (see Appendix A Figure A5). Although these fusions might add some more anionic charges overall, they would not necessarily suppress the cationic supercharged regions within the proteins of interest.

We had briefly discussed the presumptive role of the membrane potential in our previous publication [14]. Back then, we had too little evidence to explicitly state that the linear charge density of the unfolded cargo protein has a crucial role in the transport. With the evidence in this study, we have enough evidence that the positively supercharged regions within the target protein block the transportation through bacterial ABC transporters. When the cationic supercharged regions of substrates are traversing through the transporter channel [17–19], the inside-negative membrane potential applies inward electrostatic force to the polypeptide. When the electrostatic force from the repulsion is larger than the driving force, the polypeptide cannot transverse through the channel.

In Figure 3a, we performed a comparative experiment in which Arg05 clearly blocked GFP secretion, but Lys05 did not. We assume that this is related to the side-chain chemistry of these amino acids. In a solvent-exposed state, the conjugate acid pK_a value of arginine is 12.5 and the conjugate acid pK_a of lysine is 10.5 [20]. These values are much higher than the physiological cytoplasmic pH of bacteria, which is generally 7.2–7.8 [21,22], including ~7.7 for *Pseudomonas* species [23]. This means that lysine and arginine both exist in protonated states in the cytoplasm. Nonetheless, low-dielectric constant environments, such as inside the phospholipid bilayer membranes or inside proteins, often lower the conjugate acid pK_a value of basic amino acids [24]. This suggests that the hydrophobic, low-dielectric constant microenvironment inside the transporter channel might allow the deprotonation of lysine, while failing to deprotonate arginine, which have a much higher pK_a .

In fact, other transporters in the Sec transporter family exhibit a similar pattern, where cationic residues are not easily translocated [25,26], and among those cationic residues, arginine is much harder to secrete than lysine [27–29]. This could be indirect evidence that the polypeptide-secreting ABC transporters may share a similar fundamental mechanism with the Sec transporters. Both Sec transporters and ABC transporters: (i) involve the configurational changes accompanied by the ATP hydrolysis, (ii) have to "push" the polypeptides through a narrow channel, which allows no complicated structures except for the alpha helices to pass through, and (iii) primarily secrete acidic (or negatively charged) proteins. There was an attempt to explain this behavior as a "proton ratchet" effect in Sec transporters [29], which was an application of the Brownian ratchet mechanism of secretion

modes [30,31]. We propose that this theory might also be valid for polypeptide-secreting ABC transporters as well.

As applications of Hypothesis 2, we suggested a few methods to practically “fix” non-secretable cargo proteins. These methods were negative supercharging (see Figure 4), super-neutralization (see Figure 5), random mutation accompanied by activity screening (see Figure 6), and linear charge density-focused supercharging (see Figure 7). They all work by lowering the linear charge density of the target protein via mutagenesis. We demonstrated that all these methods can generate a secretable derivative of a given target protein. The latter three options had a better chance to yield mutants that reserve the original protein’s functionality. Furthermore, many of these secretable variants also showed enhanced overall expression levels (see Figures 3f,m, 5a, 6a and 7e), with few exceptions (see Figure 4a). This phenomenon suggested that removing the product protein from the cytoplasm via secretion might also have an additional benefit in terms of overall expression in some cases, further justifying the development of methods aiming to rescue the protein secretion. Finally, according to the results in Figure 8, the secretion-promoting effect of negatively supercharging cargo proteins was also observed in other polypeptide-secreting ABC transporters. This suggests that the phenomena discussed in this study might potentially be more universal, rather than being specific to the *P. fluorescens* TliDEF ABC transporter complex.

The bacterial ABC transporter’s secretion dependence on the cargo protein’s charge density was studied in various aspects. The results revealed that the absence of cationic supercharged regions in the linear charge density of the cargo proteins, not the overall negative charge, might be the actual determinant of the ABC-dependent secretion. We also provided very potent tools to make proteins secretable by the ABC transporter and thus be compatible with the subsequent secretion-based production. We hope that these results make it possible to predict whether any given protein of interest can be secreted by ABC transporters before actual testing and make it possible to produce it via secretion anyways, even if the prediction results were negative. Together, these discoveries provide a deeper understanding of the characteristics of secretable substrates of bacterial ABC transporters, as well as revolutionize the potential of bacterial ABC transporters as the platform for efficient secretion-based protein production.

4. Materials and Methods

4.1. Bacterial Strains and Growth Conditions

All major gene works, including the plasmid construction, were performed using the *E. coli* XL1-BLUE strain. Protein expression and secretion were observed in both *E. coli* XL1-BLUE and *P. fluorescens* $\Delta tliA \Delta prtA$ strains [11]. We performed the *E. coli* transformation with the standard heat shock method, while *P. fluorescens* transformation was performed via electroporation at 2.5 kV, 125 Ω , and 50 μ F, with electrocompetent cells prepared using a standard electroporation protocol [32]. M9 minimal media, Lysogeny broth (LB), and terrific broth (TB) supplemented with 30 μ g/mL of kanamycin were used for the liquid cultures, which were usually 5 mL in a 180 rpm shaking incubator if not specified otherwise. The liquid cultures were grown until the stationary phase was reached. The protein expression was not regulated by induction. The cells constantly expressed the target proteins during their growth. LB agar plates supplemented with 30 μ g/mL of kanamycin were used for the plate colony cultures. In most cases, *E. coli* and *P. fluorescens* were cultured at 37 and 25 $^{\circ}$ C, respectively, if not specified otherwise. As an exception, *E. coli* enzymatic plate assays (detailed later) were incubated at 25 $^{\circ}$ C. The cells seeded in liquid culture were incubated until they reached an A_{600} of \sim 3, which is their full growth stage. *E. coli* cells typically take about 1.5 days of growth to reach this stage, while *P. fluorescens* cells require 2 days. The proteins were analyzed for both expression and secretion by seeding the transformed cells in liquid cultures or streaking them on the solid plate activity assay.

