International Journal of Peptide Research and Therapeutics (2020) 26:1879–1891
https://doi.org/10.1007/s10989-019-09986-1

In Silico Study of Different Signal Peptides to Express Recombinant Glutamate Decarboxylase in the Outer Membrane of *Escherichia coli*

Hanieh Yarabbi[1] · Seyed Ali Mortazavi[1] · Masoud Yavarmanesh[1] · Ali Javadmanesh[2]

Accepted: 21 November 2019 / Published online: 3 December 2019

Abstract

Gamma amino butyric acid (GABA) is used as drugs, food ingredients, and dietary supplements. L-glutamate is converted to GABA by the decarboxylation reaction, which is catalyzed by the glutamate decarboxylase (GAD). *Escherichia coli* is widely being used to express proteins. However, without appropriate signal peptide, it cannot be applied for secretory proteins. Selecting a suitable signal peptide (SP) is a critical step in the secretory production of different proteins. In silico identification of suitable SP is a reliable and cost-effective alternative to experimental approaches. In previous studies, the localization of proteins was not considered and the SPs of periplasmic, membranes and extracellular were compared. Therefore, this study aimed to predict the best SP for the expression of recombinant GAD in the outer membrane of *E. coli* only. Also, we compared twelve servers to evaluate protein localization, solubility, and secretory pathway. In the present study, 127 SPs were taken from the Signal Peptide database. The localization site, physico-chemical properties, location of cleavage sites, regions and D-score of them were determined by ProtComp, ProtParam, and SignalP 3.0 and 4.1 servers, respectively. To rank SPs based on the secretion properties, PRED-TAT and SignalP 5.0 webservers were used. Based on the results, the localization site of 13 SPs was in the outer membrane of *E. coli.* Among them, the most suitable candidates seemed to be torT with a reasonably high D-score, aliphatic index, and GRAVY, followed by ccmH and then pspE. TorT accelerates GAD scale-up production and might be useful in future experimental research.

Keywords Signal peptide · Outer membrane · Glutamate decarboxylase · Expression

Introduction

γ-Aminobutyric acid (GABA) is an active biogenic substance present in the central nervous system (Cohen et al. 2002). It is involved in the regulation of the sleep–wake cycle, reducing blood pleasure (Inoue et al. 2003), prevention of diabetic condition, inducing insulin secretion from the pancreas (Adeghate and Ponery 2002; Hagiwara et al. 2004). Abnormalities in glutamate decarboxylase (GAD) function and reduced GABA levels are reported in people with many neurological disorders (Möhler 2012). GAD is a pyridoxal 5′-phosphate dependent enzyme that catalyzes l-glutamate decarboxylation to γ-aminobutyric acid (Komatsuzaki et al. 2005). Many bacterial GADs exhibit optimal activity at a pH range of 4.0–5.0, whereas at neutral pH, their activity decreases sharply. But Among the microorganisms GADs, GAD from *Enterococcus faecium* DO is active even in the neutral pH and has high performance (Hagiwara et al. 2004). The optimum temperature and pH for GAD activity were 30 C and 6–7.5, respectively (Lim et al. 2016). Km and Vmax values of GAD from Enterococcus strains were 3.26–5.26 mM and 1.20–3.45 µM/min, respectively (Chang et al. 2017; Lee et al. 2017). GAD from *E. faecium* DO has 466 amino acids with a molecular mass of 53.7 kD (NCBI_017960.1, UniProtKB- Q3Y080).

Escherichia coli is the most commonly used expression system in recombinant protein production (Rosano and Ceccarelli 2014), due to (i) fast growth (Sezonov et al. 2007); (ii) high cell density is easily attained (Shiloach and Fass

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10989-019-09986-1) contains supplementary material, which is available to authorized users.

✉ Seyed Ali Mortazavi
morteza1937@yahoo.com

[1] Department of Food Science and Technology, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

[2] Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

2005); (iii) inexpensive complex media can be used for growth (Sivashanmugam et al. 2009); (iv) well-characterized genetics, physiology and metabolism (Andersen et al. 2013); (v) simple fermentation, and favorable economics (Daegelen et al. 2009). *E. coli* strain BL21 (DE3) can direct high-level expression of cloned genes under the control of the T7 promoter (Kim et al. 2017).

The recombinant GAD enzyme has already been produced in the cytoplasm of *E. coli* (Fan et al. 2012; Yu et al. 2012), although our purpose is to express this enzyme in the outer membrane of the cell. One of the important challenges that cells face is the protein transfer from their site of synthesis in the cytosol to their sites of function. *E. coli* without suitable signal peptide, cannot be used for secretory proteins. Choosing a suitable signal peptide is a critical step in the secretory expression of different proteins (Choi and Lee 2004). Therefore, evaluation of different SP for expression recombinant glutamate decarboxylase in the outer membrane of *E. coli* is extremely crucial for increase GABA production. The secretion of recombinant GAD to the outer membrane of *E. coli* has several advantages over intracellular production. These benefits include minimizing protein degradation, simplifying downstream purification, reduces production costs, enhanced biological activity, higher product stability and solubility, and further N-terminal authenticity of the expressed peptide (Mergulhao et al. 2004). High-level expression of the recombinant GAD in cytoplasmic, periplasmic and outer membrane leads to aggregation of misfolded protein (Chang et al. 2017; Ueno 2000). Nevertheless, Santos et al. (2012) and Chang et al. (2017) mentioned that with a simple refolding process, it was converted to a folded protein with an acceptable efficiency.

In general, there are three main pathways in bacteria for translocation of proteins across the cytoplasmic into the periplasm, outer membrane or extracellular that have been classified to the general secretion pathway (Sec-pathway); the twin-arginine translocation (TAT-pathway) and the signal recognition particle pathway (SPR pathway) (De Marco 2009). It seems Sec and SRP pathways are more essential than the TAT pathway because folding and purification of secretory proteins in outer membrane space are more natural than in the cytoplasm (Choi and Lee 2004). Since the degradation of secretory proteins is less than cytoplasm, it can be concluded that the SPs using these pathways can be more appropriate than SPs which use TAT pathways (Natale et al. 2008).

The Sec machinery recognizes an N-terminal hydrophobic signal sequence. A cysteine residue follows immediately after the signal peptide cleavage site; this signal peptide is recognized and cleaved by lipoprotein signal peptidase (SPaseII or Lsp) after the N-terminal cysteine is modified with a lipid moiety, which anchors the protein to the membrane. Finally, an additional fatty acid is attached to the new N-terminus (Juncker et al. 2003). These proteins are then either retained at the cytoplasmic membrane or translocated into the outer membrane by the Lol lipoprotein-sorting pathway (Lewenza et al. 2008). Signal peptides for the sec pathway are typically 20 amino acids in length and generally consist of the following three domains: (i) a positively charged n-region that often contains Lys or Arg residues, (ii) a hydrophobic h-region and (iii) an uncharged but polar C-region (Papanikou et al. 2007). The cleavage site for the signal peptidase is located in the c-region (Green and Mecsas 2016).

