How to cite this article: Ahmadi Z, Farajnia S, Farajzadeh D, Poladi N, Pourvatan N, Karbalaeimahdi M, Shayegh F, Arya M. Optimized Signal peptide for Secretory Expression of Human Recombinant Somatropin in E. coli. Advanced Pharmaceutical Bulletin, doi: 10.34172/apb.2023.037

Research Article

Optimized Signal peptide for Secretory Expression of Human Recombinant Somatropin in E. coli

Zeynab Ahmadi[1], Safar Farajnia*[2,3], Davoud Farajzadeh*[1,4], Naser Poladi[1], Neda Pourvatan[5], Mohammad Karbalaeimahdi[3], Fahime Shayegh[2], Maryam Arya[3]

[1]Department of Biology, Faculty of Basic Sciences, Azarbaijan Shahid Madani University, Tabriz, Iran.
[2]Drug Applied Research Center, Tabriz University of Medical Science, Tabriz, Iran.
[3]Biotechnology Research Center, Tabriz University of Medical Science, Tabriz, Iran.
[4]Department of Molecular Biology and Cancer Research, Azarbaijan Shahid Madani University, Tabriz, Iran.
[5]Immunology Research Center, Tabriz University of Medical Science, Tabriz, Iran.

Davoud Farajzadeh : https://orcid.org/0000-0002-9267-4277

Safar Farajnia : https://orcid.org/0000-0002-6087-9147

*Corresponding authors: Safar Farajnia and Davoud Farajzadeh, Email: farajnias@tbzmed.ac.ir and Email: farajzadehd@gmail.com

Submitted: 15 November 2021
Revised: 20 January 2022
Accepted: 31 March 2022
ePublished: 04 April 2022

Running title: Signal Peptide for Secretory Expression of Human Recombinant Somatropin

Abstract

Purpose: The human somatropin is a single-chain polypeptide with a pivotal role in various biological processes. Although *E. coli* is considered as a preferred host for the production of human somatropin, the high expression of this protein in *E. coli* results in the accumulation of protein as inclusion bodies. Periplasmic expression using signal peptides could be used to overcome the formation of inclusion bodies; still, the efficiency of each of the signal peptides in periplasmic transportation is varied and often is protein specific. The present study aimed to use In-silico analysis to identify an appropriate signal peptide for the periplasmic expression of human somatropin in *E. coli*.

Methods: A library containing 90 prokaryotic and eukaryotic signal peptides were collected from the signal peptide database, and each signal's characteristics and efficiency in connection

with the target protein were analyzed by different software. The prediction of the secretory pathway and the cleavage position was determined by the signalP5 server. Physicochemical properties, including molecular weight, instability index, gravity, and aliphatic index, were investigated by protparam software.
Results: The results of the present study showed that among all the signal peptides studied, five signal peptides ynfB, sfaS, lolA, glnH, and malE displayed high scores for periplasmic expression of human somatropin in *E. coli*, respectively.
Conclusion: In conclusion, the results indicated that in-silico analysis could be used for the identification of suitable signal peptides for the periplasmic expression of proteins. Further laboratory studies can evaluate the accuracy of the results of insilico analysis.

Keywords: Human Somatropin, Signal peptide, *E. coli*, Secretary Expression

Introduction

Human Somatropin is a non-glycosylated single-chain polypeptide comprising of 191 amino acids, with a molecular mass of 22.1 kDa.[1] Somatropin belongs to the somatotropin/prolactin family, which plays a significant role in growth control through stimulating various tissues, mainly the liver, to secrete Insulin-like growth factor 1 (IGF-1). Besides, it is responsible for the differentiation and proliferation of myoblasts, the uptake of amino acids, and proteins' production in muscles and other tissues.[2]
Advantages such as easy genetic manipulation, low-cost media, and short culturing time have led to the use of *Escherichia coli* (*E.* coli) as the most suitable expression system for the production of many recombinant proteins.[3] However, high level expression of recombinant proteins in *E. coli* often give rise to aggregated protein molecules, known as inclusion bodies.[4] Therefore, recombinant proteins' purification encounters significant challenges, involving isolation from the cells, unfolding, refolding, and purification to produce the bioactive proteins. Various strategies have been used to overcome this problem include secretary expression by targeting the protein into the periplasmic space by an N-terminal signal peptide.[5]
Sec, SRP, and TAT are major protein secretion pathways used by prokaryotes by which proteins direct into the periplasm or extracellular space according to their signal peptides (Signal peptides).[6] Therefore, selecting an appropriate signal peptide is an aessential parameter in the secretory expression of recombinant proteins.[7] Several studies have shown that the function of signal peptides is protein-specific, and there is no unique ideal signal peptide for secretary expression of all proteins.[8] A conventional method for selecting a signal peptide for a given protein is trial and error, which is labor-intensive and time-consuming. Recently various bioinformatics programs have been developed for the analysis of the efficiency of different signal peptides, which include signalP4.1, ProtParam, SOLpro, ProtCompB, and signalP5.0. The advantages of using a bioinformatics program before starting an experimental study are reducing costs and increasing the accuracy and validity of experimental research.[9]
Secretory expression of recombinant proteins, particularly pharmaceutical proteins, in *E. coli* has many advantages. Targeting a recombinant protein to the periplasmic space or the extracellular medium, in addition to reducing costs, facilitates downstream processing, compared to the cytosolic production.[10]
The purpose of the present study was to *in silico* analysis of various signal peptides for secretary expression of Somatropin using different bioinformatic programs.

Materials and Methods

Signal peptide sequences

In this research, sequences of 90 different signal peptides were collected from the Signal Sequence Database at www.signalpeptide.de/ (Table 1) and used for further analyses.

In Silico prevision of signal peptide and prediction of h, c and n Regions

SignalP software version 4.1(www.cbs.dtu.dk/services/SignalP-4.1/) was used for the prediction of signal peptides and their sites of cleavage based on the combination of different artificial neural networks.[11] SignalP online software version 3.0 was (http://www.cbs.dtu.dk/services/SignalP-3.0/) employed for predicting n, h, and c regions of signal peptides. For this purpose, signal peptides were added to the Somatropin sequence and analyzed by the program.

Analysis of physicochemical features of signal peptides

The ProtParam program was used to evaluate the physicochemical features of the signal peptides including, theoretical pI, amino acid composition, negatively and positively charged amino acids, grand average of hydropathicity (GRAVY), instability index, aliphatic index, and molecular weight.

Analysis of protein solubility

SOLpro tool predicts the solubility of a protein upon expression in *E. coli* based on characterisstics of primary sequences. Therefore, the SOLpro at www. scratch.proteomics.ics.uci.edu/, was used to determine the protein solubility in *E. coli*. SOLpro tool has a prediction accuracy of above 74%.

Prediction of protein localization

ProtComp B server, from Softberry, Inc (http://www.softberry.com), was applied for prediction of Somatropin destination in connection with various signal peptides. It accomplishes this job using a composition of sequence homology and neural networks.[12]

Prediction of the type of signal peptides and cleavage Probability

In prokaryotes, there are three types of signal peptides, including Sec pathway cleaved by either SPase I (Sec/SPI) or SPase II (Sec/SPII), and Tat pathway cleaved by Tat/SPI.[13] SignalP5.0 server was used for discrimination of three types of signal peptides.[14] SignalP 5.0 predicts the type of signal peptides based on a deep convolutional and recurrent neural network architecture.[15] The cleavage probability was also determined by SignalP 5.0 program.