4.2. Examining the Expression and Secretion of the Proteins

We expressed the prepared plasmids in *P. fluorescens* liquid culture, mainly LB or, less frequently, TB supplemented with 60 µg/mL of kanamycin. The liquid culture was harvested after reaching the stationary phase, which occurs at approximately $A_{600} = 3\text{--}4$. The culture was then centrifuged at 18,000 rcf for 10 min, where we took the cell pellet and the liquid supernatant separately, and sample amounts equivalent to 16 µL of original cultures were loaded in SDS-PAGE using 10% polyacrylamide gels. We used nitrocellulose membranes (Amersham, Germany) for Western blotting. Polyclonal anti-LARD3 rabbit immunoglobulin G (rIgG) was utilized as the primary antibody, with 1:3000 dilution in 5% skim milk solution, and anti-rabbit recombinant goat IgG-peroxidase (anti-rIgG goat IgG-peroxidase) was used as the secondary antibody with 1:5000 dilution. The bands were then detected using a chemiluminescence agent (Advansta WesternBright Pico, San Jose, CA). Western blot images were acquired using an Azure C600 automatic detection system (Azure Biosystems, Dublin, CA). All included Western blot images are representative results from at least three different repeated experiments, starting over again from cell culturing with independent colonies. Quantifications of the Western blot images were performed with ImageJ software (<https://imagej.nih.gov/ij/>, accessed on 11 March 2019), with the unprocessed raw TIFF image data. Within the ImageJ output, the raw integrated density (RawIntDen) of each band was subtracted from a background RawIntDen measured from empty regions with the same area from the identical membrane image. This value was labeled as “calibrated intensity (*I*)”. Then, the % secretion values were calculated as follows:

$$\% \text{ Secretion} = \frac{I_{\text{supernatant}}}{I_{\text{supernatant}} + I_{\text{cell pellet}}} \times 100\%. \quad (1)$$

Statistical analyses were performed with the paired, single-tailed Student's *t*-test, where the signatures from the same image were paired together. The significance threshold (α) was set to be 0.05, which was equivalent to 5% chance of the null hypothesis (no secretion difference between the two compared groups) being true.

4.3. Examining the Enzymatic Activities of the Substrate Proteins

We utilized the LB agar plate-based enzymatic activity assay of the expressed and secreted proteins to observe the activity of the recombinant proteins. The colonies of *P. fluorescens* harboring the recombinant plasmids were streaked on the LB agar plate supplemented with enzyme substrates, which will be digested by the enzymes and exhibit visual changes (halo) around the cell colonies. The cells expressing tyrosinases (MelC2 and its derivatives) were streaked on LB agar plates supplemented with 0.1 mM CuCl₂ and 2 mM L-tyrosine, and the previously mentioned concentrations of appropriate antibiotics. The cells expressing the proteins with triglyceride lipase/esterase activities (M37, cutinase, and their derivatives) were streaked on LB agar plates supplemented with 0.5% colloidal tributyrin (glyceryl tributyrate) and antibiotics. The plates were incubated in a standing incubator at 25 °C for 3 days. The lipase/cutinase activity in the culture supernatant was measured using *p*-nitrophenol-palmitate (pNPP) and *p*-nitrophenol-butylate (pNPB). Then, 5 µL of lipase supernatant was added to 200 µL of pNPP reaction solution and 1 µL of cutinase supernatant to 200 µL of pNPB reaction solution. For the pNPP reaction solution, 10 mM of pNPP in acetonitrile, ethanol, and 50 mM Tris (pH 8.5) were mixed at a ratio of 1:4:95. Similarly for the pNPB solution, 10 mM of pNPB in acetonitrile, ethanol, and 50 mM Tris (pH 8.0) were mixed at a ratio of 1:4:95. The reaction solution was incubated for 10 min at 45 °C and monitored at $A_{405\text{nm}}$. Statistical analyses were performed with Student's *t*-test, double-tailed and homoscedastic (as the *F*-test failed to deny heteroscedasticity). The significance threshold (α) were set to be 0.05 for both the *t*-test and the *F*-test.

4.4. The Positively Charged Patch Experiments

For construction of pBR10 (which adds Arg10), pBK10 (Lys10), pBR05 (Arg05), pBK05 (Lys05), pBRD10 (ArgAsp10), pBKD10 (LysAsp10), pBRN10 (ArgAsn10), and pBKD10 (LysAsp10), we did not use the PCR-based amplification. This is primarily because the sequence we were aiming to insert was short (<30 bp). We first ordered the primers given by the upper half of Table S1. As these primers overlapped with each other entirely, they created a double-stranded region with the DNA sequence for the insert. The insert region was the desired sequences plus a few extra bases to align the reading frame. In addition, we added a new SacI site in this region, since the original SacI site in the pDART was supposed to be consumed during the In-Fusion cloning step. The inserted region was flanked by single-stranded hangovers, which can complementarily bind to the termini of the SacI-digested pDART plasmid via In-Fusion cloning (Takara, Shiga, Japan). Since these fragments already contain the single-stranded region for binding, we ran the 3'-to-5' exonuclease reaction step without the insert mix. The insert mix was added immediately after the reaction step was completed. The resulting colonies after the In-Fusion cloning were examined via DNA sequencing, to verify that they were properly constructed. In fact, due to the way the primers complementarily bind, pBKD07 (LysAsp07) and pBKD07 (LysAsp07) were constructed instead of the intended plasmids. Therefore, we repeated the above steps with different primers (also provided in Table S1) to add three additional repeats to finally construct the intended pBKD10 (LysAsp10) and pBKD10 (LysAsp10). Then, the prepared plasmids were digested with XbaI and SacI restriction enzymes to clone the genes for TliA, GFP, and GFP(-30), using the standard gene cloning method, using the multiple cloning site of the constructed plasmids.

4.5. Manual Mutagenesis

We synthesized the DNA sequence of wild-type and supercharged versions of glutathione S-transferase (GST) and streptavidin (SAv), both negatively and positively supercharged, as reported in a previous work of Liu's group [15]. Additionally, we manually created a less radically supercharged version of positively supercharged GST (GST(+19)) because the version published by Liu's group had a radical level of supercharging that would be unnecessary for our purpose. The super-neutralization processes (MelC2(Q), M37(Q)) were also performed manually. The structures of the proteins were either downloaded from the Protein Data Bank (database <https://www.rcsb.org/>, accessed on 8 January 2020) or predicted structures obtained from SWISS-MODEL (<https://swissmodel.expasy.org/>, accessed on 30 January 2020) [33]. The structures of the proteins were examined, and the charged amino acid residues exposed to the solvent were marked, and these amino acids were converted into positively charged amino acid (for the case of GST(+19)) or glutamine (for the super-neutralized proteins).

