E-NAMTILA Publishing
DLS **6** (2025) 7-16

DYSONA - Life Science
DOI: 10.30493/DLS.2024.475157

Investigation of periplasmic signal peptides for the expression of Thanatin peptide in *Escherichia coli*

Elnaz Karbaschian [1]; Ali Javadmanesh [1]*

1, Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

E-mail:
javadmanesh@um.ac.ir

Received: 11/12/2024
Acceptance: 21/02/2025
Available Online: 21/02/2025
Published: 01/03/2025

Keywords: Signal peptides, Thanatin, *Escherichia coli*

Abstract

The expression of periplasmic antimicrobial peptides is a critical aspect in cloning systems and protein expression. In the present study, bioinformatics methods were utilized to predict the optimal signal peptides for the expression of thanatin in *Escherichia coli* host. A total of 60 confirmed species-specific signal peptide sequences were obtained from signal peptide database. Since the aim of this study was to identify optimal signal peptides for periplasmic expression, the environment in which the protein operates, cleavage site, solubility, secretion probability, and the physical and chemical properties of the signal peptides were analyzed using SignalP6, SignalP4.1, Protein-Sol, and ProtParam. In the initial stage, 22 out of the 60 signal peptides passed through the SignalP6 and SignalP4.1 criteria. Solubility and secretion analysis of the signal peptides indicated that 22 signal peptides were suitable. Finally, the examination of physicochemical properties revealed that 21 signal peptides were the most suitable for the expression of thanatin. These final signal peptides were ranked by their D-Score. Among the top candidate signal peptides, E2BB, FMF4, PCOE, HSTI, FAEF were confirmed by other research.

1. Introduction

The increasing resistance of bacteria to common antibiotics necessitates the introduction of new drugs to combat infections. However, it is crucial to explore new classes of antibiotics with different mechanisms of action to inhibit bacterial growth and infection [1]. Among the wide range of antimicrobial compounds, antimicrobial proteins and peptides hold a significant position. Therefore, their investigation and evaluation are essential and highly effective [2].

Over the past few decades, the use of antibiotics in the livestock and poultry sectors for disease prevention, treatment, and growth enhancement has gained attention. The European Union Federation of Animal Health has reported that a significant portion of all antibiotics prescribed in the EU is administered to farm animals. While antibiotics offer benefits in terms of animal health and performance, their continuous use leads to multiple issues concerning both animal and human health [3]. Among these challenges, the transmission of antibiotic-resistant microorganisms holds a special place. Apart from the economic losses, this issue has created complexities in medical treatment. Moreover, it can act as a reservoir for resistance genes, posing long-term risks to human health [4]. In recent years, with the increase in resistance to conventional antimicrobial agents, numerous researchers have initiated studies to develop new antibiotics to manage and mitigate the harmful effects of infectious pathogens [5].

Antimicrobial peptides (AMPs) are a relatively new class of antibiotics of natural origin. These peptides are an integral part of the immune system of living organisms, playing a crucial role in microbial interactions and enhancing the innate immunity of organisms by eliminating invasive microorganisms. Additionally, AMPs have the ability to stimulate cells to produce chemokines, promote angiogenesis, accelerate wound healing, and influence programmed cell death in multicellular organisms [6]. Over the past decade, peptides have been recognized as promising

DYSONA Life Science ISSN: 2708-6291

treatments for combating a wide range of diseases such as cancer, diabetes, cardiovascular diseases, and infectious diseases [7].

Most antimicrobial peptides are positively charged and typically consist of 12 to 50 amino acids, with their positive charge mainly due to the presence of basic amino acids like lysine and arginine [8]. Thanatin is a cationic peptide from the beta-helix family, characterized by its positive charge, non-hemolytic nature, and amphipathic properties. This peptide was initially isolated from the insect *Podisus maculiventris* after immune system activation. It is a major component of the insect's primary defense mechanisms and has demonstrated broad efficacy against various pathogens, including Gram-positive and Gram-negative bacteria, as well as fungi [9].