Results and Discussion

In Silico prediction of signal peptide and determination of c, h, and n regions

SignalP 4.1 was applied for prediction of the most suitable signal peptide for Somatropin, enabling its secretion into the periplasmic space in *E. coli*. SignalP 4.1 identifies a signal peptide based on a discriminating score, D-score. The output was tabulated in Table 2, containing five scores of D, C, S, Y, S-mean including cleavage sites and c, h and n regions of signal peptides.

Thirty-six signal peptides were deleted from further analysis because the D-scores of them were less than the cut off value of 0.570, indicating that they are not efficient for the secretion of Somatropin protein.

Among the analyzed 90 signal peptides, four signal peptides, including pelB, flgl, nmpc, and, gfcA showed the highest D-score value of 0.910, 0.907, 0.902, and 0.902, respectively. Moreover, the results demonstrated that pelB and NPPC have the highest D-score in

prokaryotic and eukaryotic expression systems, respectively. Additionally, the lowest scores belonged to HBP and LEAP2 (0.175, 0.208) in prokaryotic and eukaryotic expression systems, respectively.

Physico-chemical features of signal peptides

Several physicochemical features of 55 remaining signal peptides containing, theoretical pI length, molecular weight, net positive charge, grand average of hydropathicity (GRAVY), instability index and aliphatic index were evaluated by ProtParam server (Table 3). The results showed that the length of signal peptides was between 18 and 28 residues. The results of in silico analysis revealed that the highest molecular weight pertained to ynfB, bcsB, lptA, and efeO (2948.71, 2853.53, 2849.47, and 2845.33 daltons, respectively).

The most hgigh GRAVY values were belonged to signal peptides flgI, thiB, OmpC and yncJ (1.935, 1.589, 1.552, and 1.541, respectively). The highest aliphatic index scores belonged to flgl, ompC, NPPC, mepA, and cysP (185.50, 171.90, 165.65, 164.74, and 164.00, respectively)

Another evaluated physicochemical feature of signal peptides was the instability index. The results demonstrated that papK, yhcN, ansB, and pilC (-2.60, -2.03, -1.15, and 1.01, respectively) were the most stable signal peptides, separately and in connection with Somatropin. The proteins whose instability index was higher than 40 were predicted as unstable, and the values under 40 might be stable.

Prediction of protein solubility

The results of Somatropin solubility in fusion with various signal peptides have shown in Table 3. The results demonstrated that the highest solubility were belonged to lamb, draA, faeG, nmpc, rbsB, and malE signal peptides (0.889, 0.885, 0.883, 0.883, 0.879, and 0.879, respectively).

Prediction of the protein localization

The analysis results for sub-cellular localization by ProtCompB server indicated that the final localization sites were the outer membrane, inner membrane, and periplasmic space for 13, 15, and 18 signal peptides, respectively. Furthermore, analysis for the final localization of Somatropin with signal peptides faeG, FimF41a, ompA, papK, prsK, lamb, nmpc, bcsB, and gfcA revealed that Somatropin could be sevcreted by these signal peptides (Table 4).

Prediction of cleavage probability and the type of signal peptides

The remaining 55 signal peptides were examined for their secretory pathway(s) by using signal P5.0 software. The results showed that except efeO (TAT pathway) and ampC (sec/SPII), all of these signal peptides were specific for the Sec/SPI pathway (Table 4). The cleavage probability of each signal peptides was tabulated in Table2.

Selection of appropriate signal peptide

First, the signal peptides with final localization in periplasmic space was selected and sorted according to the aliphatic index. Then, the stability and solubility of target protein in connection with the selected signals was examined. The signal peptides with which Somatropin remained stable and soluble were selected as the appropriate peptide signal (Table 5).

E. coli is the economical and straightforward host for the expression of recombinant proteins.[16] However, overexpression of recombinant proteins in the intracellular space of *E. coli* is usually associated with insoluble aggregate and inclusion body formation. To keep appropriate folding, the proteins should be avoided from the reductive environment of the cytoplasm. Hence, the

secretory expression has several advantages for the production of recombinant proteins, compared with cytosolic systems.

The secretion of the target protein requires transporting across the cytoplasmic membrane. In bacteria, Sec, SRP, and TAT are three major protein secretion pathways for the carriage of proteins through the plasma membrane. These protein transport systems depend on the presence of suitable signal peptides on proteins. Signal peptides are short amino terminal peptides that affect the biosynthesis, folding, and stability of the corresponding target proteins.[17] Although various signal peptides differ in their sequences, they share conserved physicochemical properties, including aliphatic index, molecular weight, instability index, Gravy, net positive charge, and theoretical pI. The three important regions of signal peptides include an amino terminal positively-charged region, a hydrophobic central region, and a carboxyl-terminal polar region that contains the cleavage site (a conserved A-X-A motif). It has demonstrated that the n region in the signal peptide has an essential role in the primary phase of protein secretion across the membranes.[18] Also the n-region responsible for the net positive charge of the signal peptide. In addition, the presence of the basic residues in this region may be indispensable for the performance of an efficient signal peptide.[19]

Further to the charge of the n-region, the c-region has an intense effect on the performance of membrane transport by both the Tat and Sec pathways. The third region of signal peptides that can affect the secretion output is the hydrophobic helical H region of the signal peptides. Also, the central h-region of signal peptides are important because the length and hydrophobic density of h-region intensify the hydrophobicity levels and facilitate the protein secretion.[19,20]

In the present study, the physicochemical features of the 90 signal peptides were analyzed for secretory expression of Somatropin in *E. coli*.

As shown in Table 3, flgI, OmpC, NPPC, mepA, and cysP showed the highest hydrophobicity levels (185.50, 171.90, 165.65, 164.74 and 164.00, respectively) among the studied signal peptides whereas, the signal peptides, bcsB, DsbC, zraP, ansB, and araF showed the lowest hydrophobicity (58.80, 78.50, 79.23, 93.64, and 93.91, respectively). Previous studies reported that OmpC has the highest aliphatic index, which is in agreement with our results.[21]

Analysis for secretory pathway revealed that all 55 Signal peptides (except efeO) are specific for the Sec pathway with reliability scores of more than 0.9 (Table 4). Therefore, our findings were consistent with some previous reports.[9,22] Sec exportome polypeptides have a cleavable, Sec-specific, n-terminal signal peptides that translocates proteins across the inner membrane (I.M.) in an unfolded state.[23,24]

There are two methods for selecting a signal peptide for any given protein, including experimental / trial and error method, and in silico analysis method. The advantages of using a bioinformatics program before starting an experimental study are increasing the precision and validity and reducing experimental research expenses.