Several articles have been published about "In silico analysis of different signal peptides for the secretory production of recombinant protein" (Mohammadi et al. 2019; Vahedi et al. 2019; Zamani et al. 2015). However, various signal peptides for the secretory production of recombinant protein, including the inner membrane (IM), periplasm, outer membrane (OM), and extracellular have been compared in one topic, and no distinction was made between them. Therefore, in the present study, in addition to seeking to find the best signal peptide, we carefully examine the protein localization and compare only the signal peptides expressing the Gad enzyme in *E. coli*'s outer membrane. This study was aimed only to predict best signal peptides to express recombinant glutamate decarboxylase in the outer membrane of *E. coli.* Also, there is no study to evaluate different signal peptides in connection with GAD and their probable effect on appropriate protein secretion. Furthermore, in this research several bioinformatics tools compared to the prediction of the subcellular localization, solubility and the secretion properties of proteins such as PSORTb, CELLO, Gneg-PLoc, ProtComp, SOLpro, PROSO II, CcSOL omics, Wilkinson and Harrison model, protein-sol, SODA, PRED-TAT, and SignalP 5.0 webservers.

Materials and Methods

Signal Sequence Collection and Study Design

In this study, an amino acid sequence encoding Glutamate decarboxylase of *E. faecium* DO was obtained from the UniProtKB server at http://www.uniprot.org/. GAD of *E. faecium* DO (UniProtKB- Q3Y080) has 466 amino acids with a molecular mass of 53.7 kD. Also, the amino acid sequences of 127 signal peptides were retrieved from the Signal Peptide Database (http://www.signalpeptide.de/). Signal sequences are listed in supplementary Table 1.

The Amino Acid Sequence of the GAD *Enterococcus faecium DO*

Translation = "MLYGKDNQEEKNYLEPIFGSASED-VDLPKYKLNKESIEPRIAYQLVQDEMLDEG-

NARLNLATFCQTYMEPEAVKLMTQTLEKNAIDK-SEYPRTTEIENRCVNMIADLWHAPNNEKFMGT-STIGSSEACMLGGMAMKFAWRKRAEKLGLDIQ-AKKPNLVISSGYQVCWEKFCVYWDVELREVPM-DEKHMSINLDTVMDYVDEYTIGIVGIMGITYTGRY DDIKGLNDLVEAHNKQTDYKVYIHVDAASGGFYAP-FTEPDLVWDFQLKNVISINSSGHKYGLVYPGVG-WVLWRDQQYLPEELVFKVSYLGGEMPTMAINFSH-SAAQLIGQYYNFVRYGFDGYRDIHQRTHDVAVY-LAKEIEKTGIFEIINDGSELPVVCYKLKEDPNREWTLY-DLSDRLLMKGWQVPAYPLPKDLDQLIIQRLVVRADF-GMNMAGDYVQDMNQAIEELNKAHIVYHKKQDVK-TYGFTH".

Computational Tools and Determine the Characteristics of Signal Peptides

Identification of Sub-Cellular Localization Site of Glutamate Decarboxylase

Gram-negative bacteria have five major subcellular localization sites: the cytoplasm, the periplasm, the inner membrane, the outer membrane, and the extracellular space. The OM is the outermost structure in Gram-negative bacteria and hence is the interface between the cell and the environment (Mogensen and Otzen 2005). Since subcellular location plays a crucial role in protein function, the availability of systems that can predict location from the sequence will be essential to the full characterization of expressed proteins. Experimental determination of subcellular location is mainly accomplished by three approaches: electron microscopy, fluorescence microscopy, and cell fractionation. These methods are very variable and time-consuming (Paladin et al. 2017). To predict signal peptides by in silico methods, different bioinformatics tools have been developed that are based on neural networks, weight matrices, or sequence alignments (Gardy et al. 2004).

Computational prediction of the Final position of proteins is a major tool for automated protein annotation and genome analysis. Due to a protein's subcellular localization can provide clues regarding its function in an organism and is critical to a wide range of studies (Yu et al. 2014). Several algorithms have been developed to the prediction of the subcellular localization of proteins such as PSORTb, CELLO, Gneg-PLoc, and ProtComp servers. The predictive websites are listed as follows (Table 1):

The performance of CELLO, PSORT-B, Gneg-mPLoc, and ProtCompB servers compared in Table 2. According to the results, ProtCompB achieved better prediction accuracy and sensitivity for all outer membrane signal peptides of *E. coli* than the other approaches. The overall prediction precision of ProtCompB reached 94.12%, which was 6.62% and 28.56% higher than CELLO (87.5%) and PSORT-B (65.56%). Noticeably, ProtCompB prediction MCC for outer membrane location (p = 96%) is higher than other predictors. In general, ProtCompB gave significantly better predictive performances for outer membrane signal peptides of *E. coli.* For this reason we used the ProtCompB server to predict the final subcellular localization of the GAD enzyme connected with different signal peptides. Precision is a measure of the ability of the system to predict only the relevant data. Accuracy of the system is defined by the closeness of its prediction toward the true values. The MCC calculates the correlation between the prediction and the observation

Table 2 The comparison of performances in outer membrane signal peptides of *Escherichia coli*

Name of server	Precision	Accuracy (%)	MCC
CELLO	87.5	84.6	0.70
PSORT-B	65.56	80.3	0.66
Gneg-mPLoc	48.57	57.6	0.59
ProtCompB	94.12	96.3	0.96

Table 1 The predictive website addresses and their features

Name of server	Website	Feature	References
CELLO	www.cello.life.nctu.edu.tw	By the composition of peptides of varying lengths (n-peptide composition)	Yu et al. (2014)
PSORTb	www.psort.org/psortb/index.html	Based on amino acid composition information, protein sequence features and sorting signal knowledge	Gardy et al. (2004)
Gneg-mPLoc	www.csbio.sjtu.edu.cn/bioinf/Gneg-multi	By incorporating the information of gene ontology, functional domain, and sequential evolution	Shen and Chou (2010)
ProtCompB	www.softberry.com	Based on N-terminal targeting sequences And discriminating the individual targeting signal peptide	http://www.softberry.com (2016)

(Gardy et al. 2004; Shen and Chou 2010; Yu et al. 2014; http://www.softberry.com 2016).

ProtComp B server was used for in silico study and prediction of the final destination of Glutamate decarboxylase linked to different signal peptides (http://www.softberry.com). ProtCompB Version 9 combines several methods of protein localization prediction—neural networks-based prediction; direct comparison with bases of homologous proteins of known localization; comparisons of pentamer distributions calculated for query and DB sequences; prediction of specific functional peptide sequences, such as signal peptides and transmembrane segments. It means that the program treats correctly only complete sequences, containing signal sequences, anchors, and other functional peptides if any. The most important point is that, in this server, if both NNets and other predictions point to the same compartment, this is a very reliable prediction. The aggregate produced by ProtCompB has been reported as one of the most precise ensemble methods in subcellular localization predictions in general (http://www.softberry.com 2016).

Prediction of n, h, and c Regions, Cleavage Site and Signal Peptide Probability

The "n, h and c" regions were predicted by the SignalP 3.0 server at http://www.cbs.dtu.dk/services/SignalP3.0/because SignalP 4.1 and SignalP 5.0 servers are not able to evaluate n, h, and c Regions. The output of SignalP 4.1 was reported as five scores. The discrimination score (D-score) and S-score recognized cleavage sites and signal peptide positions, respectively. The Y-score was the geometric average of the C-score and the slope of the S-score, which results in the more precise prediction of the cleavage sites than the raw C-score. The average of the S-score was S-mean. D-score was the average of the S-mean and Y-max, which indicated the primary distinction between secretory and non-secretory proteins (Nielsen 2017). SignalP server as the most accurate and reliable tool for identification of cleavage sites works based on a combination of several neural networks, namely artificial neural network (ANN) and hidden Markov model (HMM) and average accuracy is 87% (Petersen et al. 2011). The presence of cleavage sites, their locations in signal peptide and signal peptide probability were assigned by SignalP 4.1 and SignalP 5.0 servers.

