

METHODOLOGY

Open Access



PySupercharge: a python algorithm for enabling ABC transporter bacterial secretion of all proteins through amino acid mutation

Yerin Kim^{1†}, Danny Kim^{1†}, Nguyen-Mihn Hieu¹, Hyunjong Byun² and Jung Hoon Ahn^{1,2*}

Abstract

Background The process of producing proteins in bacterial systems and secreting them through ATP-binding cassette (ABC) transporters is an area that has been actively researched and used due to its high protein production capacity and efficiency. However, some proteins are unable to pass through the ABC transporter after synthesis, a phenomenon we previously determined to be caused by an excessive positive charge in certain regions of their amino acid sequence. If such an excessive charge is removed, the secretion of any protein through ABC transporters becomes possible.

Results In this study, we introduce 'linear charge density' as the criteria for possibility of protein secretion through ABC transporters and confirm that this criterion can be applied to various non-secretable proteins, such as SARS-CoV-2 spike proteins, botulinum toxin light chain, and human growth factors. Additionally, we develop a new algorithm, PySupercharge, that enables the secretion of proteins containing regions with high linear charge density. It selectively converts positively charged amino acids into negatively charged or neutral amino acids after linear charge density analysis to enable protein secretion through ABC transporters.

Conclusions PySupercharge, which also minimizes functional/structural stability loss of the pre-mutation proteins through the use of sequence conservation data, is currently being operated on an accessible web server. We verified the efficacy of PySupercharge-driven protein supercharging by secreting various previously non-secretable proteins commonly used in research, and so suggest this tool for use in future research requiring effective protein production.

Keywords Python Algorithm, Supercharging, Protein production, Secretion, ABC Transporter

[†]Yerin Kim, Danny Kim authors contributed equally to this work.

*Correspondence:

Jung Hoon Ahn
hoony@kaist.ac.kr

¹Department of Chemistry and Biology, Korea Science Academy of Korea Advanced Institute of Science and Technology, Busan, South Korea

²Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea



Background

Reengineering protein amino acid sequences to decrease electric charge, or “supercharging”, can improve secretion ability of the protein through the bacterial ABC transporter system, which is an improved alternative to previous protein production methods [1–3]. Here, we introduce PySupercharge, an automated Python algorithm that supercharges protein sequences to be compatible for ABC transporter secretion.

The conventional method for protein production widely used in academia and industry utilizes non-secretory *Escherichia coli* systems [4, 5]. In this process, *E. coli* is transformed with the target protein’s expression vector, and the target protein is extracted through cell lysis. This *E. coli* system is widely used for its production fidelity [6, 7]. However, contaminant proteins are inevitably added during the lysis process, making multiple steps such as gel filtration and column chromatography necessary. This disadvantage of *E. coli* protein production systems is yet to be resolved. One of the alternative methods used is the secretion of proteins into the extracellular space [8, 9]. This secretion method allows for a simpler purification process and guarantees the production of proteins in their correctly folded, active forms [10].

Many gram-negative bacteria including *E. coli* and *Pseudomonas fluorescens* have the Type I secretion system (T1SS), [11, 12]. T1SS is composed of three proteins: an ATP-binding cassette (ABC) protein, a membrane fusion protein (MFP), and an outer membrane protein (OMP) [13]. Unfolded protein is secreted into the extracellular space, where it folds back to its original form [14]. A broad range of heterogenous proteins can be secreted by T1SS, such as various growth factors, green fluorescent protein (GFP), and enzymes [2, 15–19]. Since the T1SS secretes the target protein to extracellular space and gram-negative bacteria do not secrete native proteins [20], this system is suitable for extracellular protein production. However, the bacterial secretion system including the T1SS is protein-dependent [21, 22]. Even if a secretion signal or a peptide secretion tag required by the secretion system is present, the secretion efficiency has been found to vary greatly by protein. This limitation would easily be resolved by finding the criteria for secretion, but it is still elusive. Various criteria have been studied, such as folding kinetics, the presence of an A/U rich sequence, the electric charge, and the isoelectric point [1, 23, 24]. Among these conditions, our previous research has focused on manipulating proteins’ electric charge.

In our previous examination of various secretable and non-secretable heterogenous proteins, it was found that non-secretable proteins almost always have high-density ‘excessively cationic regions’ [25]. We suggested that these regions have electrostatic limitations in being transported from the negatively charged intracellular

space to the positively charged extracellular space. It was also shown that the artificial addition of excessively cationic regions to secretable proteins greatly reduces secretion efficiency. Therefore, determining whether the target protein has an excessively cationic region and neutralizing it can be critical for efficient protein production using bacterial secretion systems. Here, we introduce an automated algorithm, PySupercharge, for this process.

Methods

Linear charge density analysis

Unlike other supercharging tools that focus on the overall charge of the target protein, our method focuses on certain segments of a polypeptide sequence since proteins exit through T1SS in unfolded polypeptide form. Therefore, we utilize the local charge densities of the linearized amino acid sequence, which we called ‘linear charge density (LCD)’ in our previous study [25]. It is essentially the average charge of amino acids in a certain range or ‘window’ [25]. To express the criteria for excessively cationic regions more clearly, we used the total charge in a window instead of the average in this study. This modified LCD value of a window starting at initial position i is defined as λ_i by the following formula

$$\lambda_i = \sum_{j=i}^{i+w-1} q_j$$

where w is the window width and q_j is the charge of the side chain of residue j at pH 7.