In this study, online bioinformatic tools were used to find suitable signal peptides for periplasmic expression of recombinant Somatropin in *E. coli*. Different signal peptides, including 17 eukaryotic and 73 prokaryotic signal peptides, were evaluated. The D-score parameter was used to determine an appropriate signal peptides. D-score is also used to sort signal peptides in the first step. According to the D-scores (Table 2), 55 out of 90 selected signal peptides were identified as signal peptides for Somatropin. Data were sorted based on the priority of D-scores, final localization, h-region length, aliphatic index, Gravy, and solubility, respectively (Table 5). According to this sorting, pelB, flgl, nmpC, GfcA, OmpW, PpiA, and OmpC showed the highest D-score. However, pelB and OmpC showed the highest D-score in other bioinformatics studies.[21] The results of analysis revealed that Somatropin in connection with 34 signal peptides was stable and directed toward the Sec pathway, 9 signal

peptides mediated the secretion, and 15 signal peptide translocated the Somatropin into the periplasmic space.
Zamani et al analyzed the secretion of Somatropin by L-asparaginase II signal sequence and reported that successful secretion of Somatropin is not achieved using the L-asparaginase II signal sequence.[22]
The expression of Somatropin with the NPR, STII and DsbA signal peptides using RRI as the host cell, showed that the DsbA was the most effective signal peptide for Somatropin gene with 80% higher expression level compared to the reference vector.[23]
Previous studies[25] demonstrated the high secretion of Somatropin with phoA signal peptide, but in our research, phoA was not the right candidate due to lower D-sore (0.688) and final localization in the inner membrane.
This study evaluated 90 different signal peptide to find the most applicable signal peptide for secreting the recombinant Somatropin protein in the *E. coli*. The results of the present study showed that ynfB, sfaS, lolA, glnH, and malE has all the features needed to be selected as suitable signal peptides for Somatropin protein

Conclusion

In this research, various signal peptides were appraised for the periplasmic expression of Somatropin in *E. coli*. The selection was based on the combination of hydrophobicity, D score, solubility, stability, and the final localization.
The results indicated that specific signal peptides, including ynfB, sfaS, lolA, glnH, and malE have the highest scores and could be used for soluble periplasmic expression of Somatropin in *E. coli*. However, the proof of these results should be verified by an experimental study.

Ethical issues

This research was approved by Iran National Committee for ethics in Biomedical Research (958751).

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgments

This study was supported by National Institute for Medical Research Development (NIMAD) grant no. 958751.

References

1. Ghasemi R. Hashemzadeh H, Razavi H, Yakhchali B. History A. Production of Recombinant Human Growth Hormone and Future Challenges. *Modares J Biotechnol* 2018;9(1):79–92.
2. Okada S, Kopchick JJ. Biological effects of growth hormone and its antagonist. Trends Mol Med 2001;7(3):126–32. doi: 10.1016/s1471-4914(01)01933-5
3. Terpe K. Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol*. 2006;72(2):211–22. doi: 10.1007/s00253-006-0465-8
4. Humer D, Spadiut O. Wanted: More monitoring and control during inclusion body processing. *World J Microbiol Biotechnol*. 2018;34(11):158. doi: 10.1007/s11274-018-2541-5
5. Rinas U, Garcia-Fruitós E, Corchero JL, Vázquez E, Seras-Franzoso J, Villaverde A.

Bacterial Inclusion Bodies: Discovering Their Better Half. *Trends Biochem Sci* 2017;42(9):726–37. doi: 10.1016/j.tibs.2017.01.005

6. Tsirigotaki A, De Geyter J, Šoštarić N, Economou A, Karamanou S. Protein export through the bacterial Sec pathway. *Nat Rev Microbiol* 2017;15(1):21–36. doi: 10.1038/nrmicro.2016.161
7. Owji H, Nezafat N, Negahdaripour M, Hajiebrahimi A, Ghasemi Y. A comprehensive review of signal peptides: Structure, roles, and applications. *Eur J Cell Biol* 2018;97(6):422–41. doi: 10.1016/j.ejcb.2018.06.003
8. Güler-Gane G, Kidd S, Sridharan S, Vaughan TJ, Wilkinson TCI, Tigue NJ. Overcoming the Refractory Expression of Secreted Recombinant Proteins in Mammalian Cells through Modification of the Signal Peptide and Adjacent Amino Acids. van Raaij MJ, editor. *PLoS One* 2016;11(5):e0155340. doi: 10.1371/journal.pone.0155340
9. Taheri-Anganeh M, Khatami SH, Jamali Z, Savardashtaki A, Ghasemi Y, Mostafavi-Pour Z. In silico analysis of suitable signal peptides for secretion of a recombinant alcohol dehydrogenase with a key role in atorvastatin enzymatic synthesis. *Mol Biol Res Commun* 2019;8(1):17–26. doi: 10.22099/mbrc.2019.31801.1372.
10. Mergulhão FJM, Summers DK, Monteiro GA. Recombinant protein secretion in Escherichia coli. *Biotechnol Adv* 2005;23(3), 177–202. doi: 10.1016/j.biotechadv.2004.11.003.
11. Nielsen H. Predicting secretory proteins with signaIP. In: Methods in Molecular Biology. *Humana Press Inc* 2017;1611:59-73. doi: 10.1007/978-1-4939-7015-5–6
12. Klee EW, Ellis LBM. Evaluating eukaryotic secreted protein prediction. *BMC Bioinformatics* 2005;6(1):256. doi: 10.1186/1471-2105-6-256
13. Dalbey RE, Wang P, Dijl JM Van. Membrane proteases in the bacterial protein secretion and quality. *Microbiol Mol Biol Rev* 2012;76(2):311-30. doi: 10.1128/MMBR.05019-11
14. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 2019;37(4):420-3. doi: 10.1038/s41587-019-0036-z
15. Nielsen H, Tsirigos KD, Brunak S, Heijne G Von. A Brief History of Protein Sorting Prediction. *Protein J* 2019;38(3):200–16. doi: 10.1007/s10930-019-09838-3
16. Schmidt FR. Recombinant expression systems in the pharmaceutical industry. *Appl Microbiol Biotechnol* 2004;65(4):363–72. doi: 10.1007/s00253-004-1656-9
17. Freudl R. Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb Cell Fact* 2018;17(1):52. doi: 10.1186/s12934-018-0901-3
18. Inouye S, Soberon X, Franceschini T, Nakamura K, Itakura K. Role of positive charge on the amino-terminal region of the signal peptide in protein secretion across the membrane. *Proc Natl Acad Sci USA* 1982;79(11):3438–41. doi: 10.1073/pnas. 79.11.3438
19. Low KO, Mahadi NM, Illias RM. Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Appl Microbiol Biotechnol* 2013;97(9):3811–26. doi: 10.1007/s00253-013-4831-z
20. Chen H, Kim J, Kendall DA. Competition between functional signal peptides demonstrates variation in affinity for the secretion pathway. *J Bacteriol* 1996;178(23):6658–64. doi:10.1128/jb.178.23.6658-6664.1996
21. Zamani M, Nezafat N, Negahdaripour M, Dabbagh F, Ghasemi Y. In Silico Evaluation of Different Signal Peptides for the Secretory Production of Human Growth Hormone in E. coli. *Int J Pept Res Ther* 2015;21(3):261–8. doi: 10.1007/s10989-015-9454-z
22. Zamani M, Nezafat N, Ghasemi Y. Evaluation of recombinant human growth hormone

secretion in E. coli using the L-asparaginase II signal peptide. *Avicenna J Med Biotechnol* 2016;8(4):182–7.

23. Soares CRJ, Gomide FIC, Ueda EKM, Bartolini P. Periplasmic expression of human growth hormone via plasmid vectors containing the λPL promoter: Use of HPLC for product quantification. *Protein Eng* 2003;16(12):1131–8. doi: 10.1093/protein/gzg114
24. Tsirigotaki A, De Geyter J, Šoštarić N, Economou A, Karamanou S. Protein export through the bacterial Sec pathway. *Nat Rev Microbiol* 2017;15(1):21–36. doi: 10.1038 /nrmicro.2016.161
25. Gray GL, Baldridge JS, McKeown KS, Heyneker HL, Chung Nan Chang. Periplasmic production of correctly processed human growth hormone in Escherichia coli: natural and bacterial signal sequences are interchangeable. *Gene* 1985;39(2–3):247–54. doi: 10.1016/0378-1119(85)90319-1

Table1. The list of signal peptides was evaluated in this study.