Today, the use of genetic engineering techniques for synthesizing antimicrobial peptides has become prevalent. To produce antimicrobial peptides through genetic engineering methods, determining the appropriate signal peptide for the optimum periplasmic expression of peptides is crucial [10]. With the advancement of technology and the development of DNA recombination techniques, the solubility of heterologous proteins has gained importance. This is because the accumulation of insoluble proteins and misfolding can lead to the formation of inclusion bodies, which, on the other hand, also reduce the efficiency of properly folded proteins [11]. To address this issue, the expression of heterologous proteins in the periplasmic space of it, is suggested in *Escherichia coli*. [12]. Translocating proteins to the periplasmic space facilitate purification since there are fewer proteins present in this space. Other advantages of using the periplasmic expression system include preventing protease attack, preventing N-terminal Met extension, and having better-folded proteins [11].

There are a large number of applications to predict the possibility of a suitable signal sequence such as SignalP, ProtParam and Protein-Sol [13]. These applications could suggest the best possible signal sequences for optimized expression of proteins or peptides in a specific host. Research by Kumari et al illustrated combined computational predictions with experimental validation to study lipoprotein and Tat signal peptides in *Anabaena sp.* PCC7120. It utilized tools like SignalP for prediction and confirmed findings through heterologous expression and activity assays, providing a practical example of integrating computational and experimental approaches in signal peptide research [14].

Therefore, the primary objective of the research is to extend the application of bioinformatics methods in identifying suitable signal peptides for periplasmic expression of thanatin in *Escherichia coli*. To achieve this goal, a comprehensive analysis of confirmed signal peptide sequences was conducted, focusing on evaluating their molecular and physicochemical characteristics using optimized online tools.

2. Material and Methods

2.1. Retrieval of thanatin sequence from the antimicrobial database

In this study, the amino acid sequence of thanatin (GSKKPVPIIYCNRRTGKCQRM) was obtained from the research conducted by [15].

2.2. Collection of signal peptides

To obtain the signal peptide sequences, the Signal Peptide Database (http://www.signalpeptide.de/) was utilized. The selection of signal peptides was based on all confirmed signal peptides for *Escherichia coli* (Table 1).

2.3. Prediction of suitable signal peptides

2.3.1. SignalP6 and SingnalP4.1

Initially, the SignalP6 (https://services.healthtech.dtu.dk/services/SignalP-6.0/) and SignalP4.1 (https://services.healthtech.dtu.dk/services/SignalP-4.1/) servers were used to analyze the 60 confirmed signal peptides intended for expression in *Escherichia coli*. The aim of this analysis was to determine the cleavage sites, probability, and Sec/Tat by SignalP6 and D, Y, S, C-Scores and S-mean parameters by SignalP4.1 for all 60 signal peptides. These web-based servers predict the presence of signal peptides and identifies their cleavage sites in Gram-positive bacteria, Gram-negative bacteria, archaea, and eukaryotes.

2.3.2. Protein-Sol

Next, the solubility of the signal peptides was evaluated using the online software Protein-Sol (http://protein-sol.manchester.ac.uk). Protein-Sol is an open-access web server that provides predicted solubility on a scale from 0 to 1, simplifying the interpretation of results [16]. Unstable and insoluble proteins were subsequently excluded.

2.3.3. ProtParam

ProtParam (https://web.expasy.org/protparam/) was used to predict the physicochemical properties related to amino acid composition, molecular weight, isoelectric point (PI), instability index, aliphatic index, positively and negatively charged amino acids, and the grand average of hydropathicity (GRAVY) of the signal peptides [17].

3. Results and Discussion

There were 60 confirmed signal peptides in the Signal peptide database (Table 1). As shown in Table 2, signal peptides evaluated using SignalP6 and SignalP4.1 servers, only 22 cases were recognized as proper signal peptides with an acceptable accuracy (Table 3). The probability was considered to be higher than 90% and the D-Score was considered larger than 0.45 [18]. Then they were selected for the next step of analysis with ProtParam and Protein-Sol servers.

Table 1. List of all confirmed signal peptides in Escherichia coli obtained from the Signal Peptide Database