From our previous study [25], we have determined that a window size of 20 amino acids is suitable for the *P. fluorescens* secretion system. This value was empirically determined based on the fact that 20 amino acids are influenced at a time by the membrane potential during transportation by the ABC transporter (Fig. 1). Also, the optimal LCD value cutoff for non-cationic-supercharged sequences was determined experimentally as $\leq (+2)$ in the window size of 20 amino acids.

The logic process of the main PySupercharge algorithm is simple. The amino acid sequence is analyzed, and the LCD value of a 20-amino-acid window is calculated for each available initial position i in the sequence. Neutral amino acids are assigned zero charge, while positive amino acids (Lys, Arg) are assigned +1 and negative amino acids (Glu, Asp) are assigned –1. Histidine charge can be optionally configured to have a user-defined positive charge, as in some cases the small charge assignment might determine an excessively cationic sequence.

From here, the algorithm previously simply converted positive Lys or Arg to a random choice between Glu and Asp if a window’s LCD value exceeded the cutoff (there is not yet any concrete evidence that shows significant

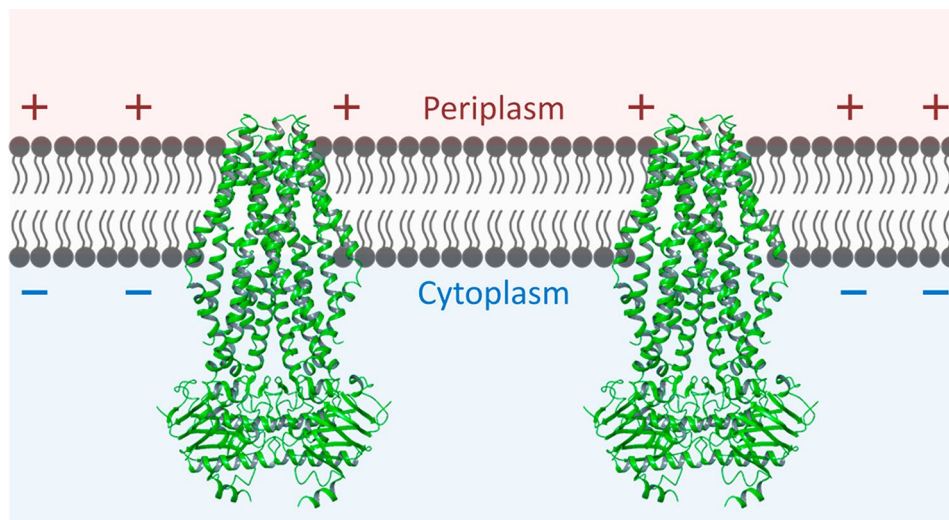


Fig. 1 ABC Transporter Proteins. Structure of the ABC transporter protein in the bacterial inner membrane. The extracellular space (and by extension, the periplasmic space) is less negative than the cytoplasm [26]. This ABC transporter can be used in extracellular secretion-based protein production

differences between Glu and Asp mutations. Cationic-to-anionic mutation was chosen for efficiency of supercharging). When the process was completed for every window, the cationic supercharged region would have been removed, resulting in a T1SS-secretable amino acid sequence.

However, this simple process of mutating amino acids based on LCD results had a limitation: changing an amino acid sequence by only using charge density information could cause losses in protein structure or function. Therefore, the effect of an amino acid mutation on a protein's stability and functionality had to be calculated for consideration, which could be done by AvNAPSA (average number of neighboring atoms per side chain atom) and Consurf, respectively.

The tool we developed for linear charge density analysis is accessible on our webserver: <https://mb.re.kr/apps/LCD>.

AvNAPSA implementation

A supercharging method, AvNAPSA (Liu et al., 2007), mutates flexible polar residues (DERKNQ) with the fewest average neighboring atoms per side chain atom [27]. We rewrote the original Perl script from the 2007 study in Python. Since AvNAPSA scores signify whether the amino acid residue has less interaction with other residues and is on the surface of the protein (a lower AvNAPSA score means fewer neighboring atoms and therefore less interaction), we utilized the AvNAPSA score to determine the effect on protein stability upon mutation. The positive-charge amino acids determined by LCD analysis were now only mutated if they had an AvNAPSA score in a certain range. An AvNAPSA cutoff of <150 has been widely used in heavy supercharging

to maintain protein stability while also allowing enough mutations to occur (a more stringent cutoff of <100 is commonly used in moderate supercharging) [28].

Consurf implementation

Consurf is a tool that traces the evolutionary history of amino acids in polypeptides and identifies conserved regions that are important to the protein function [29–31]. Since the mutation of amino acids according to the LCD analysis or AvNAPSA score could damage polypeptide regions critical to protein function, we utilized conservation scores from the Consurf webserver results. Now, when a residue had a conservation score higher than 5, it was considered important to protein function and excluded from possible mutation by LCD analysis results.