NO	Full Name	Signal Peptide	length	Source	Accession Number	Amino Acid Sequence
1	Periplasmic appA protein	appA	22	Escherichia coli (strain K12)	P07102	MKAILIPFLSLLIPLTPQSAFA
2	Cytochrome c-type biogenesis protein	ccmH	18	Escherichia coli (strain K12)	P0ABM9	MRFLLGVLMLMISGSALA
3	Protein cexE	cexE	19	Escherichia coli	A2TJI4	MKKYILGVILAMGSLSAIA
4	Thiosulfate-binding protein	cysP	25	Escherichia coli (strain K12)	P16700	MAVNLLKKNSLALVASLLLAGHVQA
5	Drhemagglutinin structural subunit	draA	21	Escherichia coli	P24093	MKKLAIMAAASMVFAVSSAHA
6	Thiol:disulfide interchange protein dsbD	dsbD	19	Escherichia coli (strain K12)	P36655	MAQRIFTLILLLCSTSVFA
7	Thiol:disulfide interchange protein dsbG	dsbG	17	Escherichia coli (strain K12)	P77202	MLKKILLLALLPAIAFA
8	K88 fimbrail protein AD	faeG	21	Escherichia coli	P14191	MKKTLIALAIAASAASGMAHA
9	Iron(III) dicitrate-binding periplasmic protein	fecB	21	Escherichia coli (strain K12)	P15028	MLAFIRFLFAGLLLVISHAFA
10	F107 fimbrail protein	fedA	21	Escherichia coli	P25394	MKRLVFISFVALSMTAGSAMA
11	F41 fimbrail protein	FimF41a	22	Escherichia coli	P11900	MKKTLIALAVAASAAVSGSVMA
12	Flagellar P-ring protein	flgI	20	Escherichia coli O1:K1 / APEC	A1A9X5	MVIKFLSALILLLVTTAAQA
13	Protein transport protein hofQ	hofQ	18	Escherichia coli (strain K12)	P34749	MKQWIAALLLMLIPGVQA

14	Outer-membrane lipoprotein carrier protein	lolA	21	Escherichia coli (strain K12)	P61316	MKKIAITCALLSSLVASSVWA
15	Lipopolysaccharide export system protein lptA	lptA	27	Escherichia coli (strain K12)	P0ADV1	MKFKTNKLSLNLVLASSLLAASIPAFA
16	Maltose-binding periplasmic protein	malE	26	Escherichia coli (strain K12)	P0AEX9	MKIKTGARILALSALTTMMFSASALA
17	Penicillin-insensitive murein endopeptidase	mepA	19	Escherichia coli O157:H7	Q8XCQ5	MNKTAIALLALLASSVSLA
18	Nickel-binding periplasmic protein	nikA	22	Escherichia coli (strain K12)	P33590	MLSTLRRTLFALLACASFIVHA
19	Cytochrome c-552	nrfA	26	Escherichia coli (strain K12	P0ABK9	MTRIKINARRIFSLLIPFFFFTSVHA
20	Outer membrane protein A	ompA	21	Escherichia coli (strain K12	P0A910	MKKTAIAIAVALAGFATVAQA
21	Outer membrane protease ompP	ompP	23	Escherichia coli (strain K12)	P34210	MQTKLLAIMLAAPVVFSSQEASA
22	Outer membrane protein W	ompW	21	Escherichia coli (strain K12)	P0A915	MKKLTVAALAVTTLLSGSAFA
23	Fimbrial adapter papK	papK	21	Escherichia coli	P62532	MIKSTGALLLFAALSAGQAIA
24	D-alanyl-D-alanine endopeptidase	pbpG	25	Escherichia coli (strain K12)	P0AFI5	MPKFRVSLFSLALMLAVPFAPQAVA
25	pectate lyase B	PelB	22	Erwinia chrysanthemi	P04959	MKYLLPTAAAGLLLLAAQPAMA
26	Alkaline phosphatase	phoA	21	Escherichia coli (strain K12)	P00634	MKQSTIALALLPLLFTPVTKA
27	Outer membrane pore protein E	phoE	21	Escherichia coli (strain K12)	P02932	MKKSTLALVVMGIVASASVQA
28	Protein prsK	prsK	21	Escherichia coli	P42191	MIKSTGALLLFAALSAGQAMA
29	Phage shock protein E	pspE	19	Escherichia coli (strain K12)	P23857	MFKKGLLALALVFSLPVFA
30	Protease 3	ptrA	23	Escherichia coli (strain K12)	P05458	MPRSTWFKALLLLVALWAPLSQA
31	S-fimbrial adhesin protein	sfaS	22	Escherichia coli O6:K15:H31	P13430	MKLKAIILATGLINCIAFSAQA
32	Taurine-binding periplasmic protein	tauA	22	Escherichia coli (strain K12)	Q47537	MAISSRNTLLAALAFIAFQAQA

33	Thiamine-binding periplasmic protein	thiB	18	Escherichia coli (strain K12)	P31550	MLKKCLPLLLLCTAPVFA
34	Periplasmic protein torT	torT	18	Escherichia coli (strain K12)	P38683	MRVLLFLLLSLFMLPAFS
35	Trimethylamine-N-oxide reductase 1	TorA	39	Escherichia coli (strain K12)	P33225	MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATA
36	sn-glycerol-3-phosphate-binding periplasmic protein ugpB	ugpB	23	Escherichia coli (strain K12)	P0AG80	MKPLHYTASALALGLALMGNAQA
37	D-xylose-binding periplasmic protein	xylF	23	Escherichia coli (strain K12)	P37387	MKIKNILLTLCTSLLLTNVAAHA
38	Uncharacterized protein yfeK	yfeK	19	Escherichia coli (strain K12)	Q47702	MKKIICLVITLLMTLPVYA
39	UPF0379 protein yhcN	yhcN	22	Escherichia coli (strain K12)	P64614	MKIKTTVAALSVLSVLSFGAFA
40	Uncharacterized protein yncJ	yncJ	22	Escherichia coli (strain K12)	P64459	MFTKALSVVLLTCALFSGQLMA
41	UPF0482 protein ynfB	ynfB	28	Escherichia coli (strain K12)	P76170	MKITLSKRIGLLAILLPCALALSTTVHA
42	Zinc resistance-associated protein	zraP	26	Escherichia coli (strain K12)	P0AAA9	MKRNTKIALVMMALSAMAMGSTSAFA
43	Beta-lactamase	ampC	19	Escherichia coli (strain K12)	P00811	MFKTTLCALLITASCSTFA
44	Heat-labile enterotoxin B chain	eltB	21	Escherichia coli	P13811	MNKVKFYVLFTALLSSLCAHG
45	Type-1 fimbrial protein, C chain	pilC	23	Escherichia coli	P62605	MKLKFISMAVFSALTLGVATNAS
46	Copper resistance protein B	pcoB	23	Escherichia coli	Q47453	MKRNLKAIPVLVAGLFTSQLSIA
47	Serine protease eatA	eatA	56	Escherichia coli	Q84GK0	MNKVFSLKYSFLAKGFIAVSELARRVSVKGKLKSASSIIISPITIAIVSYAPPSLA
48	Hemoglobin-binding protease hbp	HBP	52	Escherichia coli	O88093	MNRIYSLRYSAVARGFIAVSEFARKVHKSVRRLCFPVLLLIPVLFSAGSLA
49	Thiol:disulfide interchange protein dsbA	DsbA	19	Escherichia coli (strain K12)	POAEG4	MKKIWLALAGLVLAFSASA
50	Human G.H.	Hgh	26	Homo sapiens	P01241	MATGSRTSLL LAFGLLCLPWLQEGSA
51	Outer membrane protein C	OmpC	21	Escherichia coli (strain K12)	P06996	MKVKVLSLLVPALLVAGAANA