Investigation of Physicochemical Parameters of Signal Peptides

Physicochemical properties of signal peptides, including the length of SP sequence, molecular weight, theoretical PI, aliphatic index, instability index, grand average of hydropathicity (GRAVY), extinction coefficients, positively and negatively charged residues and estimated half-life were determined by ProtParam using the ExPASy server at http://web.expasy.org/protparam/. ProtParam computes various physicochemical properties that can be deduced from a protein sequence. No additional information is required about the protein under consideration. ProtParam, as a part of ExPASy and maintained by SIB and the European Bioinformatics Institute (EBI), is considered very trustable for computation of physicochemical properties of proteins (Gasteiger et al. 2005).

Protein Solubility Prediction

Prediction of protein solubility upon expression in *E. coli* was made by SOLpro, PROSO II, CcSOL omics, Wilkinson and Harrison model, protein-sol and SODA webservers.

SOLpro predicts protein solubility in *E. coli* using a two-stage SVM architecture based on multiple representations of the primary sequence (Cheng et al. 2005). Each classifier of the first layer takes as input a distinct set of features describing the sequence. A final SVM classifier summarizes the resulting predictions and predicts if the protein is soluble or not as well as the corresponding probability (Magnan et al. 2009). This webserver can be accessed from URL: http://scratch.proteomics.ics.uci.edu/.

PROSO II (Protein Solubility evaluator II) classifies proteins in soluble and insoluble categories at http://mbiljj45.bio.med.uni-muenchen.de:8888/prosoII/prosoII.seam. It is built on sequence composition and similarity-based model. This server can detect the subset of sequence features that possess the strongest impact on protein solubility (Smialowski et al. 2012). PROSO II employs a model based on a logistic function and an adapted Parzen window algorithm trained on experimental data extracted from the pepcDB (Berman et al. 2008) and PDB (Berman et al. 2000) databases.

CcSOL algorithm predicts protein solubility using physicochemical properties. The server also computes point mutations throughout the whole protein sequence to identify susceptible areas. CcSOL omics can be freely accessed on the web at http://service.tartaglialab.com/page/ccsol_group. In CcSOL, hydrophobicity, hydrophilicity, β-sheet, and α-helical propensities are combined into a solubility propensity score that is useful to investigate protein expression (Agostini et al. 2014).

SODA uses the propensity of the protein sequence to aggregate as well as intrinsic disorder, plus hydrophobicity and secondary structure preferences to estimate changes in the solubility. Also, SODA can evaluate difficult types of variation including point mutations, deletions, and insertions (Paladin et al. 2017). The webserver can be accessed from URL: http://protein.bio.unipd.it/soda.

The Wilkinson-Harrison model is based on two parameters: average charge, determined by the relative numbers

of Asp, Glu, Lys and Arg residues, and the content of turn-forming residues (Asn, Gly, Pro, and Ser). Protein solubility was calculated according to Wilkinson-Harrison using their webserver (http://www.biotech.ou.edu/) (Idicula-thomas et al. 2005; Smialowski et al. 2006b).

Protein-Sol is a webserver for predicting protein solubility in a graphical format. This webserver is available at http://protein-sol.manchester.ac.uk. The tool can highlight lysine and arginine content regarding modifying protein solubility (Hebditch et al. 2017).

The performance of different methods for predicting protein solubility is presented in Table 3. The protein-sol was the single best performing method in this comparison with accuracy, Matthew's correlation coefficient (MCC) and area under the receiver operating characteristic curve (AUROC) equal to 82.8%, 0.382 and 0.922, respectively (Agostini et al. 2014; Magnan et al. 2009; Paladin et al. 2017; Smialowski et al. 2012). It was followed by the ccSOL omics method. The Lowest performance was related to the Wilkinson and Harrison model. Protein-sol was proposed recently and shown to outperform previous methods in a comparative study led by the authors (Hebditch et al. 2017).

The receiver operating characteristic curve (ROC) portrays the relationship between the true positive rate and the false positive rate of the classifier (Smialowski et al. 2006b). AUROC measures the discriminating ability of the model and it takes values between 0.5 for a random drawing and 1.0 for the perfect classifier (Smialowski et al. 2012). It is often interpreted as a probability that if you randomly draw one positive and one negative instance, the one scored higher by the model will be actual positive (Frank et al. 2004).

Evaluation of the Secretion Properties of Signal Peptides

To sort SPs based on the secretion properties, PRED-TAT and SignalP 5.0 webservers were used. PRED-TAT operates based on Hidden Markov Models (HMMs) (Bagos et al. 2010). It can be accessed from http://www.compgen.org/tools/PRED-TAT/submit. PRED-TAT had MCC, CS recall and CS precision of 0.82–0.97, 0.72–0.78, 0.17–0.76 for predicting Sec pathway and Tat pathway SPs for Gram-negative bacteria, respectively (Bagos et al. 2010).

SignalP 5.0 is a deep neural network-based method combined with conditional random field classification and optimized transfer learning for improved SP prediction. SignalP 5.0 can differentiate between "standard" signal peptides translocated by the Sec translocon (Sec/SPI) and "Tat" (Twin-Arginine Translocation) signal peptides translocated by the Tat translocon (Tat/SPI) in Bacteria. In general, SignalP 5.0 distinguishes three types of signal peptides in prokaryotes: Sec substrates cleaved by SPase I (Sec/SPI), Sec substrates cleaved by SPase II (Sec/SPII), and Tat substrates cleaved by SPase I (Tat/SPI). SignalP 5.0 is available at http://www.cbs.dtu.dk/services/SignalP/index.php (Armenteros et al. 2019). To apply all webservers, each signal peptide was linked to the N-terminal of GAD amino acid sequence so that methionine residues were put in between SP and GAD amino acid sequence. SignalP 5.0 had MCCs of 0.907, 0.960 and 0.981 for predicting Sec/SPI SPs, Sec/SPII,Tat/SPI SPs for Gram-negative bacteria, respectively. Also, Regarding CS precision, the performance of SignalP 5.0 varies between 0.630 and 0.970, whereas its CS recall varies between 0.579 and 0.970. SignalP 5.0 performs as well as PRED-TAT for predicting Tat/SPI SPs in Gram-negative bacteria. SignalP 5.0 displayed the highest CS precision and CS recall scores in Gram-negative bacteria. Finally, SignalP 5.0 has the best SP discrimination in the Sec and Tat pathways (Armenteros et al. 2019).

Results and Discussion

Predicting Subcellular Localization of GAD Connected to Different Signal Peptides

ProtCompB webserver was used for predicting the subcellular location of GAD connected to different signal peptides. The predicted localization site of our protein with all signal peptides is shown in supplement's Table 3. According to the Sub-cellular localization analysis results, it can be seen that among 127 SPs, the final localization site for 13 signal peptides (RZOR, FAED, Bla, ccmH, cexE, dsbG, pspE, torT, eglS, yehD, ASPG_ERWCH, yiiX, and bcsB) were in the outer Membrane space (Table 4).