Final PySupercharge algorithm

The final version of our PySupercharge algorithm incorporates both LCD analysis and protein function/stability loss minimization methods (Fig. 2). Firstly, the program accepts the Protein Database (.pdb extension file; 'PDB format') and Consurf files (consurf_grades.txt; obtained by running process at Consurf website) for the protein of interest. Secondly, it operates LCD analysis with the established window size. If a cationic supercharged region is identified, Arg and Lys residues in that region are found and remembered. Thirdly, the program calculates the AvNAPSA score of such flagged residues and stores those that have a score in the user-defined range. Then, the algorithm sorts these residues based on the Consurf score so that only those with a score less than or equal to 5 (hence, highly variable and deemed functionally unimportant) are considered for mutation. Finally,

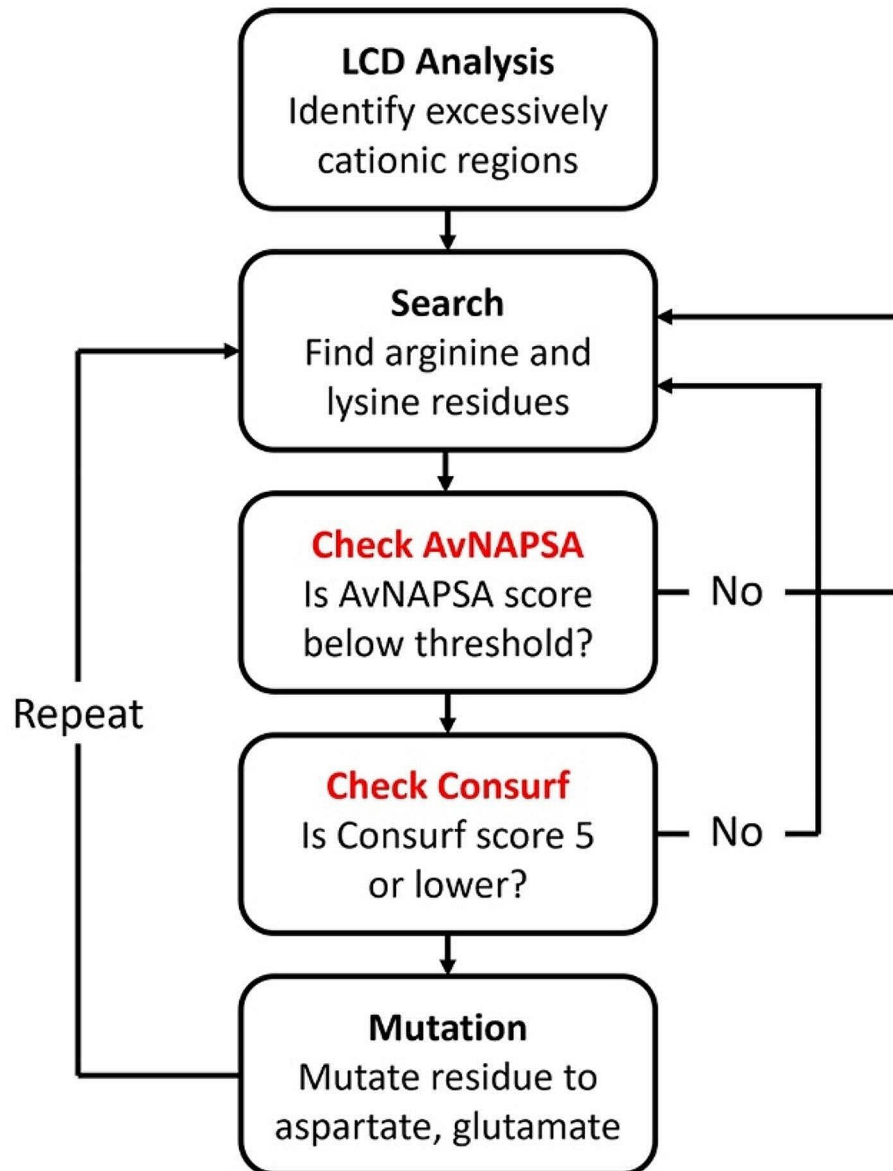


Fig. 2 Logic scheme of PySupercharge. How the PySupercharge algorithm supercharges protein sequences. PySupercharge takes amino acids' AvNAPSA and ConSurf scores into account before mutating them

the actual mutations from the approved Arg and Lys residues to Glu or Asp residues are carried out.

This final product, the PySupercharge algorithm, is available for use on our laboratory webserver, <https://mb.re.kr/apps/supercharge/>. We included an example of GFP supercharging on the webserver and in an additional file [see Additional file 3].

Protein expression analysis

We mutated multiple originally non-secretable proteins using both manual and PySupercharge-aided mutation to analyze secretion ability in *fleQ*-knockout *P. fluorescens*. MFP and OMP were co-expressed with the ABC protein in these cells for the secretion of proteins. Nitrocellulose membranes (Amersham, Germany) were used 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). The genes for protein expression were synthesized using Bionics (<http://www.bionicsro.co.kr/>) gene synthesis service. The C-terminal LARD3 secretion signal of thermostable lipase TliA was appended to every protein secreted in this study, recognized by the *P. fluorescens* TliDEF ABC transporter. (A western blot showing the secreted LARD3-only control is in an additional file [see Additional File 4]) Protein expression was carried out through constitutive expression. The amino acid sequences for all proteins used in this research are in the additional files [see Additional file 1, Additional file 2].

Results

Secretion of human growth factors by manual mutation

Various non-secretable human growth factors, namely transforming growth factor beta (TGF β), fibroblast growth factor I (FGF1), insulin-like growth factor I and II (IGF1 and IGF2), and beta-nerve growth factor (β NGF) were mutated manually before the development of PySupercharge and tested for secretion ability. The manual mutation process was simple; cationic supercharged regions were identified through LCD analysis and positive amino acids within the region were converted to negative amino acids. This approach proved successful for short sequences within growth factors but did not take function loss into account and was arduous for larger protein sequences (Fig. 3). We used AlphaFold 2 to estimate the structure of the post-mutation proteins [32]. The AlphaFold2-predicted structures of the proteins before and after mutation are included in an additional file [see Additional File 6].