4.6. Computational Design of Supercharged Proteins

We utilized Liu's group's AvNAPSA algorithm [15,34]. We re-wrote the original Perl script into a python 3.9 script and used it to create negatively supercharged versions of *Nectria haematococca* cutinase (Cuti), *Photobacterium lipolyticum* M37 lipase (M37), and *Streptomyces antibioticus* MelC2 tyrosinase (MelC2). AvNAPSA is an abbreviation for "average number of neighboring atoms per side-chain atom" [35]. It was originally developed by Liu's group to make a protein with resilience to refolding, and the supercharged proteins were also applied for cellular protein targeting in animal cells. In this paper, however, the supercharging protocol was used to generate proteins that are secreted by the ABC transporter complex. We gradually mutated positively charged residues in the increasing order of the AvNAPSA score (low AvNAPSA score means that the residue has less interaction with other amino acids and that the residue was located on the surface of the protein) until the number of mutated residues was similar to the results we obtained during the manual supercharging process. The exact AvNAPSA thresholds for our mutated proteins

are presented in Table S2. Also note that we excluded any residue proximal (closer than seven residues apart) to the active site residues, cited from their respective literature.

4.7. Synthesizing and Cloning the Genes of Supercharged Proteins

We ordered the DNA synthesis service provided by Macrogen (Seoul, South Korea), and Cosmogenetech (Seoul, South Korea). We amplified the genes with PCR and then cloned in the *E. coli*, using the pDART plasmid as the vector. The pDART plasmid was developed in our previous study [12], which acts as a shuttle vector between *E. coli* and *P. fluorescens*. The pDART plasmid also contains the gene for the LARD3 signal sequence, which is conjugated to the C-terminus of the introduced gene, for the recognition by the TliDEF transporter complex. Additionally, the vector enhances the expression level of the TliDEF transporter complex.

4.8. Various ABC Transporters Test (Double-Transformation)

We constructed pLARD3 from the commercially available pKK223-3 vector. The multiple cloning site pKK223-3 vector was fused with the *LARD3* gene so that the gene inserted to the multiple cloning site was expressed with a C-terminally fused LARD3 signal sequence. We recombined this pLARD3 with *cuti*(−) and *cuti* genes, prepared previously, to construct pLARD3-Cuti(−) and pLARD3-Cuti. Then, these plasmids were both introduced to *E. coli* along with the plasmids that encode the ABC transporter genes.

4.9. Synthesizing and Cloning the Gene of Variationally Supercharged M37

We prepared the DNA sequence of variationally supercharged M37 lipase, which we named M37(var), by replacing the codons for the surface-exposed positively charged amino acid residues with the degenerate codon. For example, the IUPAC DNA code R denotes the purine base, which is guanine or adenine. We aimed to find the Goldilocks variant somewhere between M37(−14), which had enzymatic activity but was not secreted, and M37(−23), which was secreted but had no enzymatic activity. We examined the amino acid sequences of M37(−23) and M37(−14) and marked the residues where the two differed. Then, we replaced codons for those residues with the degenerate codons. For instance, the residue Lys36 of M37 lipase was replaced by glutamic acid in M37(−23) but remained unchanged in M37(−14). Therefore, we placed the degenerate codon “RAG” at that position. For the degenerate codon RAG, there were two possible outcomes: GAG, which codes for glutamic acid, and AAG, which codes for lysine. We would have preferred to use the set “glutamic acid or arginine”, but such a combination was impossible. Therefore, we used the RAG codon in place of these residues as well. The exact sequence we ordered for the construction of M37(var) is provided in Supplementary Section A. After the DNA sequence was designed, we used the DNABworks web server (<https://hpcwebapps.cit.nih.gov/dnaworks/>, accessed on 18 January 2018) [36] to convert the sequence into a set of synthesizable primers. The parameters we used were as follows: oligo length 58 nucleotides, annealing temperature 62 °C, oligo concentration 1.00×10^{-7} M, Na^+/K^+ concentration 0.05 M, Mg^{2+} concentration 0.002 M, number of solutions 1, and no “TBIO” mode (“PTDS” mode was used instead). Then, we manually examined the output oligos and made sure that no degenerate codon was present at the end of any overlapping region between oligos. We ordered the oligos from Cosmogenetech, and we assembled them using the PCR-based DNA synthesis method described in a previous publication [37]. The obtained PCR product was purified, restricted, and then introduced into the pDART plasmid like the rest of the genes handled in this study.

4.10. Linear Charge Density Analyses

The “linear charge density (LCD)” was defined as the sum of the charges over a given number (w , representing “window”) of consecutive amino acids in a single polypeptide chain, divided by w . For example, if the window width $w = 20$, then the linear charge density at index i (i is a positive integer) is the sum of the charges of the 20 amino acids

from position i to position $i + 19$, divided by 20. We arbitrarily chose $w = 20$. One of the justifications for choosing 20 was that the average alpha helix spanning through a membrane is roughly that long. Thus, we assumed that a linearized polypeptide being translocated through the membrane channel would have at maximum 20 amino acids under the influence of the electric field formed by the membrane potential at a time. The formula for the linear charge density at position i (λ_i) was the following:

$$\lambda_i = \frac{1}{w} \sum_{j=i}^{i+w-1} q_j, \quad (2)$$

where w is window width and q_j is the average charge of the side chain of residue j at pH 7. We used -1 for aspartic acid and glutamic acid, $+1$ for arginine and lysine, and $+0.1$ for histidine. Interestingly, the non-secreted proteins all had “peaks” over $+2$ (unit was $e/20$ aa, elementary charges per 20 amino acids) at more than one position within their sequence, while almost all the secreted proteins had no such peaks (and as specified in the Discussion Section, these exceptions had their peaks mainly composed of lysines, not arginines). The LCD graphs were obtained using a script that we developed and published as a webserver (<https://mb.re.kr/apps/LCD>, accessed on 11 May 2022).