Accession number	Signal peptide	Full name	Number of amino acid	Amino acid sequence
Q47117	Q47117	n/a	23	MKLNKIIGALVLSSTFVSMGASA
P04738	FAEC	K88 fimbrial protein A	21	MKKAFLLACVFFLTGGGVSHA
P14191	FAEG3	K88 fimbrial protein AD	21	MKKTLIALAIAASAASGMAHA
P28585	BLC1	Beta-lactamase CTX-M-1	28	MVKKSLRQFTLMATATVTLLLGSVPLYA
P0AD63	BLA1	Beta-lactamase SHV-1	21	MRYIRLCIISLLATLPLAVHA
P02970	FAEG1	K88 fimbrial protein AB	21	MKKTLIALAIAASAASGMAHA
P06875	PAC	Penicillin G acylase	26	MKNRNRMIVNCVTASLMYYWSLPALA
P62593	BLAT	Beta-lactamase TEM	23	MSIQHFRVALIPFFAAFCLPVFA
P43528	E2BA	Heat-labile enterotoxin IIB, A chain	20	MAKVISFFISLFLISFPLYA
P24093	DRAA	Dr hemagglutinin structural subunit subunit	21	MKKLAIMAAASMVFAVSSAHA
P62605	FIM1C	Type-1 fimbrial protein, C chain	23	MKLKFISMAVFSALTLGVATNAS
A2TJI4	CEXE	Protein cexE	19	MKKYILGVILAMGSLSAIA
O88093	HBP	Hemoglobin-binding protease hbp	52	MNRIYSLRYSAVARGFIAVSEFARKCVHKS VRRLCFPVLLLIPVLFSAGSLA
P32890	ELBP	Heat-labile enterotoxin B chain	21	MNKVKCYVLFTALLSSLYAHG
Q03155	AIDA	Adhesin AIDA-I	49	MNKAYSIIWSHSRQAWIVASELARGHGFV LAKNTLLVLAVVSTIGNAFA
P25394	FMF7	F107 fimbrial protein	21	MKRLVFISFVALSMTAGSAMA
P02971	FMC1	CFA/I fimbrial subunit B	23	MKFKKTIGAMALTTMFVAVSASA
P13720	PAPG	Fimbrial adhesin papG	21	MKKWFPAFLFLSLSGGNDALA
P43529	E2BB	Heat-labile enterotoxin IIB, B chain	23	MSFKKIIKAFVIMAALVSVQAHA
P07965	HST3	Heat-stable enterotoxin A3/A4	19	MKKSILFIFLSVLSFSPFA
P11312	FMF3	F17 fimbrial protein	21	MQKIQFILGILAAASSSATLA
P0ABE7	C562	Soluble cytochrome b562	22	MRKSLLAILAVSSLVFSSASFA
P0A9Z7	BLA2	Beta-lactamase SHV-2	21	MRYIRLCIISLLATLPLAVHA
Q99003	F17AG	F17a-G fimbrial adhesin	22	MTNFYKVFLAVFILVCCNISQA