52	Heat-stable enterotoxin II	STII	23	Escherichia coli	P22542	MKKNIAFLLASMFVFSIATNAYA
53	L-asparaginase 2	ansB	22	Escherichia coli (strain K12)	P00805	MEFFKKTALAALVMGFSGAALA
54	Chaperone protein sfmC	sfmC	23	Escherichia coli (strain K12)	P77249	MMTKIKLLMLIIFYLIISASAHA
55	Outer membrane protein F	ompf	22	Escherichia coli (strain K12)	P02931	MMKRNILAVIVPALLVAGTANA
56	Protease 7	ompt	20	Escherichia coli (strain K12)	P09169	MRAKLLGIVLTTPIAISSFA
57	Major outer membrane lipoprotein	LPP	20	Escherichia coli (strain K12)	P69776	MKATKLVLGAVILGSTLLAG
58	Maltoporin	lamB	25	Escherichia coli (strain K12)	P02943	MMITLRKLPLAVAVAAGVMSAQAMA
59	Beta-lactamase TEM	bla	23	Escherichia coli	P62593	MSIQHFRVALIPFFAAFCLPVFA
60	D-galactose-binding periplasmic protein	mglB	23	Escherichia coli (strain K12)	P0AEE5	MNKKVLTLSAVMASMLFGAAAHA
61	Heat-stable enterotoxin ST-IA/ST-P	Sta1	19	Escherichia coli	P01559	MKKLMLAIFISVLSFPSFS
62	L-arabinose-binding periplasmic protein	araF	23	Escherichia coli (strain K12)	P02924	MHKFTKALAAIGLAAVMSQSAMA
63	Putative outer membrane porin protein	nmpc	23	Escherichia coli (strain K12)	P21420	MKKLTVAISAVAASVLMAMSAQA
64	Peptidyl-prolyl cis-trans isomerase A	ppiA	24	Escherichia coli (strain K12)	P0AFL3	MFKSTLAAMAAVFALSALSPAAMA
65	UPF0412 protein YaaI	yaaI	23	Escherichia coli (strain K12)	P28696	MKSVFTISASLAISLMLCCTAQA
66	Uncharacterized protein YhcF	yhcF	20	Escherichia coli (strain K12)	P45422	MNNVKLLIAGSAFFAMSAQA
67	Uncharacterized fimbrial-like protein YfcQ	yfcQ	18	Escherichia coli (strain K12)	P76500	MRKTFLTLLCVSSAIAHA
68	Iron uptake system component EfeO	EfeO	26	Escherichia coli (strain K12)	P0AB24	MTINFRRNALQLSVAALFSSAFMANA
69	Glutamine-binding periplasmic protein	glnH	22	Escherichia coli (strain K12)	P0AEQ3	MKSVLKVSLAALTLAFAVSSHA
70	Ribonuclease I	rna	23	Escherichia coli	P21338	MKAFWRNAALLAVSLLPFSSANA

				(strain K12)		
71	disulfide interchange protein DsbC	DsbC	20	Escherichia coli (strain K12)	P0AEG6	MKKGFMLFTLLAAFSGFAQA
72	D-ribose-binding periplasmic protein	rbsB	25	Escherichia coli (strain K12)	P02925	MNMKKLATLVSAVALSATVSANAMA
73	Cyclic di-GMP-binding protein	bcsB	25	Escherichia coli (strain K12)	P37652	MKRKLFWICAVAMGMSAFPSFMTQA
74	Threonine-rich inner membrane protein GfcA	gfcA	21	Escherichia coli (strain K12)	P75885	MKHKLSAILMAFMLTTPAAFA
75	Salivary acidic proline -rich phosphoprotein	PRH1	22	Homo sapiens	P81277	MKVLRAWLLCLLMLGLALRGAA
76	Liver -expressed antimicrobial peptide2	LEAP2	22	Homo sapiens	Q969E1	MWHLKLCAVLMIFLLLLGQIDG
77	Secreted protein C10orf99	C10orf99	24	Homo sapiens	Q6UWK7	MRLLVLSSLLCILLLCFSIFSTEG
78	Prolactin -releasing peptide	PRLH	22	Homo sapiens	P81277	MKVLRAWLLCLLMLGLALRGAA
79	Heparin sulfate proteoglycan core protein	HSPG2	21	Homo sapiens	P98160	MGWRAAGALLLALLLHGRLLA
80	Transforming growth factor beta -2	TGFB2	19	Homo sapiens	P61812	MHYCVLSAFLILHLVTVAL
81	Serine protease inhibitor Kazal –type4	SPINK4	26	Homo sapiens	O60575	MAVRQWVIALALAALLVVDREVPVAA
82	C -type natriuretic peptide	NPPC	23	Homo sapiens	P23582	MHLSQLLACALLLTLLSLRPSEA
83	Tuberoinfundibular peptide of 39 residues	PTH2	30	Homo sapiens	Q96A98	METRQVSRSPRVRLLLLLLLLLVVPWGVRT
84	Pro-neuropeptide Y	NPY	28	Homo sapiens	P01303	MLGNKRLGLSGLTLALSLLVCLGALAEA
85	Interleukin -8	CXCL8	20	Homo sapiens	P10145	MTSKLAVALLAAFLISAALC
86	Alpha -1-antitrypsin	SERPINA1	24	Homo sapiens	P01009	MPSSVSWGILLLAGLCCLVPVSLA
87	Gastrin -releasing peptide	GRP	23	Homo sapiens	P07492	MRGSELPLVLLALVLCLAPRGRA
88	Plasminogen	PLG	19	Homo sapiens	P00747	MEHKEVVLLLLLFLKSGQG
89	Transforming growth factor beta -3	TGFB3	20	Homo sapiens	P10600	MKMHLQRALVVLALLNFATV
90	Guanylate cyclase activator 2B	GUCA2B	26	Homo sapiens	Q16661	MGCRAASGLLPGVAVVLLLLLQSTQS

Table2. Signal peptide probability and c, h and n regions.