4.11. The linear Charge Density-Focused Supercharging

Linear charge densities of the target proteins (TGF β , TNF β , FGF1, and SARS-CoV-2 spike protein domains) were automatically calculated using a Python script that we created. We identified the peaks with height above $+2 e/20$ aa and aimed to remove those peaks by replacing the non-conserved, positively charged residues within those peaks. The Bayesian conservation score (a measure of the degree of conservation of a given residue among the homologs of the query protein) of each residue was calculated using the ConSurf webserver (<http://consurf.tau.ac.il/>, accessed on 8 March 2020) [38,39]. Examining the Bayesian conservation scores, the positively charged residues with low conservation scores (<4) around the high linear charge density peaks were selected and replaced. If there was any negatively charged or hydrophilic amino acid that occurred at the position in question among the homologs returned by the ConSurf analysis, we preferred to use that amino acid. Here, we made sure that the residue was solvent-accessible by examining its reported structure in the protein data bank (PDB) database (<https://www.rcsb.org/>, accessed on 8 January 2020) or predicted structure obtained from SWISS-MODEL (<https://swissmodel.expasy.org/>, accessed on 30 January 2020) [33]. In addition, we looked through the literature to make sure that the altered residues were not present on the receptor-interacting surface. Finally, we examined the stability of the interaction between the final modified protein and the receptor protein by running the docking molecular dynamics (docking MD) simulation, using the following procedure: The PDB files of the protein of interest were initially preprocessed using the default parameters of the Schrödinger Desmond program (New York, NY) and missing side chains and loops were filled in using Prime. PROPKA pH 7.0 was used to assign the H-bonds of the structure, which were further optimized by the OPLS3e field model. Lastly, the simulation of the ligand-receptor complex was run for 50 ns in Schrödinger Desmond. We published the LCD-focused supercharging tool as a webserver, which can be accessed using the following link: <https://mb.re.kr/apps/supercharge> (accessed on 11 May 2022).

4.12. Protein Structure Prediction

The structures presented in Figure 1 were predicted from the SWISS-MODEL webserver (<https://swissmodel.expasy.org/>, accessed on 25 November 2018) [33]. In the process, TliD prediction used *Aquifex aeolicus* PrtD (PDB ID 5I22.1) as the template, with sequence identity of 40.98%. TliF prediction used *Salmonella typhi* ST50 (PDB ID 5bun.1) with sequence identity of 26.52%, but with a good QMEAN score of -1.19 . LARD3 signal sequence prediction used a *Pseudomonas* species' MIS38 lipase (PDB ID 2zj7.1) with sequence

identity of 53.85%. TliE prediction used *Serratia marcescens* LipC (PDB ID 5nen.1), with sequence identity of 44.19%. In the template 5nen, the membrane fusion protein LipC was crystallized in dimer, which is different from the expected natural homo-hexameric state of the ABC transporter's membrane fusion proteins, as specified by the authors of 5nen [40]. The structure of the homo-hexamer complex of TliE was predicted using the Galaxy-Homomer webserver (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=HOMOMER>, accessed on 12 February 2020) [41], using the SWISS-MODEL predicted structure of monomeric TliE as the input, with the default parameters. We further optimized the result by repeating the optimization service within the GalaxyHomomer server.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23126700/s1> [42–46].

Author Contributions: Conceptualization, J.H.A. and H.L.; methodology, H.B.; software, H.B. and M.H.N.; validation, J.H.A. and H.L.; formal analysis, H.B.; investigation, H.B., J.P., B.U.F., J.B., and J.H.A.; resources, J.H.A.; data curation, H.B. and J.P.; writing—original draft preparation, H.B.; writing—review and editing, H.L. and J.H.A.; visualization, H.B.; supervision, J.H.A. and H.L.; project administration, J.H.A.; funding acquisition, J.H.A. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Global Frontier Project from the Intelligent Synthetic Biology Center and the Korea Science Academy of KAIST, with funds from the Ministry of Science and ICT of the Republic of Korea.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The sequences of the constructed mutant polypeptides are provided in FASTA format in Appendix A. The Western blot raw data are provided in Supplementary Text S2.

Acknowledgments: We are very grateful to David R. Liu for providing the AvNAPSA Perl script.

Conflicts of Interest: The authors declare no competing financial interests or other conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

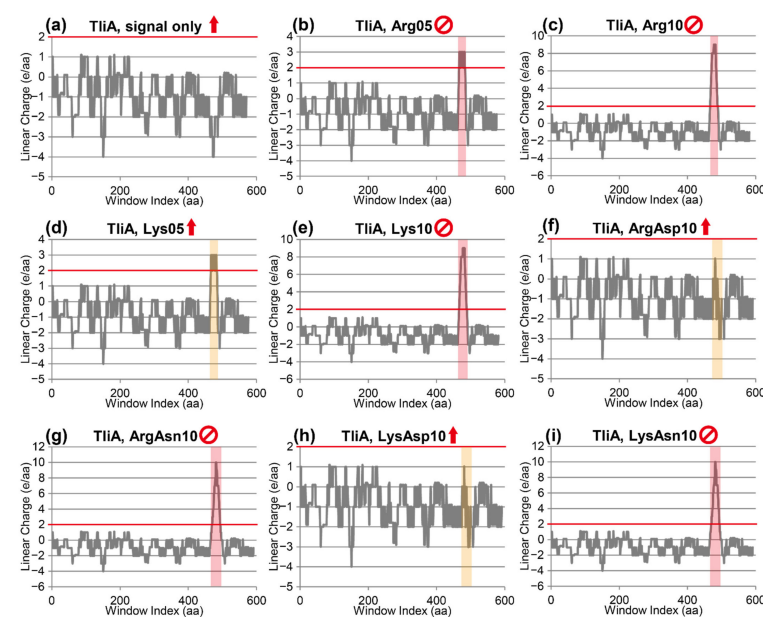


Figure A1. Linear charge density graphs of the artificial oligopeptide patch experiments, presented in Figures 3 and A2. A red arrow indicates that the protein was secreted, while a red cancellation

sign indicates that the protein was not secreted. Note that insertion of artificial oligopeptides that blocked the secretion, namely Arg05, Arg10, Lys10, ArgAsn10, and LysAsn10, all created significant peaks $\geq +2 e/aa$. Curiously, Lys05 in panel (d) did create a peak over the threshold (yellow background), but nonetheless allowed the secretion, illustrating the difference between lysine and arginine residues.