Table 1. Continued

Accession number	Signal peptide	Full name	Number of amino acid	Amino acid sequence
P25447	FAEF	K88 minor fimbrial subunit faeF	22	MKKTMMAAALVLSALSIQSALA
P07111	PAPH	PAP fimbrial minor pilin protein	22	MRLRFSVPLFFFGCVFVHGVFA
P09181	LYS4	Lysis protein for colicin N	17	MCGKILLILFFIMTLSA
P06970	FAED	Outer membrane usher protein faeD	35	MKKYVTTKSVQPVAFRLTTLSLVMSAVLG SASVIA
Q47459	PCOE		20	MKKILVSFVAIMAVASSAMA
P22542	HSTI	Heat-stable enterotoxin II	23	MKKNIAFLLASMFVFSIATNAYA
P05818	BMAE	M-agglutinin	24	MNLKKIAIASSVFAGITMALTCHA
P17543	PAPJ	Protein papJ	27	MVVNKTTAVLYLIALSLSGFIHTFLRA
Q47692	TSH	Temperature-sensitive hemagglutinin tsh	52	MNRIYSLRYSAVARGFIAVSEFARKCVHKS VRRLCFPVLLLIPVLFSAGSLA
Q47196	Q47196	n/a	21	MRKIQFILGILAAASSSATLA
Q07685	Q07685	n/a	40	MAHSFYVNFGKKNIWRMSEMKKTFIASA IAITINTGSAIA
O32591	ESPP	Serine protease espP	55	MNKIYSLKYSHITGGLIAVSELSGRVSSRA TGKKKHKRILALCFLGLLQSSYSFA
P11900	FMF4	F41 fimbrial protein	22	MKKTLIALAVAASAAVSGSVMA
P07110	PAPC	Outer membrane usher protein papC	24	MKDRIPFAVNNITCVILLSLFCNA
Q9XD84	TIBA	Adhesin/invasin tibA	54	MNKVYNTVWNESTGTWVVTSELTRKGG LRPRQIKRTVLAGLIAGLLMPSMPALA
P62607	FMF2	F7-2 fimbrial protein	21	MIKSVIAGAVAMAVVSFGAYA
P13810	E2AA	Heat-labile enterotoxin IIA, A chain	18	MIKHVLLFFVFISFSVSA
P15319	PAPD	Chaperone protein papD	21	MIRKKILMAAIPLFVISGADA
P13811	ELBH	Heat-labile enterotoxin B chain	21	MNKVKFYVLFTALLSSLCAHG
P20861	FANG	Protein fanG	21	MKKLYKAITVICILMSNLQSA
P20862	FANH	Protein fanH	20	MIKKVPVLLFFMASISITHA
Q47066	BLT1	Beta-lactamase Toho-1	29	MMTQSIRRSMLTVMATLPLLFSSATLHA Q
O68900	PET	Serine protease pet	52	MNKIYSIKYSAATGGLIAVSELAKKVICKT NRKISAALLSLAVISYTNIIYA
P18103	FANC	K99 fimbrial protein	20	MKKTLLAIILGGMAFATTNASA
Q03961	KPSD1	Polysialic acid transport protein kpsD	22	MKLFKSILLIAACHAAQASA
P21413	FM98	Fimbrial protein 987P	23	MRMKKSALTLAVLSSLFSGYSLA
P14542	IUTA	Ferric aerobactin receptor	25	MMISKKYTLWALNPLLLTMMAPAVA
P0ABW7	CSOA	CS1 fimbrial subunit A	23	MKLKKTIGAMALATLFATMGASA
P01559	HST1	Heat-stable enterotoxin ST-IA/ST-P	19	MKKLMLAIFISVLSFPSFS
P06717	ELAP	Heat-labile enterotoxin A chain	18	MKNITFIFFILLASPLYA
P13812	E2AB	Heat-labile enterotoxin IIA, B chain	19	MSSKKIIGAFVLMTGILSG
P14190	FAEG2	K88 fimbrial protein AC	21	MKKTLIALAIAASAASGMAHA
P04127	PAPA	Pap fimbrial major pilin protein	22	MIKSVIAGAVAMAVVSFGVNNA
P46739	NFAA	Non-fimbrial adhesin 1	28	MKAKKYENQIYNENGRRCQRHGRRLAIA
Q7BS42	PIC	Serine protease pic	55	MNKVYSLKYCPVTGGLIAVSELARRVIKK TCRRLTHILLAGIPAICLCYSQISQA
P15488	FMS3	CS3 fimbrial subunit A	22	MLKIKYLLIGLSLSAMSSYSLA