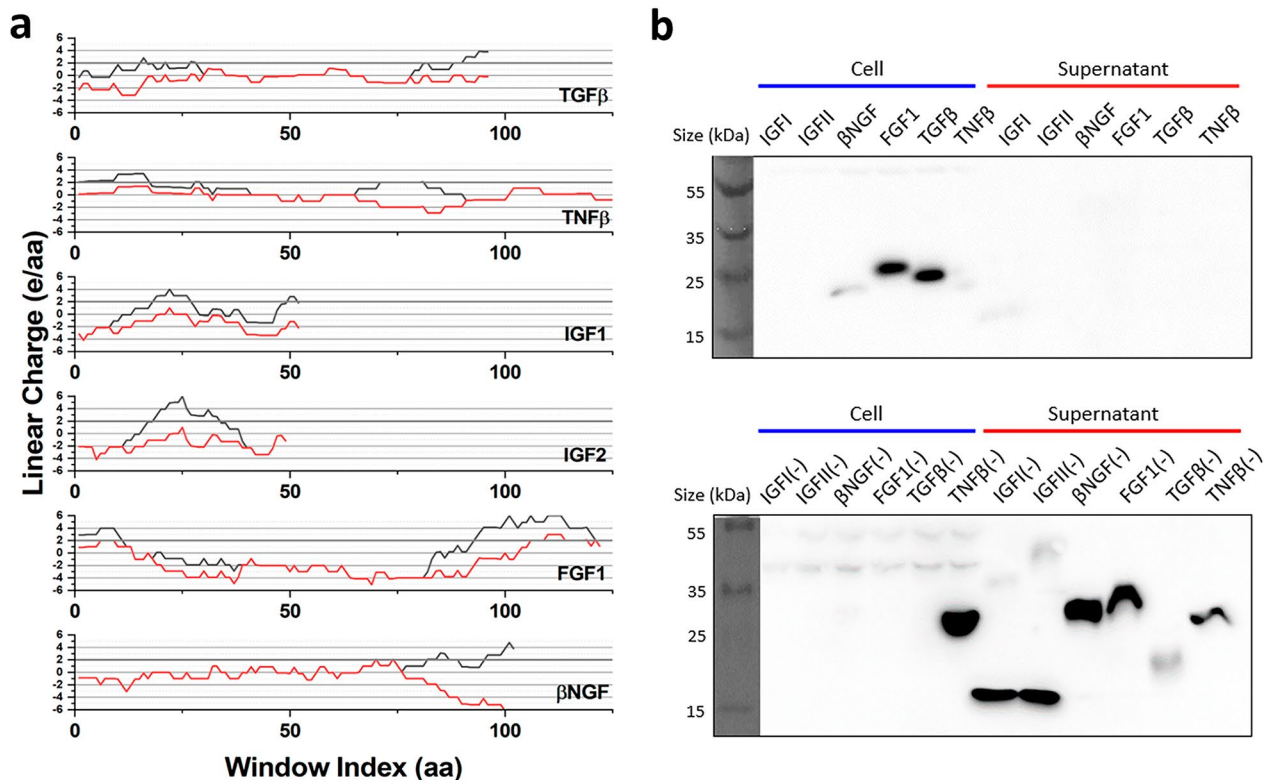


Fig. 3 Secretion results of manually supercharged human growth factors. **(a)** LCD analysis results of wildtype sequences (shown in grey) and manually supercharged sequences (shown in red) of various human growth factors. Each plotted point matches an amino acid position to the LCD value of the 20-amino-acid window starting from that position. **(b)** Adapted from "Utilizing the ABC Transporter for Growth Factor Production by *fleQ* deletion mutant of *Pseudomonas fluorescens*", Fabia et al., *Biomedicines* 2022, 9(6), p. 679 [2]. Adapted with permission- label sizes and colors have been modified. The western blotting results of manually mutated growth factors (bottom row) show enhanced secretion ability compared to the wildtype proteins (top row). A separate high-exposure western blot of wildtype IGF1 and IGF2 is in an additional file [see Additional File 4] due to low visibility in this figure

Secretion of SARS-CoV-2 spike proteins with PySupercharge

The two domains of the SARS-CoV-2 spike protein homotrimer S1 (PDB ID 6ZP0) are the N-terminal domain (NTD) and receptor binding domain (RBD). LCD analysis identified two cationic supercharged regions of LCD value over 2 in each domain, which could possibly inhibit bacterial secretion. To investigate the correlation between the linear charge density and secretion ability, both NTD and RBD sequences were negatively supercharged to two LCD cutoff values: ≤ 2 and ≤ 1 (Fig. 4). In the case of the NTD, there was a total of 26 arginine and lysine residues within the coding region. Among these residues, PySupercharge mutated 6 and 5 for NTD and RBD respectively. The AlphaFold2-predicted structures of the wildtype, LCD ≤ 2 and LCD ≤ 1 versions of NTD and RBD are included in an additional file [see Additional File 6].

Analyzing the western blotting results for both the NTD and RBD, supercharged proteins are found in much greater amounts in the supernatant after secretion than the wildtype proteins (Fig. 4c). This implies the greater secretion ability of supercharged proteins. Notably, LCD ≤ 1 cutoff mutant proteins are shown to have

considerably better secretion ability than those of the ≤ 2 cutoff value. SDS-PAGE data for NTD and RBD is in an additional file [see Additional File 5].