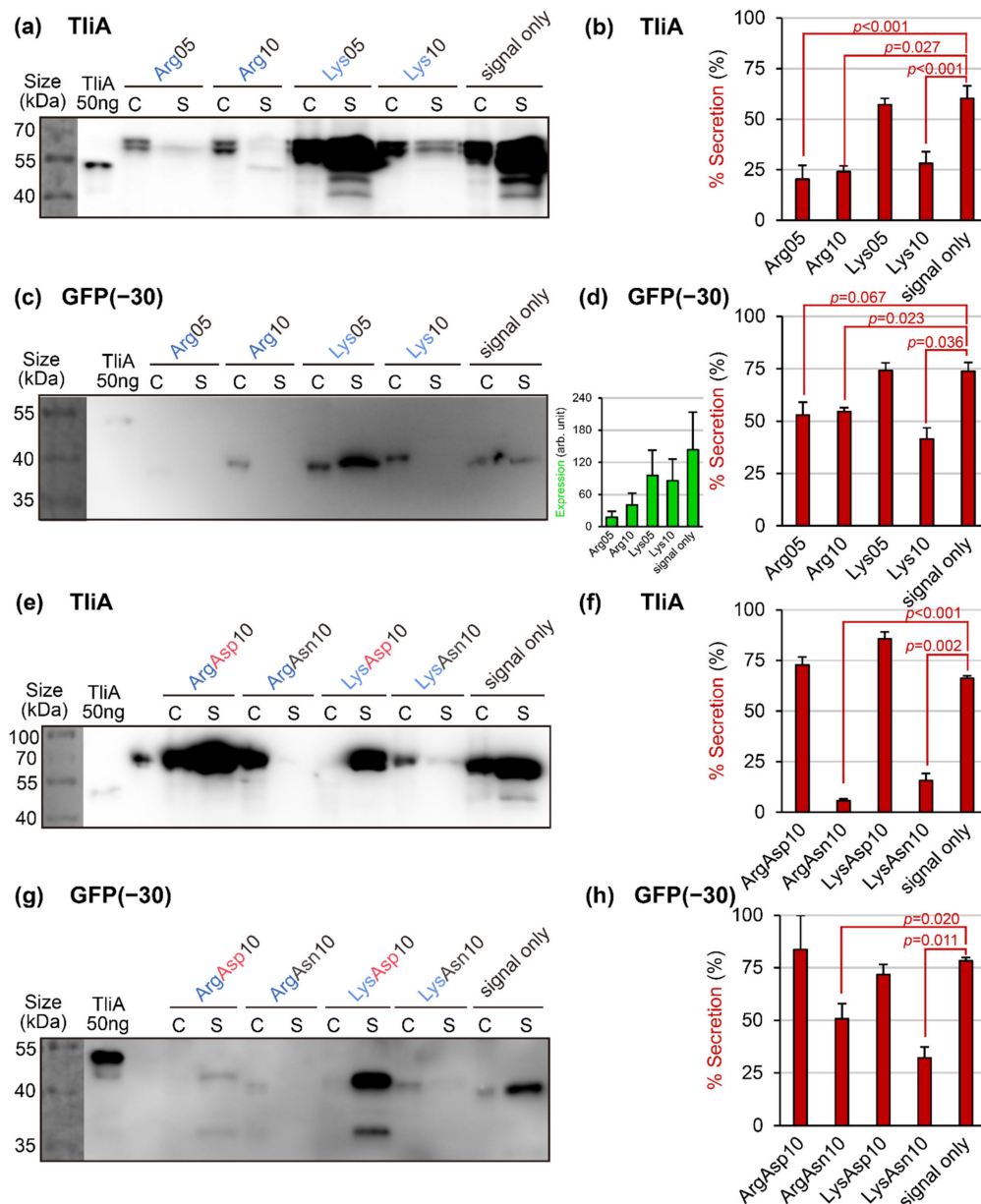


Figure A2. Artificial oligopeptide insertion experiments, repeated with other substrate proteins. These are reiterations of the experiments presented in Figure 3. The additional substrates are: *Pseudomonas fluorescens* thermostable lipase (TliA) and the negatively supercharged GFP (GFP(-30)). Note that both of these substrates are usually secreted better than GFP. These two substrates exhibited a very similar pattern to that of GFP (see Figure 3). (a,b) A reiteration of the experiment in Figure 3f,g using TliA as the substrate. (c,d) A reiteration of the experiment in Figure 3f,g using GFP(-30) as the substrate. Note that the Arg05 sample exhibited a very low total expression level (as can be seen from panel (d), smaller chart), which might have caused the lower statistical significance. (e,f) A reiteration of the experiment in Figure 3m,n using TliA as the substrate. (g,h) A reiteration of the experiment in Figure 3m,n using GFP(-30) as the substrate.