Table 2. Analysis of all signal peptides using SingnalP6 and SignalP4.1

Signal peptide	Cleavage site between pos	Probability	Sec/TAT	D-score	S-score	Y-score	C-score	S-mean
FAEC	21 - 22	0.968210	Sec	0.672	0.893	0.612	0.541	0.738
BLC1	28 - 29	0.920558	Sec	0.388	0.669	0.382	0.496	0.397
BLA1	21 - 22	0.978446	Sec	0.808	0.958	0.745	0.646	0.879
FAEG1	21 - 22	0.968949	Sec	0.765	0.927	0.695	0.611	0.843
PAC	26 - 27	0.887768	Sec	0.452	0.585	0.467	0.545	0.428
BLAT	16 - 17	0.762495	Others	0.347	0.447	0.352	0.438	0.339
E2BA	-	-	Others	0.322	0.443	0.351	0.527	0.272
DRAA	21 - 22	0.977010	Sec	0.765	0.945	0.661	0.518	0.882
FIM1C	22 - 23	0.955511	Sec	0.608	0.939	0.366	0.158	0.882
CEXE	-	-	Others	0.459	0.664	0.448	0.495	0.478
HBP	52 - 53	0.495193	Others	0.221	0.440	0.224	0.195	0.217
ELBP	21 - 22	0.855603	Sec	0.715	0.900	0.692	0.685	0.742
AIDA	49 - 50	0.539807	Others	0.217	0.280	0.253	0.352	0.157
FMF7	21 - 22	0.962829	Sec	0.770	0.947	0.684	0.566	0.866
FMC1	23 - 24	0.976038	Sec	0.702	0.959	0.959	0.354	0.879
PAPG	21 - 22	0.899796	Sec	0.510	0.852	0.431	0.371	0.600
E2BB	23 - 24	0.973732	Sec	0.836	0.948	0.827	0.835	0.847
HST3	21 - 22	0.950200	Sec	0.717	0.904	0.620	0.477	0.828
FMF3	21 - 22	0.879422	Sec	0.587	0.872	0.436	0.273	0.759
C562	22 - 23	0.972444	Sec	0.798	0.961	0.705	0.569	0.903
F17AG	22 - 23	0.979892	Sec	0.681	0.906	0.618	0.526	0.752
FAEF	22 - 23	0.974962	Sec	0.817	0.969	0.734	0.608	0.910
PAPH	22 - 23	0.974962	Sec	0.256	0.400	0.230	0.209	0.301
LYS4	20 - 21	0.497634	Others	0.270	0.561	0.200	0.200	0.389
FAED	-	-	Others	0.423	0.786	0.368	0.342	0.516
PCOE	20 - 21	0.973792	Sec	0.824	0.955	0.754	0.641	0.903
HSTI	23 - 24	0.974405	Sec	0.817	0.948	0.781	0.769	0.858
BMAE	21 - 22	0.619527	Others	0.650	0.865	0.590	0.551	0.718
PAPJ	-	-	0	0.247	0.538	0.221	0.211	0.292
TSH	52 - 53	0.495193	Others	0.221	0.440	0.224	0.195	0.217
Q47196	21 - 22	0.897869	Sec	0.624	0.901	0.457	0.276	0.812
Q07685	40 - 41	0.707339	Others	0.211	0.343	0.215	0.240	0.206
ESPP	-	-	Others	0.350	0.436	0.422	0.614	0.229
FMF4	22 - 23	0.970273	Sec	0.827	0.945	0.803	0.816	0.853
PAPC	-	-	Others	0.296	0.497	0.265	0.249	0.349
TIBA	54 - 55	0.386910	Others	0.328	0.550	0.409	0.550	0.190
FMF2	-	-	Others	0.385	0.717	0.366	0.416	0.417
E2AA	18 - 19	0.971150	Sec	0.379	0.600	0.362	0.383	0.408
PAPD	21 - 22	0.965723	Sec	0.605	0.894	0.473	0.317	0.753
ELBH	17 - 18	0.801002	Others	0.675	0.925	0.570	0.441	0.793

Table 2. Continued

Signal peptide	Cleavage site between pos	Probability	Sec/TAT	D-score	S-score	Y-score	C-score	S-mean
FANG	23 - 24	0.803277	Others	0.456	0.818	0.298	0.149	0.633
FANH	20 - 21	0.976958	Sec	0.670	0.890	0.607	0.524	0.742
BLT1	28 - 29	0.803734	TAT	0.796	0.953	0.727	0.644	0.873
PET	-	-	Others	0.224	0.409	0.175	0.185	0.306
FANC	22 - 23	0.969995	Sec	0.696	0.905	0.593	0.467	0.812
KPSD1	12 - 13	0.614438	Others	0.660	0.888	0.563	0.443	0.769
FM98	23 - 24	0.970732	Sec	0.802	0.926	0.753	0.699	0.858
IUTA	25 - 26	0.839579	Others	0.751	0.962	0.620	0.444	0.899
CSOA	23 - 24	0.977405	Sec	0.710	0.954	0.549	0.353	0.892
HST1	19 - 20	0.808960	Others	0.633	0.875	0.494	0.320	0.790
ELAP	18 - 19	0.974561	0.9978	0.311	0.524	0.294	0.289	0.341
E2AB	0.118	0.829254	Others	0.224	0.430	0.177	0.118	0.304
PAPA	-	-	Others	0.534	0.868	0.399	0.274	0.685
NFAA	-	-	0	0.221	0.340	0.169	0.102	0.279
PIC	55 - 56	0.226302	Others	0.227	0.321	0.242	0.286	0.202
FMS3	22 - 23	0.906635	Sec	0.772	0.915	0.737	0.693	0.812

It has been reported that the SignalP6 server has the highest accuracy among other prediction softwares. SignalP6 can predict additional types of signal peptides that previous versions have been unable to detect while better extrapolating to both proteins distantly related to those used to create the model and metagenomic data of unknown origin. In addition, it is capable of identifying the subregions of signal peptides [19]. Then this server and SignalP4.1 were used for present study. As shown in Table 2, among the types of protein secretion systems (types I, II, and III), type II is the most commonly used. The type II system consists of three secretory pathways (SecB-dependent (Sec), signal recognition particle (SRP), and dual arginine transport (TAT) pathways) and is widely distributed among gram-negative bacteria [20].