Prediction of n, h and c-Regions and Signal Peptide Probability

The results showed that SPs' D-scores were between 0.642 (RZOR) and 0.893 (pspE) (Table 5). The most important

Table 3 Evaluation of performances of SOLpro and PROSO II servers in prediction of protein solubility

Method	SOLpro	PROSO II	CcSOL omics	Wilkinson and Harrion model	Protein-sol	SODA
Accuracy	74%	71%	74%	56.2%	82.8%	59.2%
MCC	0.487	0.422	0.519	0.127	0.382	0.458
ROC	0.742	0.785	0.855	0.601	0.922	0.734

Table 4 Identifying the sub-cellular location of GAD connected to different signal peptides by ProtComp server

Name signal peptide	LocDB/PotLocDB				Neural nets				Pentamers				Integral				Final
	CP	OM	EX	PP	CP	OM	EX	PP	CP	OM	EX	PP	CP	OM	EX	PP	
RZOR	0	0	0	0	0.53	2.17	0.04	0.26	0.28	2.43	0	0.5	0.86	7.58	0	1.55	OM
FAED	0	0	0	0	0.98	1.22	0.07	0.73	0.51	1.4	0	1.13	1.67	4.6	0	3.73	OM
Bla	0	0	0	0	0.46	2.16	0.1	0.28	0.31	2.46	0	0.42	0.96	7.73	0	1.31	OM
ccmH	0	0	0	0	0.58	1.97	0.17	0.27	0.37	2.27	0	0.51	1.19	7.2	0	1.62	OM
cexE	0	0	0	0	0.93	1.38	0.26	0.43	0.49	1.4	0	1.27	1.54	4.44	0	4.02	OM
dsbG	0	0	0	0	0.11	2.73	0.05	0.11	0.07	2.74	0	0.37	0.24	8.61	0	1.15	OM
pspE	0	0	0	0	0.48	2	0.29	0.23	0.28	1.6	0	1.29	0.89	5.04	0	4.07	OM
torT	0	0	0	0	0.11	2.72	0.05	0.11	0.07	2.33	0	0.79	0.23	7.31	0	2.46	OM
ASPG_ ERWCH	0	0	0	0	0.06	1.93	1.01	0	0.51	2	0	0.55	1.68	6.53	0	1.79	OM
eglS	0	0	0	0	0.5	2.07	0.19	0.25	0.27	2.32	0	0.58	0.86	7.32	0	1.83	OM
yehD	0	0	0	0	0.39	2.11	0.31	0.19	0.34	2.32	0	0.48	1.08	7.38	0	1.54	OM
yiiX	0	0	0	0	0.3	2.43	0.08	0.2	0.16	2.76	0	0.23	0.5	8.78	0	0.72	OM
bcsB	0	0	0	0	1.1	1.33	0.15	0.42	0.62	2.07	0	0.31	2.07	6.88	0	1.05	OM

CP cytoplasmic, *OM* outer membrane, *EX* extracellular, *PP* periplasmic

parameter for the diagnosis of a SP is the discriminating score (D-score) which is usually described with a cut-off value of 0.5. Only when an SP sequence has a D-score above 0.50, it is considered. In silico analysis results of the SignalP server have also indicated that the highest D-score belonged to pspE, ccmH, ASPG_ERWCH and yiiX, respectively (Table 5).

The sequences with a D-score higher than 0.57 were classified as putative signal peptides, whereby sequences possessing a D-score above 0.7 had a high probability that they did so. The used setting was *E. coli,* default D-cutoff value of 0.57 and standard graphics output. To use the server, for the evaluations on the whole secretory candidate protein, each SP sequence was connected to the N-terminal of glutamate decarboxylase amino acid sequence and methionine residues were inserted between each SP and GAD amino acid sequence.

For in silico investigation of n, h and c regions, SignalP version 3.0 was applied. The results showed that the collected SPs' n-region length was between 3 and 17, h-region length was between 7 and 12, and c-region length was between 2 and 10 amino acids. It seemed all SP sequences in our study not only had a D-score above 0.50, but also contained distinct n, h and c regions (Table 5).

The N and h-regions play a critical role in transferring recombinant proteins into outer membrane space, while c-region plays a vital role as a cleavable site which can be distinguished by signal peptidase enzyme. Therefore a reliable SP sequence should have clear n, h and c regions (Owji et al. 2018). The hydrophobicity factor extremely relies on the length of h-region. The increase in the length of h-region would improve the level of hydrophobicity (Papanikou et al. 2007). Accordingly, there has been a significant diversity in the length of SPs h-region (7–12). Considering h-regions in Table 5, which indicate the hydrophobicity levels of the signal peptides torT, RZOR, FAED, eglS, yehD, and bcsB have the highest hydrophobicity levels among all 13 signal peptides.

Cleavage Site Prediction

According to the results (Table 5), all 13 signal peptides implying that signal peptidase enzyme correctly identified their cleavage sites. The c-region is the site of signal peptide cleavage by the signal peptidase. An "A-x-A" box sequence is believed to govern the cleavage motif in *E. coli*, which is characterized by the presence of alanine amino acid at the positions − 3 and − 1 relative to the signal peptidase cleavage site (Von Heijne and Abrahmsèn 1989). According to consensus motif A-X-A, the "x" is a large bulky residue like Phe, Tyr, Leu, and His at position -2 (Pratap and Dikshit 1998). Six of our SPs have AxA motif in their cleavage sites, including ccmH, cexE, dsbG, ASPG_ERWCH, eglS, and yiiX (Table 5).

Investigation of Physicochemical Parameters

The different physico-chemical properties of signal peptides, including the length of SP sequence, molecular weight, theoretical PI, aliphatic index, instability index, GRAVY and positively and negatively charged residues were evaluated by the ProtParam server, as shown in Table 6 and supplementary Table 4. The in silico results showed that the SP lengths were between 17 (dsbG) to 35

Table 5 Signal peptide probability and n, h and c regions of signal sequences

No.	Signal peptide	Predicted cleavage location	Pos. cleavage site	Cleavage site probability (%)	C score	Y score	S score	S mean	D score	n-region	h-region	c-region
1	RZOR	VVG-MM	19–20	0.992	0.161	0.388	0.958	0.929	0.642	6 (1–6)	11 (7–17)	2 (18–19)
2	FAED	VIA-MM	35–36	1	0.379	0.573	0.953	0.851	0.703	17 (1–17)	11 (18–28)	7 (29–35)
3	Bla	VFA-MM	23–24	0.997	0.501	0.665	0.956	0.88	0.781	7 (1–7)	9 (8–16)	7 (17–23)
4	ccmH	ALA-MM	18–19	1	0.782	0.836	0.931	0.886	0.859	3 (1–3)	9 (4–12)	7 (12–18)
5	cexE	AIA-MM	19–20	0.999	0.583	0.718	0.925	0.869	0.789	5 (1–5)	10 (6–15)	4 (16–19)
6	dsbG	AFA-MM	17–18	0.998	0.355	0.571	0.93	0.915	0.733	5 (1–5)	10 (6–15)	2 (16–17)
7	pspE	VFA-MM	19–20	0.993	0.828	0.874	0.959	0.915	0.893	4 (1–4)	9 (5–13)	6 (14–19)
8	torT	AFS-MM	18–19	0.926	0.455	0.654	0.951	0.936	0.787	4 (1–4)	12 (5–16)	2 (17–18)
9	ASPG_ERWCH	ASA-MM	pos. 21–22	0.998	0.732	0.816	0.944	0.899	0.855	6 (1–6)	10 (7–16)	5 (17–21)
10	eglS	ASA-MM	pos. 29–30	0.993	0.506	0.661	0.984	0.905	0.776	8 (1–8)	11 (9–19)	10 (20–29)
11	yehD	VFA-MM	pos. 22–23	0.991	0.751	0.805	0.954	0.884	0.843	4 (1–4)	11 (5–15)	7 (16–22)
12	yiiX	AFA-MM	pos. 18–19	1	0.782	0.831	0.933	0.881	0.855	4 (1–4)	7 (5–12)	6 (13–18)
13	bcsB	TQA-MM	pos. 25–26	0.924	0.311	0.511	0.976	0.892	0.69	6 (1–6)	11 (7–17)	8 (18–25)