Secretion of botulinum toxin light chain (BoNT/A)

Among the 7 serotypes of botulinum toxin, type A is the most widely used in both academia and the therapeutic industry. This neurotoxin is consisted of a heavy chain that enables internalization of the toxin into the presynaptic terminal and a light chain that cleaves synaptosome-associated proteins (SNAP25) [33]. LCD analysis of the light chain, BoNT/A, identified 6 cationic supercharged regions with the LCD greater than 2. Two mutants, BoNT (LCD ≤ 2) and BoNT (LCD ≤ 1), were created with the LCD cutoff value of 2 and 1 respectively using PySupercharge. BoNT (LCD ≤ 1) was mutated with stricter criteria to compare the secretion ability. The AlphaFold2-predicted structures of BoNT, BoNT (LCD ≤ 2) and BoNT (LCD ≤ 1) are included in an additional file [see Additional File 6].

The Western blotting results for BoNT in the cell and supernatant show that the secretion ability of both of the two supercharged proteins is much greater than that of the wildtype protein (Fig. 5). The LCD ≤ 1 cutoff mutant

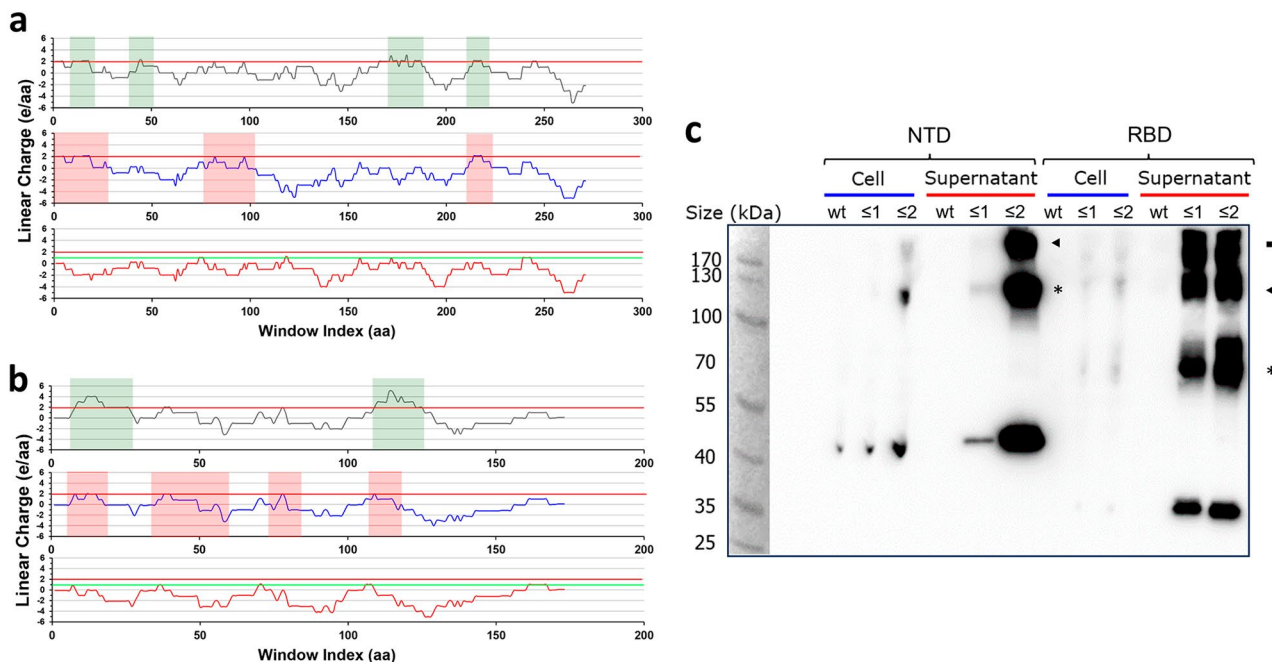


Fig. 4 Secretion of automatically supercharged SARS-CoV-2 S1 protein NTD and RBD. **(a)** Top graph: LCD analysis result of wildtype SARS-CoV-2 S1 protein NTD sequence (cationic supercharged regions of LCD > 2 shaded green). Middle graph: Sequence mutated by PySupercharge to make NTD (LCD ≤ 2). Note that all the shaded regions from the top graph have been modified to stay below or on the LCD = 2 reference line. LCD > 1 regions have been shaded in red for comparison with NTD (LCD ≤ 1). Bottom graph: Sequence of NTD (LCD ≤ 1), mutated in a similar manner but with a cutoff value of 1. The LCD values for the SARS-CoV-2 spike protein domains were calculated with histidine charge configuration +0.1. All of the regions have been mutated to stay below or on the green LCD = 1 reference line. **(b)** Similar set of graphs for SARS-CoV-2 S1 protein RBD. **(c)** The western blot analysis results of the NTD (1st–6th lane) and RBD (7th–12th lane). Supercharged NTD and RBD sequences clearly display greater secretion ability. A separate high-exposure western blot of wildtype RBD is in an additional file [see Additional File 4] due to low visibility in this figure. (*, dimer protein; ▶, trimer protein; ■, tetramer protein)