The most important parameter for the diagnosis of a signal peptide is the discriminating score (D-score) which is usually described with a cut-of value of 0.45, and when the D score of a signal peptide is more than 0.45, the sequence is considered as the signal peptide [10]. D score is known as one the most important parameters to identify a signal peptide [18]. According to the data presented in Table 2, 22 signal peptides with D score higher than 0.45 were considered as proper signal peptides (Table 3).

Table 3. Analysis of selected signal peptides using ProtParam and Protein-Sol servers

Signal peptide	*MW	*PI	Gravity	Net charge	*AI	Stability	*H	Solubility
FAEC	4714.75	10.33	0.223	+8	79.30	Stable	>10	0.752
BLA1	5016.25	10.55	0.389	+8	115.23	Stable	>10	0.635
FAEG1	4544.54	10.66	0.149	+8	86.51	Stable	>10	0.826
DRAA	4652.70	10.66	0.226	+8	79.53	Stable	>10	0.853
FIM1C	4917.99	10.66	0.238	+8	86.67	Stable	>10	0.835
FMF7	4748.83	11.06	0.288	+8	81.63	Stable	>10	0.778
FMC1	4922.04	10.72	0.144	+9	71.56	Stable	>10	0.896
HST3	4692.77	10.66	0.298	+8	90.24	Unstable	>10	0.837
E2BB	5021.20	10.72	0.302	+9	95.33	Stable	>10	0.741
C562	4815.84	11.06	0.339	+8	97.50	Stable	>10	0.778
F17AG	5042.15	9.85	0.368	+7	90.68	Stable	>10	0.726
FAEF	4766.88	10.66	0.257	+8	95.45	Stable	>10	0.826
PCOE	4679.69	11.06	0.198	+8	97.67	Stable	>10	0.827
HSTI	5069.16	10.40	0.198	+8	82.44	Stable	>10	0.735
FMF4	4607.64	10.66	0.343	+8	93.18	Stable	>10	0.827
PAPD	4775.92	10.64	0.240	+9	102.09	Stable	>10	0.846
FM98	4992.07	10.70	0.040	+9	86.67	Stable	>10	0.794
FANH	4763.91	10.66	0.307	+8	97.38	Stable	>10	0.829
FANC	4740.79	10.66	0.143	+8	86.59	Stable	>10	0.812
CSOA	4843.97	10.72	0.113	+9	78.22	Stable	>10	0.848
Q47196	4679.69	11.06	0.198	+8	97.67	Stable	>10	0.779
FMS3	4920.05	10.24	0.220	+8	106.36	Stable	>10	0.722

* MW: Molecular weight, PI: Isoelectric point, AI: Aliphatic index, H: Half-life considered in *Escherichia coli*, in vivo.

The results of the analysis of the physical and chemical properties of the signal peptides, including amino acid composition, molecular weight (MW), isoelectric point (PI), instability index, aliphatic index, positive and negative charged residues, and the grand average of hydropathicity (GRAVY) as determined by the ProtParam server. ProtParam, a part of ExPASy and the European Bioinformatics Institute (EIB), is used for reliability of estimating the physical and chemical properties of proteins [21]. ProtParam computes various physicochemical properties that can be deduced from a protein sequence. The current study showed that the highest and lowest molecular weights (MW) corresponded to HSTI (5069.16) and FAEG1 (4544.54), respectively. Studies have shown that signal peptides with lower pI values (below 5.5) are often associated with higher secretion efficiencies due to their negative charge, which may promote better interactions with positively charged membrane components [22]. Results of the current research indicated that the PI for all selected signal peptide were above 10. The grand average of hydropathicity (GRAVY) is used to compare the hydrophobicity of signal peptides [23]. The findings demonstrated that the lowest GRAVY belonged to FM98 (0.040) and the highest was attributed to BLA1 (0.389).