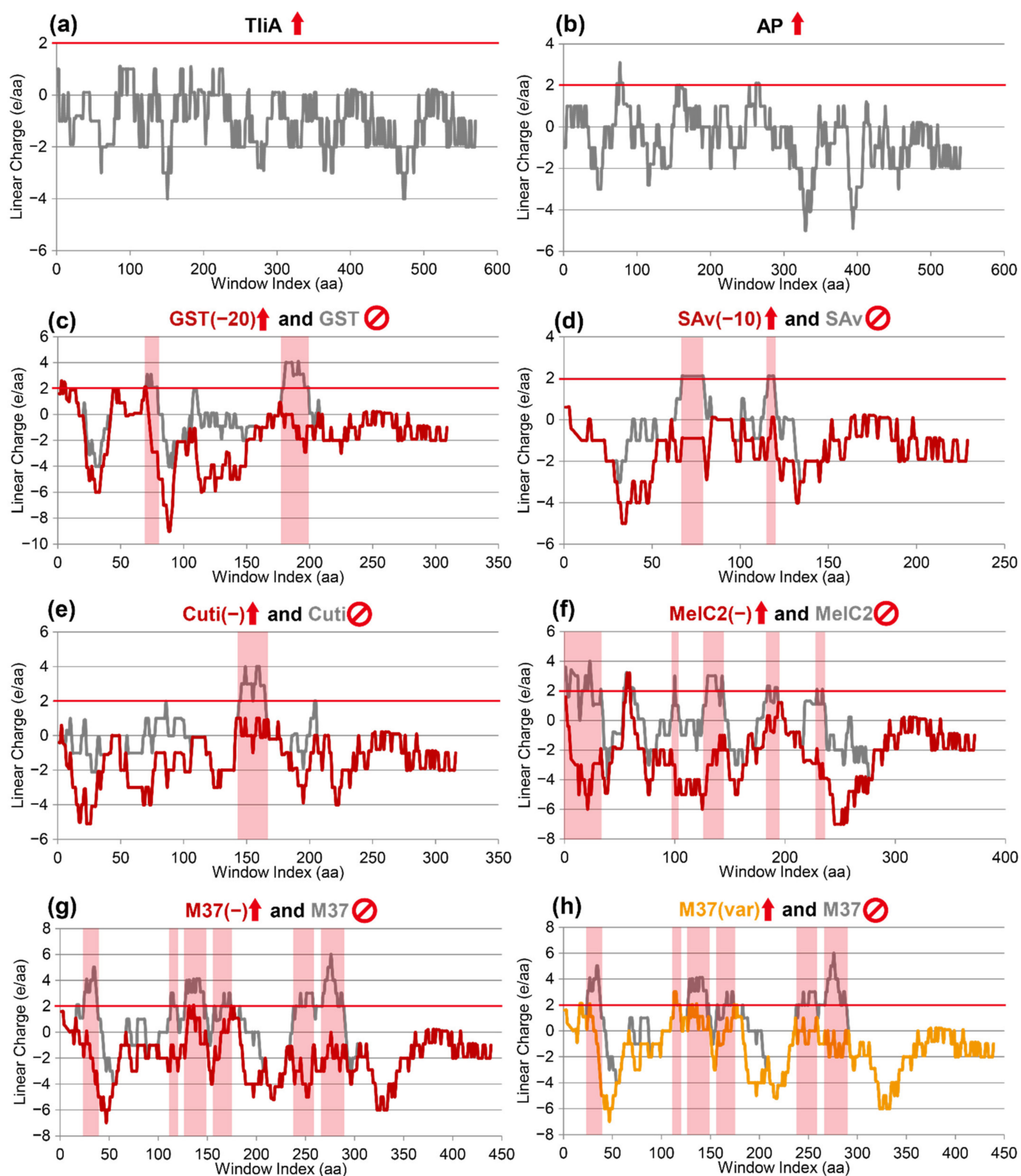


Figure A3. The linear charge density graphs omitted from the main text. The horizontal red lines indicate the arbitrary $+2 e/aa$ threshold which we set. The prominent peaks are marked with the red background color. A red arrow indicates that the protein was secreted, while a red cancellation sign indicates that the protein was not secreted. (a) The LCD graph of TliA, the natural substrate of the TliDEF ABC transporter that we used in this experiment. (b) The LCD graph of a typical secretable protein, *E. coli* alkaline phosphatase (AP). Note that the peaks above $+2$ are not prominent. (c,d) The

LCD graphs of glutathione S-transferase (GST) and streptavidin (SAv), and their negatively supercharged variants, GST(-20) and SAv(-10). The secretion tests of these proteins are presented in Figure 4a. (e–g) The LCD graphs of cutinase (Cuti), MelC2 tyrosinase, and M37 lipase, and their negatively supercharged variants Cuti(-), MelC2(-), and M37(-). The secretion tests of these proteins are presented in Figure 4c. (h) The LCD graphs of M37 lipase and its variant obtained from a mutation and screening approach, M37(var). The secretion tests of these proteins are presented in Figure 6a.

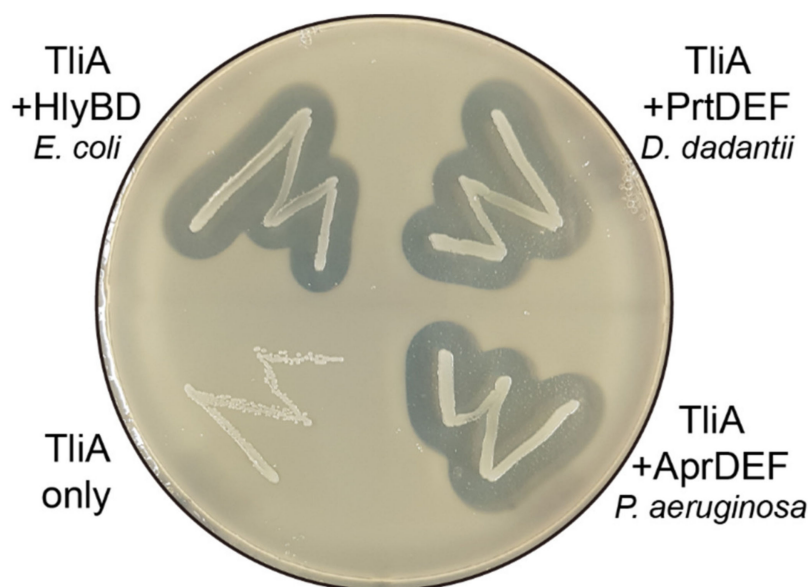


Figure A4. The other bacterial T1SS transporters are compatible with the LARD3 signal sequence used throughout the experiments in this study. Three different bacterial T1SS transporters, *P. aeruginosa* AprDEF, *D. dadantii* PrtDEF, and *E. coli* HlyBD + TolC, can successfully secrete TliA thermostable lipase by recognizing the LARD3 signal sequence naturally present in TliA. The secreted TliA lipase creates a visible halo on the plate lipase activity assay. The negative control colonies only harbor the *tliA* gene and no genes for the transporters, so the expressed TliA lipase stayed in the cytoplasm and no halo was formed.

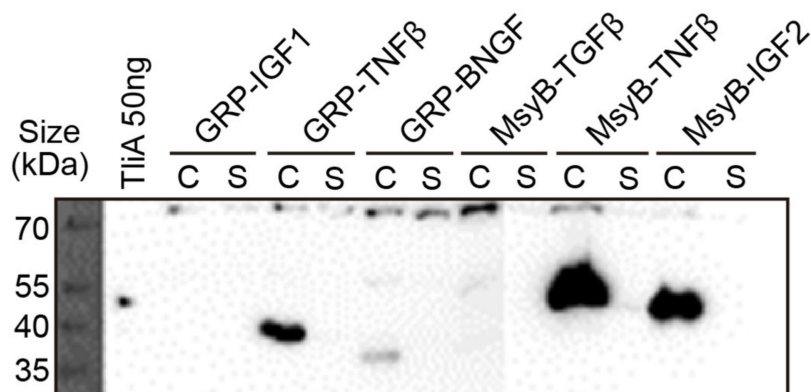


Figure A5. The N-terminal fusion of acidic proteins could not rescue the secretion of non-secretable proteins. Two different very anionic (acidic) proteins, *Arabidopsis thaliana* glycine-rich protein (GRP, pI = 3.07) and *E. coli* MsyB (pI = 3.60), were attached to multiple different cationic (basic) proteins. None of them could rescue the secretion of these proteins, as clearly shown in all the samples. All the target proteins were localized in the cell pellet portions (“C”), and no signal was detected in the culture supernatant portions (“S”).