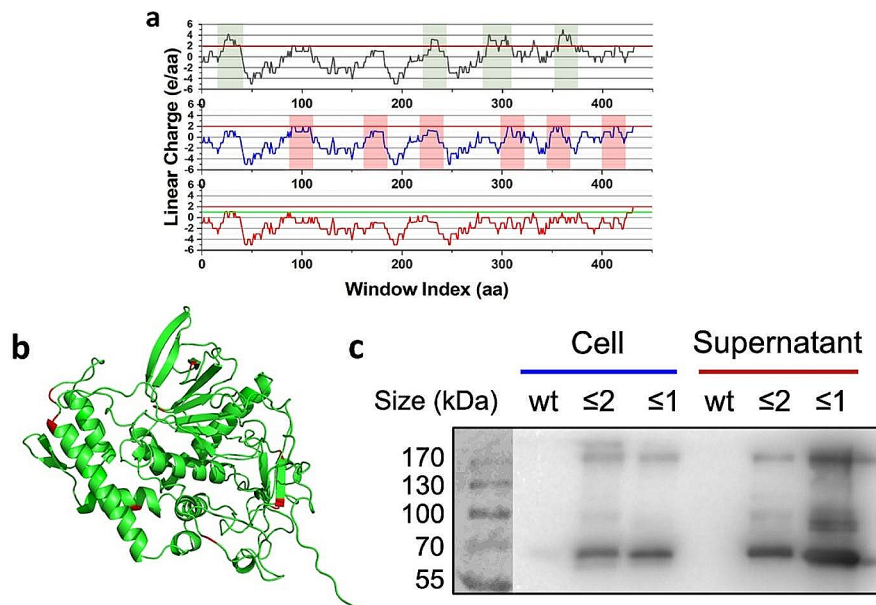


Fig. 5 Secretion of supercharged botulinum toxin type A. **(a)** Top graph: LCD analysis result of BoNT wildtype sequence (cationic supercharged regions of $LCD > 2$ shaded green). Middle graph: Mutated sequence of BoNT ($LCD \leq 2$). Note that all the shaded regions from the top graph have been modified to stay below or on the red $LCD = 2$ reference line. $LCD > 1$ regions have been shaded in red for easy comparison with BoNT ($LCD \leq 1$). Bottom graph: Mutated sequence of BoNT ($LCD \leq 1$). Note that all the shaded regions from the top graph have been modified to stay below or on the green $LCD = 1$ reference line. **(b)** BoNT/A wildtype protein structure with residues mutated in BoNT ($LCD \leq 2$) highlighted in red. **(c)** The western blot analysis results of BoNT. Supercharged BoNT ($LCD \leq 2$) and BoNT ($LCD \leq 1$) clearly display higher secretion ability, the latter being even higher than the former

performs even better than the ≤ 2 mutant, similar to the SARS-CoV-2 spike protein analysis. SDS-PAGE data for BoNT is in an additional file [see Additional File 5].

Discussion

Protein supercharging can improve secretion of proteins, enabling them to be produced through bacterial secretion systems using ABC transporters. In most previous supercharging-related studies, the focus was on the surface charge of the protein [28, 34]. We focused instead on the charge of the unfolded polypeptide chain, modifying it to aid in its secretion. In our previous study, we defined a new criterion for protein supercharging, the linear charge density. Based on the LCD calculations, protein sequences were manually mutated and tested for secretion ability. Then we took a step further, automating the LCD analysis and mutation process. We also implemented two separate tools, AvNAPSA and Consurf, in our LCD analysis.

AvNAPSA and Consurf are two analysis processes utilized to prevent unwanted effects from mutation by LCD analysis. Since mutation using only LCD analysis is entirely dependent on amino acid charge, structural stability or functionality could be greatly disrupted by mutation. AvNAPSA was originally designed for finding and mutating surface residues to manipulate protein surface charge [27]. We translated it to Python and used

it to separate the surface and inner residues, now only mutating these surface residues to minimize the destabilization of protein structure by high mutation load. Consurf traces the evolutionary history of the residues and returns the conservation score [29–31]. A residue with a high conservation score has a higher chance of being critical to protein function, and therefore is exempt from mutation.

PySupercharge aims to minimize protein function loss through exclusion of evolutionarily conserved residues from mutation. However, it does not guarantee preservation of protein activity. Additionally, our present algorithm does not consider the types and significance of the chemical bonds formed around particular amino acid residues. So, we risk mutating a surface residue that had formed important hydrogen bonds when it was present. This could result in the loss of structural integrity. Therefore, a potential direction for further development of the algorithm would be to refine the mutation candidate selection to residues not engaging in hydrogen bonding with others. As for the current algorithm, we have added an option for excluding user-selected residues in the supercharging process by their indexes (e.g., a user input of ‘35, 136’ in the Exclusion field prevents the 35th and 136th residues from being mutated). Users may choose to exclude certain residues if they are predicted to be functionally significant.

Overall, we have developed PySupercharge, a one-click tool to mutate amino acid sequences based on linear charge density analysis. Given the PDB and Consurf files of the target protein, PySupercharge can also calculate AvNAPSA and Consurf scores and use them to maintain structural stability and exclude functionally important residues from mutation, respectively. We verified that non-secretable protein sequences mutated with PySupercharge showed increased secretion efficiency, and so suggest this tool for future research in effective protein secretion. Our tool is available on the webserver <https://mb.re.kr/apps/supercharge/>.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12934-024-02342-z>.

Additional file 1: Amino acid sequences of wildtype proteins expressed. Sequences of all of the wildtype proteins we expressed and secreted in the study.