The aliphatic index, which represents the relative volume occupied by aliphatic side chains in an amino acid sequence, can be considered a positive factor for increasing the thermal stability of globular proteins. The variations in this index for the signal peptides that were more appropriate including 95.33, 93.18, 97.67, 82.44, 95.45 for E2BB, FMF4, PCOE, HSTI, FAEF respectively. Generally, proteins with an instability index above 40 are considered unstable, but all peptides excluding HST3 were reported as stable, indicating that HST3 has a high likelihood of degradation [23].

Overall, while most characteristics between these 22 signal peptides were similar, E2BB, FMF4, PCOE, HSTI, FAEF showed more favorable values in terms of solubility, GRAVY, and aliphatic index.

Various web servers have been generated to predict the solubility of proteins in *Escherichia coli* and Protein Sol server is one of the most accurate solubility prediction servers [7]. The numerical value reported in the Protein Sol server is between zero and one [16]. The results of the solubility prediction for signal peptides are reported in Table 3.

Table 4: List of signal peptides suitable for secretory expression of thanatin peptide

Signal peptide	Sequence	Length	Probability	D-score
E2BB	MSFKKIIKAFVIMAALVSVQAHA	23	0.973732	0.836
FMF4	MKKTLIALAVAASAAVSGSVMA	22	0.970273	0.827
PCOE	MRKIQFILGILAAASSSATLA	20	0.973792	0.824
HSTI	MKKNIAFLLASMFVFSIATNAYA	23	0.974405	0.817
FAEF	MKKTMMAAALVLSALSIQSALA	22	0.974962	0.817
BLA1	MRYIRLCIISLLATLPLAVHA	21	0.978446	0.808
FM98	MRMKKSALTLAVLSSLFSGYSLA	23	0.970732	0.802
C562	MRKSLLAILAVSSLVFSSASFA	22	0.972444	0.798
FMS3	MLKIKYLLIGLSLSAMSSYSLA	21	0.906635	0.772
FMF7	MKRLVFISFVALSMTAGSAMA	21	0.962829	0.770
FAEG1	MKKTLIALAIAASAASGMAHA	21	0.968949	0.765
DRAA	MKKLAIMAAASMVFAVSSAHA	21	0.977010	0.765
CSOA	MKLKKTIGAMALATLFATMGASA	23	0.977405	0.710
FMC1	MQKIQFILGILAAASSSATLA	23	0.976038	0.702
FANC	MKKTLLAIILGGMAFATTNASA	20	0.969995	0.696
F17AG	MTNFYKVFLAVFILVCCNISQA	22	0.979892	0.681
FAEC	MKKAFLLACVFFLTGGGVSHA	21	0.968210	0.672
FANH	MIKKVPVLLFFMASISITHA	20	0.976958	0.670
Q47196	MRKIQFILGILAAASSSATLA	21	0.897869	0.624
FIM1C	MKLKFISMAVFSALTLGVATNAS	23	0.955511	0.608
PAPD	MIRKKILMAAIPLFVISGADA	21	0.965723	0.605

The length of all signal peptides originated from gram-negative bacteria and ranged between 17 and 23 amino acids. Regularly, a signal peptide sequence in gram-negative bacteria contains 18-30 amino acids. Therefore, it can be stated that the studied signal peptides are in a proper range [24]. Results of this study also confirmed the ideal length of signal peptides which is between 17 and 23 amino acids (Table 4). Summary of all possible signal peptides (21 signal peptides), sorted based on D-Score, are presented in Table 4. Among top signal peptides, E2BB, FMF4, PCOE, HSTI, and FAEF also were used to express periplasmic recombinant proteins in *Escherichia coli* [9]. Out of 60 signal peptides from table 1, only 21 were predicted to be suitable for preplasmic expression of thanatin peptide in *E. coli*, these large number of predicted suitable signals are much higher than other studies [10][13]. The reason could be that in this research, we used an updated database which only confirmed signal peptides were selected. In previous studies, researchers used mixture of predicted and experimental signal peptides. The present study aimed to identify and predict suitable signal peptides for the periplasmic expression of the peptide thanatin in *Escherichia coli*.

4. Conclusion

The findings from this study can be utilized to design expression constructs for the periplasmic expression of thanatin in *Escherichia coli* as host. However, laboratory validations are recommended for final confirmation.

5. Acknowledgments

This study was supported by Ferdowsi University of Mashhad, grant number: 2/47173.

Conflict of interest statement

The authors declared no conflict of interest.