NO	SP	n-region	h-region	c-region	Cleavage site	cleavage probability	C-score	Y-score	S-score	S-mean	D-score
1	appA	4	12	7	AFA	0.9807	0.801	0.786	0.938	0.808	0.797
2	ccmH	3	9	7	ALA	0.9806	0.773	0.568	0.655	0.472	0.532s
3	cexE	4	8	7	AIA	0.995	0.691	0.551	0.665	0.504	0.534
4	cysP	9	9	6	VQA	0.999	0.757	0.770	0.896	0.821	0.794
5	draA	4	10	7	AHA	0.990	0.717	0.807	0.971	0.921	0.860
6	dsbD	4	9	7	VFA	0.9705	0.829	0.605	0.649	0.503	0.567
7	dsbG	4	9	7	AFA	0.9021	0.417	0.447	0.712	0.536	0.480
8	faeG	5	10	7	AHA	0.9921	0.762	0.814	0.970	0.891	0.851
9	fecB	6	9	7	AFA	0.9354	0.601	0.424	0.514	0.355	0.398
10	fedA	5	9	8	AMA	0.9844	0.739	0.815	0.972	0.911	0.860
11	FimF41a	5	11	7	VMA	0.9827	0.873	0.869	0.978	0.896	0.882
12	flgI	5	9	7	AQA	0.9692	0.824	0.880	0.981	0.937	0.907
13	hofQ	4	8	7	VQA	0.9938	0.643	0.474	0.436	0.357	0.430
14	lolA	5	10	7	VWA	0.9948	0.715	0.675	0.874	0.724	0.693
15	lptA	11	9	7	AFA	0.9840	0.801	0.711	0.905	0.753	0.726
16	malE	8	10	9	ALA	0.9270	0.718	0.810	0.988	0.924	0.863
17	mepA	4	9	7	SLA	0.9500	0.790	0.726	0.860	0.717	0.722
18	nikA	7	9	7	VHA	0.9155	0.740	0.604	0.710	0.563	0.589
19	nrfA	10	10	7	VHA	0.9611	0.549	0.408	0.514	0.369	0.394
20	ompA	4	10	7	AQA	0.9814	0.800	0.841	0.968	0.891	0.865
21	ompP	6	10	6	ASA	0.8765	0.618	0.649	0.870	0.740	0.692
22	ompW	5	10	7	AFA	0.9924	0.808	0.863	0.966	0.923	0.891
23	papK	5	10	7	AIA	0.9415	0.721	0.642	0.837	0.659	0.648
24	pbpG	6	12	7	AVA	0.9542	0.681	0.753	0.985	0.890	0.817
25	PelB	6	10	6	AMA	0.9905	0.792	0.875	0.981	0.949	0.910
26	phoA	5	9	7	TKA	0.9648	0.496	0.613	0.845	0.722	0.688
27	phoE	5	10	6	VQA	0.9875	0.761	0.807	0.948	0.855	0.829
28	prsK	5	10	7	AMA	0.9805	0.837	0.854	0.950	0.881	0.867
29	pspE	4	9	7	VFA	0.9743	0.811	0.593	0.687	0.514	0.564
30	ptrA	8	9	7	SQA	0.9750	0.699	0.579	0.582	0.504	0.522
31	sfaS	5	11	7	AQA	0.9551	0.695	0.763	0.961	0.841	0.800
32	tauA	7	9	7	AQA	0.9441	0.832	0.820	0.947	0.834	0.827
33	thiB	4	8	6	VFA	0.9667	0.611	0.757	0.962	0.927	0.837
34	TorT	3	9	6	AFS	0.8362	0.435	0.413	0.593	0.442	0.424
35	TorA	18	15	7	ATA	0.9628	0.259	0.211	0.286	0.202	0.208
36	ugpB	7	10	7	AQA	0.9861	0.826	0.821	0.924	0.830	0.825
37	xylF	6	11	7	AHA	0.9446	0.726	0.806	0.973	0.903	0.851
38	yfeK	4	10	6	VYA	0.9878	0.711	0.490	0.571	0.398	0.456
39	yhcN	6	10	7	AFA	0.9780	0.714	0.596	0.793	0.602	0.598
40	yncJ	5	11	7	LMA	0.8738	0.798	0.851	0.962	0.904	0.876
41	ynfB	10	12	7	VHA	0.9723	0.819	0.623	0.789	0.590	0.611
42	zraP	7	12	8	AFA	0.9535	0.786	0.838	0.994	0.929	0.881
43	ampC	4	10	6	TAS-CS.	0.6246	0.788	0.848	0.942	0.910	0.877
44	eltB	6	9	7	AHG	0.6339	0.647	0.747	0.954	0.874	0.807
45	pilC	5	11	7	TNA-SF.	0.8309	0.171	0.392	0.973	0.909	0.635
46	pcoB	7	10	7	SIA	0.9063	0.369	0.378	0.585	0.449	0.404
47	eatA	37	13	7	-	-	0.230	0.166	0.329	0.286	0.210
48	HBP	34	12	7	SLA	0.6063	0.243	0.179	0.262	0.168	0.175
49	DsbA	4	10	6	ASA-	0.9419	0.572	0.616	0.837	0.717	0.654
50	Hgh	7	12	6	GSA	0.8990	0.200	0.237	0.539	0.318	0.267
51	OmpC	5	10	7	ANA	0.9648	0.827	0.863	0.973	0.918	0.889
52	STII	5	12	7	AYA	0.9604	0.856	0.856	0.971	0.892	0.873
53	ansB	7	9	7	ALA	0.9587	0.838	0.644	0.707	0.555	0.611
54	sfmC	7	10	7	AHA	0.9601	0.806	0.595	0.576	0.439	0.537
55	ompf	6	10	7	ANA	0.981	0.839	0.862	0.946	0.902	0.880
56	ompt	5	9	7	SFA	0.9250	0.293	0.335	0.538	0.414	0.364
57	LPP	6	9	5	LLA-GF	0.4598	0.145	0.214	0.581	0.472	0.309
58	lamB	9	10	7	AMA	0.8549	0.785	0.819	0.981	0.894	0.854
59	bla	7	10	7	VFA	0.9203	0.624	0.413	0.465	0.334	0.384
60	mglB	5	12	7	AHA	0.9717	0.767	0.834	0.986	0.923	0.876
61	Sta1	4	9	7	SFS	0.8744	0.492	0.664	0.939	0.888	0.769
62	araF	6	11	7	AMA	0.987	0.804	0.844	0.958	0.874	0.858
63	nmpc	5	12	7	AQA	0.9833	0.835	0.876	0.981	0.930	0.902
64	ppiA	5	13	7	AMA	0.9564	0.785	0.846	0.989	0.939	0.890
65	yaaI	6	11	7	AQA	0.7641	0.721	0.806	0.957	0.913	0.856