Additional file 2: Amino acid sequences of negatively supercharged proteins. Sequences of all of the supercharged proteins (either manually or by PySupercharge) we expressed and secreted in the study.

Additional file 3: Example negative supercharging and LCD analysis of green fluorescent protein (GFP). Amino acid sequences for wildtype GFP (upper text) and negatively supercharged GFP (lower text). Four amino acids have been mutated. LCD analysis graph of the same (wildtype: blue, LCD \leq 2: orange). A 3D model of the wildtype GFP with AvNAPSA <150 residues highlighted in red is shown below.

Additional file 4: High-exposure western blotting images of wildtype proteins. High-exposure western blotting images of wildtype IGF1, IGF2, and SARS-CoV-2 RBD.

Additional file 5: SDS-PAGE images of SARS-CoV-2 domains and BoNT. SDS-PAGE data for supercharged SARS-CoV-2 domains and BoNT in the study.

Additional File 6: AlphaFold2-generated structures of wildtype and supercharged proteins. Superimposed AlphaFold2-generated protein structures of all proteins in the study.

Acknowledgements

Not applicable.

Author contributions

YK wrote and edited the new online version of the Python algorithm. DK contributed to hosting the algorithm on the webserver and was a major contributor in writing and editing the manuscript. NMH wrote the original offline version of the algorithm and was a major contributor in writing the manuscript. HB performed linear charge density analysis on the proteins mentioned in this study and generated the linear charge density graph data shown for each protein. JHA was a contributor in writing and editing the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by Korea Science Academy of KAIST with funds from the Ministry of Science and ICT.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 26 November 2023 / Accepted: 19 February 2024

Published online: 20 April 2024

References

- Byun H, Park J, Kim SC, Ahn JH. A lower isoelectric point increases signal sequence-mediated secretion of recombinant proteins through a bacterial ABC transporter. *J Biol Chem*. 2017. <https://doi.org/10.1074/jbc.M117.786749>.
- Fabia B-U, Bingwa J, Park J, Hieu N-M, Ahn J-H. Utilizing the ABC transporter for growth factor production by fleQ deletion mutant of *Pseudomonas fluorescens*. *Biomedicines*. 2021;9(6):679. <https://doi.org/10.3390/biomedicines9060679>. PubMed PMID: .
- Holland IB, Schmitt L, Young J. Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway (review). *Mol Membr Biol*. 2005;22(1–2):29–39. Epub 2005/08/12. doi: 10.1080/09687860500042013. PubMed PMID: 16092522.
- Chen R. Bacterial expression systems for recombinant protein production: *E. Coli* and beyond. *Biotechnol Adv*. 2012;30(5):1102–7. Epub 2011/10/05. .013. PubMed PMID: 21968145.
- Sorensen HP, Mortensen KK. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb Cell Fact*. 2005;4(1):1. <https://doi.org/10.1186/1475-2859-4-1>. Epub 2005/01/05.
- Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol*. 2014;5:172. <https://doi.org/10.3389/fmicb.2014.00172>. Epub 2014/05/27.
- Slade KM, Baker R, Chua M, Thompson NL, Pielak GJ. Effects of recombinant protein expression on green fluorescent protein diffusion in *Escherichia coli*. *Biochemistry*. 2009;48(23):5083–9. <https://doi.org/10.1021/bi9004107>. Epub 2009/05/06.
- Freudl R. Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb Cell Fact*. 2018;17(1):52. <https://doi.org/10.1186/s12934-018-0901-3>. Epub 2018/03/31.
- Kleiner-Grote GRM, Risse JM, Friehs K. Secretion of recombinant proteins from *E. Coli*. *Eng Life Sci*. 2018;18(8):532–50. <https://doi.org/10.1002/elsc.201700200>. Epub 2018/04/14.
- 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(2):e2911.
- Green ER, Meccas J. Bacterial Secretion Systems: An Overview. *Microbiol Spectr*. 2016;4(1). Epub2016/03/22.10.1128/microbiolspec.VMBF-0012-2015. PubMed PMID: 26999395; PubMed Central PMCID: PMC4804464.
- Thomas S, Holland IB, Schmitt L. The type 1 secretion pathway - the hemolysin system and beyond. *Biochim Biophys Acta*. 2014;1843(8):1629–41. <https://doi.org/10.1016/j.bbamcr.2013.09>. Epub 2013/10/17.
- Letoffe S, Delepelaire P, Wandersman C. Protein secretion in gram-negative bacteria: assembly of the three components of ABC protein-mediated exporters is ordered and promoted by substrate binding. *EMBO J*. 1996;15(21):5804–11.
- Young J, Holland IB. ABC transporters: bacterial exporters-revisited five years on. *Biochim et Biophys Acta (BBA)-Biomembranes*. 1999;1461(2):177–200.
- Ahn JH, Pan JG, Rhee JS. Identification of the tliDEF ABC transporter specific for lipase in *Pseudomonas fluorescens* SIK W1. *J Bacteriol*. 1999;181(6):1847–52. Epub 1999/03/12. PubMed PMID: 10074078; PubMed Central PMCID: PMC93584.
- Park J, Eom GT, Oh JY, Park JH, Kim SC, Song JK, et al. High-level production of bacteriotoxic phospholipase A1 in bacterial host *Pseudomonas fluorescens* Via ABC transporter-mediated secretion and Inducible expression. *Microorganisms*. 2020;8(2). <https://doi.org/10.3390/microorganisms8020239>. Epub 2020/02/15.
- Park Y, Moon Y, Ryoo J, Kim N, Cho H, Ahn JH. Identification of the minimal region in lipase ABC transporter recognition domain of *Pseudomonas fluorescens* for secretion and fluorescence of green fluorescent protein. *Microb Cell Fact*. 2012;11:60. <https://doi.org/10.1186/1475-2859-11-601475-2859-11-60>. Epub 2012/05/15.