References

1. Baneyx, F. Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **1999**, *10*, 411–421. [[CrossRef](#)]
2. Structural Genomics Consortium; China Structural Genomics Consortium; Northeast Structural Genomics Consortium; Gräslund, S.; Nordlund, P.; Weigelt, J.; Hallberg, B.M.; Bray, J.; Gileadi, O.; Knapp, S.; et al. Protein production and purification. *Nat. Methods* **2008**, *5*, 135–146.
3. Chen, R. Bacterial expression systems for recombinant protein production: *E. Coli* and beyond. *Biotechnol. Adv.* **2012**, *30*, 1102–1107. [[CrossRef](#)] [[PubMed](#)]
4. Ihling, N.; Uhde, A.; Scholz, R.; Schwarz, C.; Schmitt, L.; Büchs, J. Scale-up of a type i secretion system in *e. Coli* using a defined mineral medium. *Biotechnol. Prog.* **2020**, *36*, e2911. [[CrossRef](#)] [[PubMed](#)]
5. Burdette, L.A.; Leach, S.A.; Wong, H.T.; Tullman-Ercek, D. Developing gram-negative bacteria for the secretion of heterologous proteins. *Microb. Cell Factories* **2018**, *17*, 196. [[CrossRef](#)]
6. Von Graevenitz, A.; Weinstein, J. Pathogenic significance of *pseudomonas fluorescens* and *pseudomonas putida*. *Yale J. Biol. Med.* **1971**, *44*, 265–273.
7. Huang, K.-X.; Badger, M.; Haney, K.; Evans, S.L. Large scale production of *bacillus thuringiensis* ps149b1 insecticidal proteins cry34ab1 and cry35ab1 from *pseudomonas fluorescens*. *Protein Expr. Purif.* **2007**, *53*, 325–330. [[CrossRef](#)]
8. Squires, C.H.; Retallack, D.M.; Chew, L.C.; Ramseier, T.M.; Schneider, J.C.; Talbot, H.W. Heterologous protein production in *p. Fluorescens*. *BioProcess Int.* **2004**, *2004*, 54–58.
9. Ahn, J.H.; Pan, J.G.; Rhee, J.S. Identification of the *tlid* abc transporter specific for lipase in *pseudomonas fluorescens* sik w1. *J. Bacteriol.* **1999**, *181*, 1847–1852. [[CrossRef](#)]
10. Park, Y.; Moon, Y.; Ryoo, J.; Kim, N.; Cho, H.; Ahn, J.H. Identification of the minimal region in lipase abc transporter recognition domain of *pseudomonas fluorescens* for secretion and fluorescence of green fluorescent protein. *Microb. Cell Factories* **2012**, *11*, 60. [[CrossRef](#)]
11. Son, M.; Moon, Y.; Oh, M.J.; Han, S.B.; Park, K.H.; Kim, J.-G.; Ahn, J.H. Lipase and protease double-deletion mutant of *pseudomonas fluorescens* suitable for extracellular protein production. *Appl. Environ. Microbiol.* **2012**, *78*, 8454–8462. [[CrossRef](#)]
12. Ryu, J.; Lee, U.; Park, J.; Yoo, D.-H.; Ahn, J.H. A vector system for abc transporter-mediated secretion and purification of recombinant proteins in *pseudomonas* species. *Appl. Environ. Microbiol.* **2015**, *81*, 1744–1753. [[CrossRef](#)]
13. Park, J.; Eom, G.T.; Young Oh, J.; Hyun Park, J.; Kim, S.C.; Kwang Song, J.; Hoon Ahn, J. High-level production of bacteriotoxic phospholipase a1 in bacterial host *pseudomonas fluorescens* via abc transporter-mediated secretion and inducible expression. *Microorganisms* **2020**, *8*, 239. [[CrossRef](#)] [[PubMed](#)]
14. Byun, H.; Park, J.; Kim, S.C.; Ahn, J.H. A lower isoelectric point increases signal sequence-mediated secretion of recombinant proteins through a bacterial abc transporter. *J. Biol. Chem.* **2017**, *292*, 19782–19791. [[CrossRef](#)]
15. Lawrence, M.S.; Phillips, K.J.; Liu, D.R. Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* **2007**, *129*, 10110–10112. [[CrossRef](#)] [[PubMed](#)]
16. Delepelaire, P. Type i secretion in gram-negative bacteria. *Biochim. Biophys. Acta* **2004**, *1694*, 149–161. [[CrossRef](#)] [[PubMed](#)]
17. Bakkes, P.J.; Jenewein, S.; Smits, S.H.J.; Holland, I.B.; Schmitt, L. The rate of folding dictates substrate secretion by the *escherichia coli* hemolysin type 1 secretion system. *J. Biol. Chem.* **2010**, *285*, 40573–40580. [[CrossRef](#)]
18. Schwarz, C.K.W.; Lenders, M.H.H.; Smits, S.H.J.; Schmitt, L. Secretion of slow-folding proteins by a type 1 secretion system. *Bioengineered* **2012**, *3*, 289–292. [[CrossRef](#)]
19. Bumba, L.; Masin, J.; Macek, P.; Wald, T.; Motlova, L.; Bibova, I.; Klimova, N.; Bednarova, L.; Veverka, V.; Kachala, M.; et al. Calcium-driven folding of rtx domain beta-rolls ratchets translocation of rtx proteins through type i secretion ducts. *Mol. Cell* **2016**, *62*, 47–62. [[CrossRef](#)]
20. Berg, J.M.; Tymoczko, J.L.; Stryer, L. *Biochemistry*, 6th ed.; W.H. Freeman: New York, NY, USA, 2007.
21. Booth, I.R. Regulation of cytoplasmic pH in bacteria. *Microbiol. Rev.* **1985**, *49*, 359–378. [[CrossRef](#)]
22. Wilks, J.C.; Slonczewski, J.L. Ph of the cytoplasm and periplasm of *escherichia coli*: Rapid measurement by green fluorescent protein fluorimetry. *J. Bacteriol.* **2007**, *189*, 5601–5607. [[CrossRef](#)] [[PubMed](#)]
23. Arce-Rodriguez, A.; Volke, D.C.; Bense, S.; Häussler, S.; Nickel, P.I. Non-invasive, ratiometric determination of intracellular pH in *pseudomonas* species using a novel genetically encoded indicator. *Microb. Biotechnol.* **2019**, *12*, 799–813. [[CrossRef](#)] [[PubMed](#)]
24. Panahi, A.; Brooks, C.L. Membrane environment modulates the pKa values of transmembrane helices. *J. Phys. Chem. B* **2015**, *119*, 4601–4607. [[CrossRef](#)] [[PubMed](#)]
25. Andersson, H.; von Heijne, G. Positively charged residues influence the degree of secA dependence in protein translocation across the *e. Coli* inner membrane. *FEBS Lett.* **1994**, *347*, 169–172. [[CrossRef](#)]
26. Summers, R.G.; Knowles, J.R. Illicit secretion of a cytoplasmic protein into the periplasm of *escherichia coli* requires a signal peptide plus a portion of the cognate secreted protein. Demarcation of the critical region of the mature protein. *J. Biol. Chem.* **1989**, *264*, 20074–20081. [[CrossRef](#)]
27. Summers, R.G.; Harris, C.R.; Knowles, J.R. A conservative amino acid substitution, arginine for lysine, abolishes export of a hybrid protein in *escherichia coli*. Implications for the mechanism of protein secretion. *J. Biol. Chem.* **1989**, *264*, 20082–20088. [[CrossRef](#)]
28. Boyd, D.; Beckwith, J. The role of charged amino acids in the localization of secreted and membrane proteins. *Cell* **1990**, *62*, 1031–1033. [[CrossRef](#)]