66	yhcF	6	8	7	AQA	0.9636	0.737	0.748	0.897	0.777	0.761
67	yfcQ	4	8	7	AHA	0.9790	0.712	0.783	0.932	0.854	0.816
68	EfeO	9	11	7	ANA	0.9450	0.585	0.705	0.973	0.875	0.785
69	glnH	6	10	7	SHA	0.9779	0.740	0.814	0.965	0.910	0.859
70	rna	7	10	7	ANA	0.9760	0.784	0.835	0.975	0.912	0.871
71	DsbC	4	10	7	AQA	0.9809	0.764	0.825	0.971	0.898	0.859
72	rbsB	6	12	8	AMA	0.6795	0.798	0.818	0.979	0.893	0.854
73	bcsB	6	11	9	TQA	0.8993	0.455	0.615	0.985	0.889	0.744
74	gfcA	5	10	7	AFA	0.9834	0.8441	0.882	0.985	0.925	0.902
75	PRH1	6	10	7	RGA	0.542	0.195	0.324	0.657	0.553	0.409
76	LEAP2	6	10	7	LLG	0.324	0.139	0.165	0.359	0.281	0.208
77	C10orf99	4	12	7	IFS	0.593	0.255	0.298	0.499	0.384	0.329
78	PRLH	6	10	7	RGA	0.542	0.195	0.324	0.657	0.553	0.409
79	HSPG2	5	10	7	LLA	0.986	0.330	0.269	0.356	0.246	0.260
80	TGFB2	3	10	7	PLS	0.049	0.129	0.186	0.443	0.355	0.248
81	SPINK4	6	12	8	-	0.051	0.143	0.190	0.453	0.322	0.239
82	NPPC	6	11	7	SEA	0.9791	0.398	0.566	0.889	0.804	0.678
83	PTH2	15	9	7	VRT	0.5430	0.156	0.172	0.354	0.310	0.233
84	NPY	7	13	7	AEA	0.6235	0.578	0.465	0.504	0.413	0.446
85	CXCL8	5	16	7	ALC	0.5500	0.343	0.420	0.816	0.603	0.488
86	SERPINA1	7	11	7	SLA	0.8489	0.402	0.289	0.395	0.260	0.278
87	GRP	6	11	7	GRA	0.8903	0.268	0.242	0.381	0.244	0.243
88	PLG	-	-	-	GQG	0.4277	0.207	0.239	0.444	0.246	0.242
89	TGFB3	7	10	10	-	0.147	0.146	0.224	0.610	0.519	0.333
90	GUCA2B	5	15	7	TQS	0.7021	0.320	0.249	0.369	0.270	0.257

Table 3. The physicochemical characteristics of the signal peptides that were analyzed in the study.

NO	Signal peptides	Length	M.W. (Da)	P.I.	Net positive charge	GRAVY	Aliphatic index	Instability (Separately)	Instability with hGH*	Stability*	Solubility
1	appA	22	2384.99	8.5	0.9	1.405	155.45	53.16	42.9	u	0.782
2	cysP	25	2575.15	10	2.1	1.064	164.00	11.14	37.38	S	0.765
3	draA	21	2135.63	10	2.1	1.162	98.10	16.49	38.41	S	0.885
4	faeG	21	2027.47	10	2.1	1.005	112.38	11.36	37.90	S	0.883
5	fedA	21	2231.76	11	1.9	1.290	102.38	29.55	39.70	S	0.869
6	FimF41a	22	2090.57	10	1.9	1.355	124.55	15.15	38.17	S	0.863
7	flgI	20	2116.67	8.5	0.9	1.935	185.50	10.64	37.96	S	0.806
8	lolA	21	2192.70	9.3	0.9	1.324	139.52	16.67	38.43	S	0.764
9	lptA	27	2849.47	10.3	2.9	0.881	130.37	17.32	37.91	S	0.831
10	malE	26	2698.34	11.1	2.9	1.012	113.08	2.85	36.27	S	0.879
11	mepA	19	1887.31	8.5	0.9	1.479	164.74	32.07	40.03	u	0.833
12	nikA	22	2434.99	10.3	0.9	1.350	137.73	60.45	42.85	u	0.790
13	ompA	21	2046.50	10	1.9	1.295	121.43	9.52	37.72	S	0.857
14	ompP	23	2406.88	5.7	1.9	0.904	114.78	44.47	41.21	u	0.798
15	ompW	21	2093.55	10	1.9	1.210	125.71	1.44	36.92	S	0.824
16	papK	21	2047.48	8.5	1.9	1.390	140.00	-2.60	36.52	S	0.849
17	pbpG	25	2705.36	11	1.9	1.228	117.20	57.99	42.81	u	0.800
18	PelB	22	2228.78	8.3	0.9	1.191	138.18	41.42	40.88	u	0.802
19	phoA	21	2256.82	10	0.9	0.971	139.52	56.02	42.33	u	0.769
20	phoE	21	2104.59	10	0.9	1.195	130.00	1.44	36.92	S	0.834

21	prsK	21	2065.52	8.5	0.9	1.267	121.43	3.27	37.10	S	0.859
22	sfaS	22	2290.85	9.3	0.9	1.314	146.82	5.41	37.16	S	0.844
23	tauA	22	2308.72	9.5	0.9	1.055	120.45	34.41	40.16	u	0.824
24	thiB	18	1974.60	8.8	0.9	1.589	157.22	65.64	42.96	u	0.608
25	ugpB	23	2342.80	8.3	0.9	0.622	110.87	18.01	38.37	S	0.844
26	xylF	23	2482.08	9.3	0.9	1.083	161.30	33.61	40.04	u	0.781
27	yhcN	22	2254.76	10	0.9	1.418	128.64	-2.03	36.39	S	0.764
28	yncJ	22	2344.91	7.9	0.9	1.541	128.64	15.15	38.17	S	0.795
29	ynfB	28	2948.71	10	0.9	1.239	163.93	29.32	39.35	S	0.774
30	zraP	26	2733.37	11.1	0.9	0.746	79.23	28.75	39.37	S	0.834
31	ampC	19	2022.46	7.8	0.9	1.342	97.89	25.22	39.41	u	0.783
32	eltB	21	2342.84	9.1	0.9	0.890	111.43	31.10	39.86	S	0.803
33	pilC	23	2400.92	10	0.9	1.104	110.43	1.01	36.54	S	0.794
34	DsbA	19	1990.48	10	0.9	1.416	144.21	11.50	38.17	S	0.842
35	OmpC	21	2078.63	10	0.9	1.552	171.90	14.37	38.20	S	0.797
36	STII	23	2552.09	9.7	1.9	1.026	102.17	32.43	39.92	S	0.861
37	ansB	22	2274.76	8.3	1.9	1.136	93.64	-1.15	36.48	S	0.846
38	ompF	22	2266.83	11	1.9	1.259	150.91	67.18	43.54	u	0.876
39	sta1	19	2159.72	10	1.9	1.368	123.16	25.28	39.41	S	0.841
40	lamB	25	2545.22	11	1.9	1.332	125.20	42.97	41.07	u	0.889
41	mglB	23	2362.89	10	1.9	0.952	102.17	14.15	37.95	S	0.865
42	araF	23	2348.87	10	1.9	0.878	93.91	96.71	46.83	u	0.876
43	nmpc	23	2292.84	10	1.9	1.243	119.13	30.34	39.69	S	0.883
44	ppiA	24	2371.90	8.5	1.9	1.438	98.33	39.94	40.72	u	0.841
45	yaaI	23	2389.93	7.8	1.9	1.365	114.78	23.74	38.98	S	0.842
46	yhcF	20	2084.48	8.5	1.9	0.915	98.00	25.79	39.39	S	0.860
47	yfcQ	18	1962.40	9.5	1.9	1.006	119.44	13.91	38.50	S	0.792
48 38	efeO	26	2845.33	12	1.9	0.654	94.23	54.20	42.42	u	0.865
49	glnH	22	2244.72	10	1.9	1.209	133.18	10.58	37.70	S	0.846
50	rna	23	2478.94	11	1.9	0.757	106.52	40.05	40.74	u	0.809
51	DsbC	20	2179.67	10	1.9	1.000	78.50	5.25	37.45	S	0.836
52	rbsB	25	2494.02	10	1.9	0.948	109.60	11.14	37.38	S	0.879
53	bcsB	25	2853.53	10	1.9	0.688	58.80	48.06	41.66	u	0.874
54	gfcA	21	2293.87	10	1.9	1.019	98.10	40.98	40.83	u	0.842
55	NPPC	23	2494.05	6.5	1.9	1.07	165.65	95.44	46.69	u	0.737

*S=Stable, U=Unstable
*The proteins whose instability index was higher than 40 were predicted as unstable and the values under 40 might be stable.