Table 6 Physico-chemical properties of the GAD connected to signal peptides determined by ProtParam

No.	Signal peptides	Amino acid length	MW (Da)	PI	Net positive charge	Net negative charge	Charge GRAVY	Aliphatic Index	Instability	Classify
1	RZOR	486	56021.34	5.15	51	69	– 0.31	85.41	34.61	Stable
2	FAED	502	57526.83	5.2	52	69	– 0.302	85.42	35.19	Stable
3	Bla	490	56454.57	5.09	49	69	– 0.296	84.35	36.49	Stable
4	ccmH	485	55751.82	5.05	49	69	– 0.305	85.81	34.39	Stable
5	cexE	486	55807.86	5.1	50	69	– 0.316	85.84	35.29	Stable
6	dsbG	484	55667.79	5.1	50	69	– 0.302	87.42	35.44	Stable
7	pspE	486	55893.98	5.1	50	69	– 0.305	85.64	34.8	Stable
8	torT	485	55940.07	5.05	49	69	– 0.296	86.41	35.18	Stable
9	ASPG_ ERWCH	488	56367.43	5.06	50	70	– 0.312	84.08	35.26	Stable
10	eglS	496	56869.14	5.1	50	69	– 0.287	85.89	37.46	Stable
11	yehD	489	56167.16	5.1	50	69	– 0.309	83.13	37.2	Stable
12	yiiX	485	55813.85	5.1	50	69	– 0.321	86.41	35.74	Stable
13	bcsB	492	56681.88	5.15	51	69	– 0.332	81.83	36.15	Stable

(FAED) amino acid for 13 sequences, with an average of 22 amino acids (Supplementary Table 4). Also, the lowest and the highest Mw belonged to dsbG ($Mw_{sp} = 1839.44$, $Mw_{\text{sp connected to GAD}} = 55667.79$) and FAED ($Mw_{sp} = 3698.48$, $Mw_{\text{sp connected to GAD}} = 57526.83$), respectively (Table 6 and supplementary Table 4).

All the selected SPs had net positive charges (Arg-Lys) of 1–4 and negative charges (Asp-Glu) of 0–1 based on ProtParam results, whereas the range of PI signal peptide and PI of the signal peptides connected to GAD were between 8.02 (Bla)—11 (yehD, yiiX) and 5.05 (ccmH, torT)—5.2 (FAED), respectively (Table 5 and supplementary Table 4). A net charge of at least one is assumed essential for the efficient export of the recombinant protein and different signal peptides may require different magnitudes of positive charge for maximum efficiency (Low et al. 2013). A net positive charge in the N region (arginines and/or lysines) enhances the processing and translocation rates protein to the outer membrane (Guo et al. 2018).

As it is observed, the lowest and the highest GRAVY belonged to bcsB and eglS, respectively (Table 6). The grand average of hydropathy score (GRAVY) for a protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence (Kyte and Doolittle 1982). A positive GRAVY is a positive indicator of hydrophobicity and a negative indicator of hydrophilicity. Therefore, in addition to presenting the hydrophobicity of the protein, it can show an association with its solubility. A more hydrophobicity implies a higher ability of the protein in hydrogen bonding formation with water molecules and higher solubility (Gasteiger et al. 2005; Low et al. 2013).

The aliphatic index is another factor, which indicates the hydrophobicity value. The highest aliphatic index belonged to dsbG and the lowest belonged to bcsB (Table 6). It seems, according to our results, all SPs have appropriate GRAVY and aliphatic index to use. The aliphatic index is defined as the relative volume occupied by the aliphatic side chains (i.e., alanine, valine, isoleucine, and leucine) in an amino acid sequence. Consequently, the SPs which have a high GRAVY and aliphatic index are much better to apply (Gasteiger et al. 2005).

Instability index of five signal peptides (Separately) including Bla, eglS, yehD, yiiX, and bcsB were more than 40, so they were predicted as unstable (supplementary Table 4). However, according to our results in Table 6, the instability index of signal peptides in connection with GAD was between 34.39 (ccmH) and 37.46 (eglS). Instability index all the signal peptides in connection with GAD were less than 40 and predicted as stable. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable (Gamage et al. 2019).

Protein Solubility Prediction by Several Computational Methods

We evaluated our signal peptides by directly applying the SOLpro, PROSO II, ccSOL omics, Wilkinson and Harrison model, protein-sol and SODA webservers. The solubility of glutamate decarboxylase in connection with the 13 studied signal peptides analysis results showed that GAD was insoluble and Insolubility probability in *E. coli* was between 0.566 (ccmH) and 0.593 (pspE) out of 1 (Table 7).

High-level expression of the recombinant GAD in cytoplasmic, periplasmic and outer membrane leads to aggregation of misfolded protein (Chang et al. 2017; Ueno 2000).

Table 7 Solubility of the signal peptides predicted by SOLpro, PROSO II, ccSOL omics, Wilkinson and Harrison model, protein-sol and SODA servers

No.	Method signal peptides	SOLpro	PROSO II	CcSOL omics	Wilkinson and Harrison model	Protein-sol	SODA	Final result, the mean of probabilities
1	RZOR	Insoluble; 0.658560	Insoluble; 0.528	Insoluble; 0.94	Insoluble	Insoluble; 0.519	Insoluble	Insoluble; 0.568
2	FAED	Insoluble; 0.713758	Insoluble; 0.533	Soluble; 1.0	Insoluble	Insoluble; 0.500	Insoluble	Insoluble; 0.582
3	Bla	Insoluble; 0.715333	Insoluble; 0.526	Insoluble; 0.54	Insoluble	Insoluble; 0.519	Insoluble	Insoluble; 0.586
4	ccmH	Insoluble; 0.651040	Insoluble; 0.516	Soluble; 0.79	Insoluble	Insoluble; 0.532	Insoluble	Insoluble; 0.566
5	cexE	Insoluble; 0.725601	Insoluble; 0.521	Insoluble; 0.93	Insoluble	Insoluble; 0.527	Insoluble	Insoluble; 0.590
6	dsbG	Insoluble; 0.719120	Insoluble; 0.518	Insoluble; 0.90	Insoluble	Insoluble; 0.506	Insoluble	Insoluble; 0.581
7	pspE	Insoluble; 0.724931	Insoluble; 0.530	Soluble; 1.0	Insoluble	Insoluble; 0.527	Insoluble	Insoluble; 0.593
8	torT	Insoluble; 0.726721	Insoluble; 0.509	Insoluble; 0.86	Insoluble	Insoluble; 0.516	Insoluble	Insoluble; 0.583
9	ASPG_ERWCH	Insoluble; 0.708111	Insoluble; 0.521	Insoluble; 0.99	Insoluble	Insoluble; 0.540	Insoluble	Insoluble; 0.589
10	eglS	Insoluble; 0.736770	Insoluble; 0.493	Insoluble; 0.54	Insoluble	Insoluble; 0.492	Insoluble	Insoluble; 0.589
11	yehD	Insoluble; 0.688305	Insoluble; 0.524	Insoluble; 0.99	Insoluble	Insoluble; 0.529	Insoluble	Insoluble; 0.580
12	yiiX	Insoluble; 0.718173	Insoluble; 0.526	Insoluble; 0.85	Insoluble	Insoluble; 0.533	Insoluble	Insoluble; 0.592
13	bcsB	Insoluble; 0.686927	Insoluble; 0.525	Soluble; 0.61	Insoluble	Insoluble; 0.517	Insoluble	Insoluble; 0.576

As in our experiments, the Gad enzyme was expressed as an inclusion body. As Santos et al. (2012) and Chang et al. (2017) mentioned, with a simple refolding process which has acceptable efficiency, is converted to a folded protein.