29. Allen, W.J.; Corey, R.A.; Watkins, D.W.; Pereira, G.C.; Oliveira, A.S.F.; Hards, K.; Cook, G.M.; Collinson, I. Rate-limiting transport of positive charges through the Sec-machinery is integral to the mechanism of protein transport. *bioRxiv* **2022**. [[CrossRef](#)]
30. Simon, S.M.; Peskin, C.S.; Oster, G.F. What drives the translocation of proteins? *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3770–3774. [[CrossRef](#)]
31. Allen, W.J.; Corey, R.A.; Oatley, P.; Sessions, R.B.; Baldwin, S.A.; Radford, S.E.; Tuma, R.; Collinson, I. Two-way communication between secy and seca suggests a brownian ratchet mechanism for protein translocation. *eLife* **2016**, *5*, e15598. [[CrossRef](#)]
32. Seidman, C.E.; Struhl, K.; Sheen, J.; Jessen, T. Introduction of plasmid DNA into cells. In *Current Protocols in Molecular Biology*; Ausubel, M.F., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K., Eds.; John Wiley & Sons, Inc.: New York, NY, USA, 1987; Volume 1.
33. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L. Swiss-model: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [[CrossRef](#)] [[PubMed](#)]
34. Der, B.S.; Kluwe, C.; Miklos, A.E.; Jacak, R.; Lyskov, S.; Gray, J.J.; Georgiou, G.; Ellington, A.D.; Kuhlman, B. Alternative computational protocols for supercharging protein surfaces for reversible unfolding and retention of stability. *PLoS ONE* **2013**, *8*, e64363. [[CrossRef](#)] [[PubMed](#)]
35. Chew, L.C.; Ramseier, T.M.; Retallack, D.M.; Schneider, J.C.; Squires, C.H.; Talbot, H.W. *Pseudomonas fluorescens*. In *Production of Recombinant Proteins*; Gellissen, G., Ed.; Wiley-VCH: Weinheim, Germany, 2005; pp. 45–62.
36. Hoover, D.M.; Lubkowsky, J. Dnaworks: An automated method for designing oligonucleotides for pcr-based gene synthesis. *Nucleic Acids Res.* **2002**, *30*, e43. [[CrossRef](#)] [[PubMed](#)]
37. Xiong, A.-S.; Yao, Q.-H.; Peng, R.-H.; Li, X.; Fan, H.-Q.; Cheng, Z.-M.; Li, Y. A simple, rapid, high-fidelity and cost-effective pcr-based two-step DNA synthesis method for long gene sequences. *Nucleic Acids Res.* **2004**, *32*, e98. [[CrossRef](#)]
38. Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. Consurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **2016**, *44*, W344–W350. [[CrossRef](#)]
39. Landau, M.; Mayrose, I.; Rosenberg, Y.; Glaser, F.; Martz, E.; Pupko, T.; Ben-Tal, N. Consurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **2005**, *33*, W299–W302. [[CrossRef](#)]
40. Murata, D.; Okano, H.; Angkawidjaja, C.; Akutsu, M.; Tanaka, S.I.; Kitahara, K.; Yoshizawa, T.; Matsumura, H.; Kado, Y.; Mizohata, E.; et al. Structural basis for the serratia marcescens lipase secretion system: Crystal structures of the membrane fusion protein and nucleotide-binding domain. *Biochemistry* **2017**, *56*, 6281–6291. [[CrossRef](#)]
41. Baek, M.; Park, T.; Heo, L.; Park, C.; Seok, C. Galaxyhomomer: A web server for protein homo-oligomer structure prediction from a monomer sequence or structure. *Nucleic Acids Res.* **2017**, *45*, W320–W324. [[CrossRef](#)]
42. Yang, K.-S.; Sung, B.-H.; Park, M.-K.; Lee, J.-H.; Lim, K.-J.; Park, S.-C.; Kim, S.-J.; Kim, H.-K.; Sohn, J.-H.; Kim, H.-M.; et al. Recombinant Lipase Engineered with Amphipathic and Coiled-Coil Peptides. *ACS Catal.* **2015**, *5*, 5016–5025. [[CrossRef](#)]
43. Kim, D.; Jeon, C.; Kim, J.-H.; Kim, M.-S.; Yoon, C.-H.; Choi, I.-S.; Kim, S.-H.; Bae, Y.-S. Cytoplasmic transduction peptide (CTP): New approach for the delivery of biomolecules into cytoplasm in vitro and in vivo. *Exp. Cell Res.* **2006**, *312*, 1277–1288. [[CrossRef](#)]
44. Radaev, S.; Zou, Z.; Huang, T.; Lafer, E.M.; Hinck, A.P.; Sun, P.D. Ternary complex of transforming growth factor-beta1 reveals isoform-specific ligand recognition and receptor recruitment in the superfamily. *J. Biol. Chem.* **2010**, *285*, 14806–14814. [[CrossRef](#)] [[PubMed](#)]
45. Brown, A.; Robinson, C.J.; Gallagher, J.T.; Blundell, T.L. Cooperative heparin-mediated oligomerization of fibroblast growth factor-1 (FGF1) precedes recruitment of FGFR2 to ternary complexes. *Biophys. J.* **2013**, *104*, 1720–1730. [[CrossRef](#)] [[PubMed](#)]
46. Mori, S.; Wu, C.-Y.; Yamaji, S.; Saegusa, J.; Shi, B.; Ma, Z.; Kuwabara, Y.; Lam, K.S.; Isseroff, R.R.; Takada, Y.K.; et al. Direct Binding of Integrin $\alpha v \beta 3$ to FGF1 Plays a Role in FGF1 Signaling. *J. Biol. Chem.* **2008**, *283*, 18066–18075. [[CrossRef](#)] [[PubMed](#)]