Table 4. Analysis of secretion pathways and final localization of human somatropin mediated by different signal peptides.

NO	sp	Secretion pathway	Reliability Score	Cytoplasmic	Membrane	Secreted	Periplasmic	Final Prediction Site
1	appA	Sec/SPI	0.9925	1.68	4.70	0.00	3.62	Inner membrane
2	cysP	Sec/SPI	0.9795	1.42	6.26	0.00	2.33	Outer Membrane
3	draA	Sec/SPI	0.9984	0.86	4.74	0.48	3.92	Outer Membrane
4	faeG	Sec/SPI	0.9984	0.53	1.75	5.03	2.69	Secreted
5	fedA	Sec/SPI	0.9963	0.32	7.13	2.55	0.00	Inner Membrane
6	FimF41a	Sec/SPI	0.9963	0.00	2.40	6.31	1.29	Secreted
7	flgI	Sec/SPI	0.9892	1.09	5.84	0.00	3.07	Inner Membrane
8	lolA	Sec/SPI	0.9975	0.43	2.34	0.00	7.23	periplasmic
9	lptA	Sec/SPI	0.9846	0.55	6.03	0.00	3.42	Outer Membrane
10	malE	Sec/SPI	0.9909	0.71	3.44	0.00	5.85	Periplasmic
11	mepA	Sec/SPI	0.9925	0.58	7.14	0.00	2.29	Outer Membrane
12	nikA	Sec/SPI	0.9001	0.8	5.47	0.00	3.73	Inner membrane
13	ompA	Sec/SPI	0.9977	0.13	1.07	5.21	3.58	Sedreted
14	ompP	Sec/SPI	0.9834	1.76	7.82	0.00	0.42	Outer membrane
15	ompW	Sec/SPI	0.9965	0.00	6.16	2.12	1.72	Outer Membrane
16	papK	Sec/SPI	0.978	0.11	1.83	7.41	0.65	secreted
17	pbpG	Sec/SPI	0.9844	0.64	2.43	0.00	6.93	Periplasmic
18	PelB	Sec/SPI	0.9967	1.29	1.42	3.33	3.96	Periplasmic
19	phoA	Sec/SPI	0.9924	1.15	7.68	0.00	1.17	Inner membrane
20	phoE	Sec/SPI	0.9973	0.28	8.63	0.43	0.66	Inner Membrane
21	prsK	Sec/SPI	0.9929	0.00	2.13	6.21	1.66	Secreted
22	sfaS	Sec/SPI	0.9831	1.52	3.49	0.00	4.99	Periplasmic
23	tauA	Sec/SPI	0.9096	0.74	5.50	0.00	3.75	Outer Membrane
24	thiB	Sec/SPI	0.9867	0.80	2.85	0.00	6.35	Periplasmic
25	ugpB	Sec/SPI	0.995	0.55	3.17	0.00	6.29	Periplasmic
26	xylF	Sec/SPI	0.9969	1.40	3.81	0.00	4.80	periplasmic
27	yhcN	Sec/SPI	0.9896	0.26	8.20	1.54	0.00	Inner membrane
28	yncJ	Sec/SPI	0.9078	1.21	7.34	0.00	1.45	Inner membrane
29	ynfB	Sec/SPI	0.9881	0.00	2.65	0.98	6.37	periplasmic
30	zraP	Sec/SPI	0.9931	0.57	2.46	0.00	6.97	Peri plasmic
31	ampC	Sec/SPII	0.6243	0.93	2.63	0.00	6.39	Periplasmic
32	eltB	Sec/SPI	0.7337	0.97	7.60	0.00	1.43	Outer membrane
33	pilC	Sec/SPI	0.9545	0.99	8.63	0.29	0.10	Outer membrane
34	DsbA	Sec/SPI	0.9875	0.00	8.44	0.68	0.89	Inner membrane
35	OmpC	Sec/SPI	0.9874	0.33	6.55	1.58	1.54	Inner membrane
36	STII	Sec/SPI	0.9953	0.11	8.42	1.47	0.00	Outer membrane
37	ansB	Sec/SPI	0.9641	0.60	6.46	0.00	2.94	Inner membrane
38	ompF	Sec/SPI	0.9896	0.62	8.19	0.74	0.45	Inner membrane
39	sta1	Sec/SPI	0.9672	0.08	9.51	0.41	0.00	Inner membrane
40	lamB	Sec/SPI	0.9865	0.32	3.71	3.88	2.09	Secreted
41	mglB	Sec/SPI	0.9971	0.80	5.63	0.00	3.57	Inner membrane
42	araF	Sec/SPI	0.9941	0.22	3.73	0.00	6.05	Periplasmic
43	nmpc	Sec/SPI	0.9964	0.00	0.96	7.84	1.20	Secreted
44	ppiA	Sec/SPI	0.9934	0.54	5.45	0.00	4.01	Outer membrane
45	yaaI	Sec/SPI	0.78	0.18	4.43	2.80	2.59	Inner membrane
46	yhcF	Sec/SPI	0.9801	0.86	8.13	0.00	1.01	Outer membrane
47	yfcQ	Sec/SPI	0.9956	1.58	7.04	0.37	1.01	Inner membrane
48	efeO	TAT	0.5377	0.25	0.49	0.00	9.26	Periplasmic
49	glnH	Sec/SPI	0.9959	0.18	3.97	0.00	5.85	Periplasmic
50	rna	Sec/SPI	0.9914	0.75	8.88	0.37	0.00	Outer membrane
51	Dsbc	Sec/SPI	0.9955	0.46	5.80	0.00	3.75	Inner membrane
52	rbsB	Sec/SPI	0.9969	0.00	2.61	2.76	4.63	periplasmic
53	bcsB	Sec/SPI	0.9793	0.02	2.28	7.17	0.53	Secreted
54	gfcA	Sec/SPI	0.9959	0.19	2.21	6.76	0.85	Secreted
55	NPPC	Sec/SPI	0.9877	1.33	7.54	0.00	1.12	Inner membrane

Table 5. Characteristics of most efficient signal peptides for periplasmic expression of human somatropin based on their determinant features

no	sp	Aliphatic index	Gravy	D-score	stability	solubility
1	ynfB	163.93	1.239	0.611	39.35	0.774
2	xylF	161.30	1.083	0.851	40.04	0.781
3	thiB	157.22	1.589	0.837	42.96	0.608
4	sfaS	146.82	1.314	0.800	37.16	0.844
5	lolA	139.52	1.324	0.693	38.43	0.764
6	PelB	138.18	1.191	0.910	40.88	0.802
7	glnH	133.18	1.209	0.859	37.70	0.846
8	pbpG	117.20	1.228	0.817	42.81	0.800
9	malE	113.08	1.012	0.863	36.27	0.879
10	ugpB	110.87	0.622	0.825	38.37	0.844
11	rbsB	109.60	0.948	0.854	37.38	0.879
12	ampC	97.89	1.342	0.877	39.41	0.783
13	efeO	94.23	0.654	0.785	42.42	0.865
14	araF	93.91	0.878	0.858	46.83	0.876
15	zraP	79.23	0.746	0.881	39.37	0.834