As Chang et al. (2013) mentioned, the solubility of passenger proteins seems essential for efficient outer membrane expression, considering that the insoluble proteins may misfold or form inclusion bodies in this cellular compartment.

These insoluble proteins need to be solubilized and refolded to obtain functional proteins (Paladin et al. 2017). The researchers observed that insoluble proteins more frequently contained hydrophobic stretches of 20 or more residues, had lower glutamine content (Gln composition < 4%), fewer negatively charged residues (Asp +Glu composition < 17%) and a higher percentage of aromatic amino acids (aromatic composition > 7.5%) than soluble proteins (Smialowski et al. 2006a).

Changing the growth conditions, such as growth temperature, pH of the culture medium, concentration of inducer and induction time can be effective in decreasing the formation of inclusion bodies and improve the solubility of glutamate decarboxylase (Fan et al. 2012). At the isoelectric point (pI),

Table 8 Secretion sorting of SPs by PRED-TAT and SignalP 5.0 servers

No.	Signal peptides	PRED-TAT		SignalP 5.0	
		Type of SP	Reliability score	Type of SP	Reliability score
1	Rzor	Sec	0.912	Sec/SPII	0.3657
2	FAED	Sec	0.955	Sec/SPI	0.4654
3	Bla	Sec	0.926	Sec/SPI	0.5722
4	CcmH	Sec	0.955	Sec/SPI	0.8999
5	CexE	Sec	0.983	Sec/SPI	0.7685
6	DsbG	Sec	0.995	Sec/SPI	0.6186
7	PspE	Sec	0.983	Sec/SPI	0.8827
8	torT	Sec	0.981	Sec/SPI	0.6923
9	Aspg-erwch	Sec	0.923	Sec/SPI	0.4946
10	EglS	Sec	0.997	Sec/SPI	0.8803
11	YehD	Sec	0.987	Sec/SPI	0.8358
12	YiiX	Sec	0.990	Sec/SPI	0.845
13	BcsB	Sec	0.978	Sec/SPI	0.9675

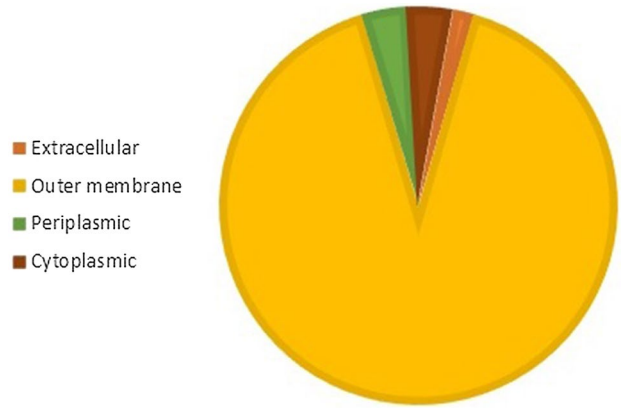

Fig. 1 Localization prediction for GAD connected to torT signal peptide

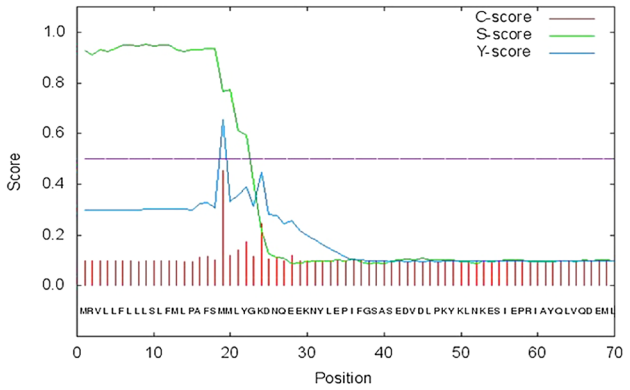

Fig. 2 Prediction the presence and location of signal peptide cleavage sites in GAD amino acid sequence linked with torT signal peptide

proteins have a net zero charge, attractive forces predominate, and molecules tend to associate, resulting in insolubility (Gromiha 2010). Also, most proteins could be expressed as a soluble protein in the presence of sorbitol, arginine, and trehalose or chemical additives in the expression medium (Godbey 2014). These materials can suppress the formation of inclusion bodies through decrease the non-covalent interactions between protein molecules. Thus, increase the solubility of target protein in *E. coli* overexpression systems (Gromiha 2010).

Secretion Sorting of Signal Peptides

The classification was confirmed by detection of signal peptides based on the secretion properties using the PRED-TAT and SignalP 5.0 servers. The results demonstrated that all 13 SPs belonged to the Sec pathway (Table 8).

Overall Considerations and Selection of the Best Potential SPs

Based on the results, sub-cellular localization sites of 13 signal peptides were in the outer membrane of *E. coli,* where the signal peptidase enzyme properly identified their cleavage sites. Also, according to the computational analysis, the most suitable candidates seemed to be torT with a reasonably high D-score, aliphatic index and GRAVY, followed by ccmH and then pspE (Figs. 1, 2 and 3).

There is a need for increased protein solubility to produce proteins on a large scale for industrial purposes. Overexpression of proteins in *E. coli* leads to the formation of insoluble protein or inclusion bodies, because bacteria lack the necessary system for protein folding in the natural form. Therefore, protein produced by in vitro conditions needs to be refolded.

There are different techniques for refolding of the inclusion body proteins including adding accelerant, chromatography, dialysis, dilution, and ultrafiltration, etc. (Godbey 2014). Commonly used chemical additives for protein refolding are denaturants [urea, guanidinium chloride (GdnHCl)], detergents (Triton X-100, CHAPS, SDS, N-lauroylsarcosine and CTAB Detergents with cycloamylose or cyclodextrin) and inhibitors (arginine hydrochloride, arginine amide, glycine amide, proline) (Gromiha 2010).

Conclusion

γ-Aminobutyric acid has broad potential for application as a bioactive additive in the food and pharmaceutical industries. GABA is biosynthesized from L-glutamate and this reaction is catalyzed by glutamate decarboxylase. The best approach for the transfer of GAD to outer membrane space is using a suitable signal peptide. The identification of suitable SPs is one of the most vital steps to produce secretory proteins as a recombinant protein in *E. coli*. The computational method provides the ability to rapidly predict possible secretory SPs and other features in the efficient secretion. A list of secretory SPS can provide an opportunity to select the best option based on efficient secretion.