18. Ryu J, Lee U, Park J, Yoo DH, Ahn JH. A vector system for ABC transporter-mediated secretion and purification of recombinant proteins in *Pseudomonas* species. *Appl Environ Microbiol*. 2015;81(5):1744–53. <https://doi.org/10.1128/AEM.03514-14>. Epub 2014/12/31.
19. Son M, Moon Y, Oh MJ, Han SB, Park KH, Kim JG, et al. Lipase and protease double-deletion mutant of *Pseudomonas fluorescens* suitable for extracellular protein production. *Appl Environ Microbiol*. 2012;78(23):8454–62. <https://doi.org/10.1128/AEM.02476-12>. PubMed PMID: 23042178; PubMed Central PMCID: PMC3497380.
20. Dalbey RE, Kuhn A. Protein traffic in Gram-negative bacteria—how exported and secreted proteins find their way. *FEMS Microbiol Rev*. 2012;36(6):1023–45.
21. Burdette LA, Leach SA, Wong HT, Tullman-Ercek D. Developing Gram-negative bacteria for the secretion of heterologous proteins. *Microb Cell Fact*. 2018;17(1):196. <https://doi.org/10.1186/s12934-018-1041-5>. Epub 2018/12/24.
22. Chung CW, You J, Kim K, Moon Y, Kim H, Ahn JH. Export of recombinant proteins in *Escherichia coli* using ABC transporter with an attached lipase ABC transporter recognition domain (LARD). *Microb Cell Fact*. 2009;8(1):11. <https://doi.org/10.1186/1475-2859-8-11>.
23. Bakkes PJ, Jenewein S, Smits SH, Holland IB, Schmitt L. The rate of folding dictates substrate secretion by the *Escherichia coli* hemolysin type 1 secretion system. *J Biol Chem*. 2010;285(52):40573–80. <https://doi.org/10.1074/jbc.M110.173658>. Epub 2010/10/26.
24. Khosa S, Scholz R, Schwarz C, Trilling M, Hengel H, Jaeger KE, et al. An A/U-Rich enhancer region is required for high-level protein secretion through the HlyA type I Secretion System. *Appl Environ Microbiol*. 2018;84(1). <https://doi.org/10.1128/AEM.01163-17>. PubMed PMID: 29030442; PubMed Central PMCID: PMC5734041. Epub 2017/10/17.
25. Byun H, Park J, Fabia BU, Bingwa J, Nguyen MH, Lee H et al. Generalized Approach towards Secretion-Based Protein Production via Neutralization of Secretion-Preventing Cationic Substrate Residues. *Int J Mol Sci*. 2022;23(12). Epub 20220615. <https://doi.org/10.3390/ijms23126700>. PubMed PMID: 35743142; PubMed Central PMCID: PMC9223453.
26. Novo D, Perlmutter NG, Hunt RH, Shapiro HM. Accurate flow cytometric membrane potential measurement in bacteria using diethyloxycarbocyanine and a ratiometric technique. *Cytometry*. 1999;35(1):55–63. [https://doi.org/10.1002/\(sici\)1097-0320\(19990101\)35:1<55::aid-cyto8>3.0.co;2-2](https://doi.org/10.1002/(sici)1097-0320(19990101)35:1<55::aid-cyto8>3.0.co;2-2). Epub 1999/11/30.
27. Lawrence MS, Phillips KJ, Liu DR. Supercharging proteins can impart unusual resilience. *J Am Chem Soc*. 2007;129(33):10110–2.
28. Der BS, Kluwe C, Miklos AE, Jacak R, Lyskov S, Gray JJ, et al. Alternative computational protocols for supercharging protein surfaces for reversible unfolding and retention of stability. *PLoS ONE*. 2013;8(5):e64363. <https://doi.org/10.1371/journal.pone.0064363>. Epub 2013/06/07.
29. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016;44(W1):W344–W50.
30. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*. 2010;38(suppl2):W529–W33.
31. Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, et al. ConSurf: using evolutionary data to raise testable hypotheses about protein function. *Isr J Chem*. 2013;53(3–4):199–206.
32. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9. Epub 2021/07/16. <https://doi.org/10.1038/s41586-021-03819-2>. PubMed PMID: 34265844; PubMed Central PMCID: PMC8371605 have filed non-provisional patent applications 16/701,070 and PCT/EP2020/084238, and provisional patent applications 63/107,362, 63/118,917, 63/118,918, 63/118,921 and 63/118,919, each in the name of DeepMind Technologies Limited, each pending, relating to machine learning for predicting protein structures. The other authors declare no competing interests.
33. Dressler D, Saberi FA. Botulinum toxin: mechanisms of action. *Eur Neurol*. 2005;53(1):3–9.
34. Thompson DB, Cronican JJ, Liu DR. Engineering and identifying supercharged proteins for macromolecule delivery into mammalian cells. *Methods Enzymol*. 2012;503:293–319. <https://doi.org/10.1016/B978-0-12-396962-0.00012-4>. Epub 2012/01/11.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.