The secretory SPs' D-scores were between 0.642 (RZOR) and 0.893 (pspE). Considering h-regions in Table 4, which indicate the hydrophobicity levels of the signal peptides torT, RZOR, FAED, eglS, yehD, and bcsB have the highest hydrophobicity levels among all 13 signal peptides. All 13 signal peptides implying that signal peptidase enzyme correctly identify their cleavage sites. The secretory SPs having the highest GRAVY were eglS, torT, Bla, dsbG, and FAED. Instability index all the signal peptides in connection with GAD were less than 40 and predicted as stable. Six of our SPs have AxA motif in their cleavage sites, including ccmH,

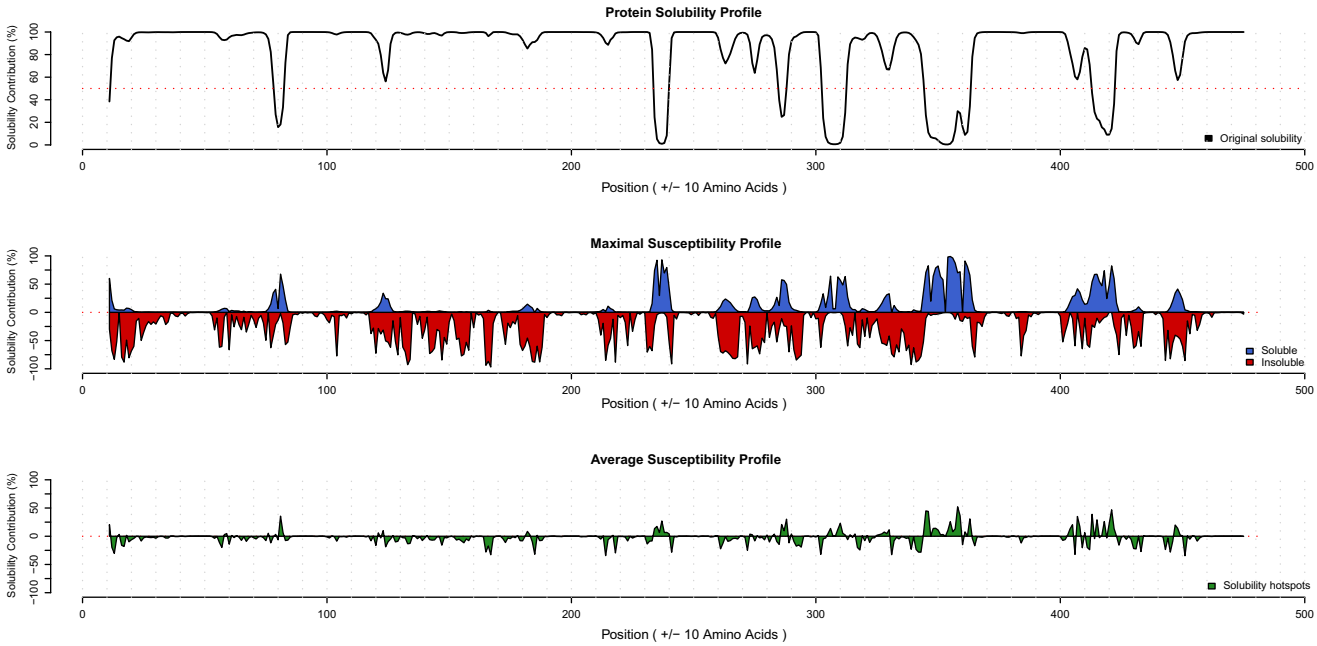

Fig. 3 In silico distribution of GAD solubility attached to torT signal peptide

cexE, dsbG, ASPG_ERWCH, eglS, and yiiX. Finally, the most suitable candidates seemed to be torT with a fairly high D-score, aliphatic index, and GRAVY, followed by ccmH and then pspE, which are Sec-pathway SPs. torT accelerates GAD scale-up production and might be useful in future experimental research.

Acknowledgements This work was supported by funds from the Ferdowsi University of Mashhad (Grant # 3/48396) and the Iran National Science Foundation (Grant # 98000180).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Adeghate E, Ponery AS (2002) GABA in the endocrine pancreas: cellular localization and function in normal and diabetic rats. Tissue Cell 34(1):1–6

Agostini F, Cirillo D, Livi CM, Delli Ponti R, Tartaglia GG (2014) cc SOL omics: a webserver for solubility prediction of endogenous and heterologous expression in *Escherichia coli*. Bioinformatics 30(20):2975–2977

Andersen KR, Leksa NC, Schwartz TU (2013) Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins 81(11):1857–1861

Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37(4):420

Bagos PG, Nikolaou EP, Liakopoulos TD, Tsirigos KD (2010) Combined prediction of Tat and Sec signal peptides with hidden Markov models. Bioinformatics 26(22):2811–2817

Berman H, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242

Berman HM, Westbrook JD, Gabanyi MJ, Tao W, Shah R, Kouranov A, Kopp J (2008) The protein structure initiative structural genomics knowledgebase. Nucleic Acids Res 37:D365–D368

Chang CCH, Song J, Tey BT, Ramanan RN (2013) Bioinformatics approaches for improved recombinant protein production in *Escherichia coli*: protein solubility prediction. Brief Bioinform 15(6):953–962

Chang C, Zhang J, Ma SH, Wang L, Wang D, Zhang J, Gao Q (2017) Purification and characterization of glutamate decarboxylase from *Enterococcus raffinosus* TCCC11660. J Ind Microbiol Biotechnol 44(6):817–824

Cheng J, Randall AZ, Sweredoski MJ, Baldi P (2005) SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res 33:W72–W76

Choi JH, Lee SY (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*. Appl Microbiol Biotechnol 64(5):625–635

Cohen I, Navarro V, Clemenceau S, Baulac M, Miles R (2002) On the origin of interictal activity in human temporal lobe epilepsy in vitro. Science 298(5597):1418–1421

Daegelen P, Studier FW, Lenski RE, Cure S, Kim JF (2009) Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21 (DE3). J Mol Biol 394(4):634–643

De Marco A (2009) Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. Microb Cell Fact 8(1):26

Fan E, Huang J, Hu S, Mei L, Yu K (2012) Cloning, sequencing and expression of a glutamate decarboxylase gene from the GABA-producing strain *Lactobacillus brevis* CGMCC 1306. Ann Microbiol 62(2):689–698

Frank E, Hall M, Trigg L, Holmes G, Witten IH (2004) Data mining in bioinformatics using Weka. Bioinformatics 20(15):2479–2481

Gamage DG, Gunaratne A, Periyannan GR, Russell TG (2019) Applicability of instability index for in vitro protein stability prediction. Protein Pept Lett 26(5):339–347

Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FS (2004) PSORTb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics 21(5):617–623

Gasteiger E, Hoogland C, Gattiker A, Wilkins, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) The proteomics protocols handbook. Humana Press, New Jersey, pp 571–607

Godbey WT (2014) Chapter 2 – Proteins. An Introduction to Biotechnology. *The Science, Technology and Medical Applications*. pp 9–33

Green ER, Mecsas J (2016) Bacterial secretion systems—an overview. Microbiology spectrum 4(1):1–32

Gromiha MM (2010) Chapter 1 – Proteins. *Protein bioinformatics: from sequence to function*. pp 1–27

Guo H, Sun J, Li X, Xiong Y, Wang H, Shu H, Wang Y (2018) Positive charge in the n-region of the signal peptide contributes to efficient post-translational translocation of small secretory preproteins. J Biol Chem 293(6):1899–1907

Hagiwara H, Seki T, Ariga T (2004) The effect of pre-germinated brown rice intake on blood glucose and PAI-1 levels in streptozotocin-induced diabetic rats. Biosci Biotechnol Biochem 68(2):444–447

Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J (2017) Protein–Sol: a web tool for predicting protein solubility from sequence. Bioinformatics 33(19):3098–3100

Idicula-Thomas S, Kulkarni AJ, Kulkarni BD, Jayaraman VK, Balaji PV (2005) A support vector machine-based method for predicting the propensity of a protein to be soluble or to form inclusion body on overexpression in *Escherichia coli*. Bioinformatics 22(3):278–284

Inoue K, Shirai T, Ochiai H, Kasao M, Hayakawa K, Kimura M, Sansawa H (2003) Blood-pressure-lowering effect of a novel fermented milk containing γ-aminobutyric acid (GABA) in mild hypertensives. Eur J Clin Nutr 57(3):490

Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci 12(8):1652–1662

Kim S, Jeong H, Kim EY, Kim JF, Lee SY, Yoon SH (2017) Genomic and transcriptomic landscape of *Escherichia coli* BL21 (DE3). Nucleic Acids Res 45(9):5285–5293

Komatsuzaki N, Shima J, Kawamoto S, Momose H, Kimura T (2005) Production of γ-aminobutyric acid (GABA) by *Lactobacillus paracasei* isolated from traditional fermented foods. Food Microbiol 22(6):497–504

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157(1):105–132

Lee KW, Shim JM, Yao Z, Kim JA, Kim HJ, Kim JH (2017) Characterization of a glutamate decarboxylase (GAD) from *Enterococcus avium* M5 isolated from jeotgal, a Korean fermented seafood. J Microbiol Biotechnol 27:1216–1222

Lewenza S, Mhlanga MM, Pugsley AP (2008) Novel inner membrane retention signals in *Pseudomonas aeruginosa* lipoproteins. J Bacteriol 190(18):6119–6125

Lim HS, Cha IT, Lee H, Seo MJ (2016) Optimization of γ-aminobutyric acid production by *Enterococcus faecium* JK29 isolated from a traditional fermented foods. Microbiol Biotechnol Lett 44:26–33

Low KO, Mahadi NM, Illias RM (2013) Optimization of signal peptide for recombinant protein secretion in bacterial hosts. Appl Microbiol Biotechnol 97(9):3811–3826

Magnan CN, Randall A, Baldi P (2009) SOLpro: accurate sequence-based prediction of protein solubility. Bioinformatics 25(17):2200–2207

Mergulhao FJ, Monteiro GA, Cabral JM, Taipa MA (2004) Design of bacterial vector systems for the production of recombinant proteins in *Escherichia coli*. J Microbiol Biotechnol 14(1):1–14

Mogensen JE, Otzen DE (2005) Interactions between folding factors and bacterial outer membrane proteins. Mol Microbiol 57(2):326–346

Mohammadi S, Mostafavi-Pour Z, Ghasemi Y, Barazesh M, Pour SK, Atapour A, Morowvat MH (2019) In silico analysis of different signal peptides for the excretory production of recombinant NS3-GP96 fusion protein in *Escherichia coli*. Int J Pept Res Ther 25(4):1279–1290

Möhler H (2012) The GABA system in anxiety and depression and its therapeutic potential. Neuropharmacology 62(1):42–53

Natale P, Brüser T, Driessen AJ (2008) Sec-and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms. Biochimica et Biophysica Acta (BBA) 1778(9):1735–1756

Nielsen H (2017) Predicting secretory proteins with SignalP. In: Kihara D (ed) Protein function prediction. Humana Press, New York, pp 59–73

Owji H, Nezafat N, Negahdaripour M, Hajiebrahimi A, Ghasemi Y (2018) A comprehensive review of signal peptides: structure, roles, and applications. Eur J Cell Biol 97(6):422–441

Paladin L, Piovesan D, Tosatto SC (2017) SODA: prediction of protein solubility from disorder and aggregation propensity. Nucleic Acids Res 45(W1):W236–W240

Papanikou E, Karamanou S, Economou A (2007) Bacterial protein secretion through the translocase nanomachine. Nat Rev Microbiol 5(11):839

Petersen TN, Brunak S, Von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785

Pratap J, Dikshit KL (1998) Effect of signal peptide changes on the extracellular processing of streptokinase from *Escherichia coli*: requirement for secondary structure at the cleavage junction. Mol Gen Genet MGG 258(4):326–333

Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol 5:172

Santos CA, Beloti LL, Toledo MA, Crucello A, Favaro MT, Mendes JS, Souza AP (2012) A novel protein refolding protocol for the solubilization and purification of recombinant peptidoglycan-associated lipoprotein from *Xylella fastidiosa* overexpressed in *Escherichia coli*. Protein Expr Purif 82(2):284–289

Sezonov G, Joseleau-Petit D, d'Ari R (2007) *Escherichia coli* physiology in Luria-Bertani broth. J Bacteriol 189(23):8746–8749

Shen HB, Chou KC (2010) Gneg-mPLoc: a top-down strategy to enhance the quality of predicting subcellular localization of Gram-negative bacterial proteins. J Theor Biol 264(2):326–333

Shiloach J, Fass R (2005) Growing *E. coli* to high cell density—a historical perspective on method development. Biotechnol Adv 23(5):345–357

Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, Li Q (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. Protein Sci 18(5):936–948

Smialowski P, Schmidt T, Cox J, Kirschner A, Frishman D (2006a) Will my protein crystallize? A sequence-based predictor. Proteins 62(2):343–355

Smialowski P, Martin-Galiano AJ, Mikolajka A, Girschick T, Holak TA, Frishman D (2006b) Protein solubility: sequence based prediction and experimental verification. Bioinformatics 23(19):2536–2542

Smialowski P, Doose G, Torkler P, Kaufmann S, Frishman D (2012) PROSO II—a new method for protein solubility prediction. FEBS J 279(12):2192–2200

Ueno H (2000) Enzymatic and structural aspects on glutamate decarboxylase. J Mol Catal B 10(1–3):67–79

Vahedi F, Nassiri M, Ghovvati S, Javadmanesh A (2019) Evaluation of different signal peptides using bioinformatics tools to express recombinant erythropoietin in mammalian cells. Int J Pept Res Ther 25(3):989–995

Von Heijne G, Abrahmsèn L (1989) Species-specific variation in signal peptide design Implications for protein secretion in foreign hosts. FEBS Lett 244(2):439–446

Yu K, Lin L, Hu S, Huang J, Mei L (2012) C-terminal truncation of glutamate decarboxylase from *Lactobacillus brevis* CGMCC 1306 extends its activity toward near-neutral pH. Enzyme Microb Technol 50(4–5):263–269

Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, Lu CH (2014) CELLO2GO: a web server for protein subcellular localization prediction with functional gene ontology annotation. PLoS ONE 9(6):e99368

Zamani M, Nezafat N, Negahdaripour M, Dabbagh F, Ghasemi Y (2015) In silico evaluation of different signal peptides for the secretory production of human growth hormone in *E. coli*. Int J Pept Res Ther 21(3):261–268