Funding statement

The authors declared that this study was supported by Ferdowsi University of Mashhad, grant number: 2/47173.

Data availability statement

The authors declared that all related data are included in the article.

References

1. Mancuso G, Midiri A, Gerace E, Biondo C. Bacterial antibiotic resistance: the most critical pathogens. Pathogens. 2021;10(10):1310. DOI
2. Rashidian Z, Roshanak S, Sekhavati MH, Javadmanesh A. Synergistic effects of nisin and CLF36 antimicrobial peptides *in vitro*. Vet. Res. Biol. Prod. 2023;36(3):44-50. DOI
3. Pokharel S, Shrestha P, Adhikari B. Antimicrobial use in food animals and human health: time to implement 'One Health' approach. Antimicrob. Resist. Infect. Control. 2020;9:1-5. DOI
4. Bobate S, Mahalle S, Dafale NA, Bajaj A. Emergence of environmental antibiotic resistance: Mechanism, monitoring and management. Environ. Adv. 2023;11:100409. DOI
5. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Aguilar GR, Gray A, Han C, Bisignano C, Rao P, Wool E, Johnson SC. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399(10325):629-55. DOI
6. Zhang QY, Yan ZB, Meng YM, Hong XY, Shao G, Ma JJ, Cheng XR, Liu J, Kang J, Fu CY. Antimicrobial peptides: mechanism of action, activity and clinical potential. Military Med. Res. 2021;8:1-25. DOI
7. Wang L, Wang N, Zhang W, Cheng X, Yan Z, Shao G, Wang X, Wang R, Fu C. Therapeutic peptides: current applications and future directions. Signal Transduct. Target Ther. 2022;7(1):48. DOI
8. Clark S, Jowitt TA, Harris LK, Knight CG, Dobson CB. The lexicon of antimicrobial peptides: a complete set of arginine and tryptophan sequences. Commun. Biol. 2021;4(1):605. DOI
9. Mousavi P, Morowvat MH, Mostafavi-Pour Z, Aram F, Malekzadeh K, Nezafat N, Ghasemi Y. Experimental analysis of E2BB (LTIIb) signal peptide in secretory production of reteplase in *Escherichia coli*. Int. J. Pept. Res. Ther. 2021;27:209-18. DOI
10. Vahedi F, Nassiri M, Ghovvati S, Javadmanesh A. Evaluation of different signal peptides using bioinformatics tools to express recombinant erythropoietin in mammalian cells. Int. J. Pept. Res. Ther. 2019;25:989-95. DOI
11. Khatami SH, Taheri-Anganeh M, Arianfar F, Savardashtaki A, Sarkari B, Ghasemi Y, Mostafavi-Pour Z. Analyzing signal peptides for secretory production of recombinant diagnostic antigen B8/1 from *Echinococcus granulosus*: an *in silico* approach. Mol. Biol. Res. Commun. 2020;9(1):1. DOI
12. Schlegel S, Rujas E, Ytterberg AJ, Zubarev RA, Luirink J, De Gier JW. Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. Microb. Cell Fact. 2013;12:1-12. DOI
13. Forouharmehr A, Nassiri M, Ghovvati S, Javadmanesh A. Evaluation of different signal peptides for secretory production of recombinant bovine pancreatic ribonuclease A in gram-negative bacterial system: an *in silico* study. Curr. Proteomics. 2018;15(1):24-33. DOI
14. Kumari S, Chaurasia AK. *In silico* analysis and experimental validation of lipoprotein and novel Tat signal peptides processing in *Anabaena* sp. PCC7120. J. Microbiol. 2015;53:837-46. DOI

15. Javadmanesh A, Mohammadi E, Mousavi Z, Azghandi M, Tanhaiean A. Antibacterial effects assessment on some livestock pathogens, thermal stability and proposing a probable reason for different levels of activity of thanatin. Sci. Rep. 2021;11(1):10890. DOI

16. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein–Sol: a web tool for predicting protein solubility from sequence. Bioinformatics. 2017;33(19):3098-100. DOI

17. Ali F, Yao Z, Li W, Sun L, Lin W, Lin X. In-silico prediction and modeling of the quorum sensing LuxS protein and inhibition of AI-2 biosynthesis in *Aeromonas hydrophila*. Molecules. 2018;23(10):2627. DOI

18. Roshanak S, Yarabbi H, Shahidi F, Tabatabaei Yazdi F, Movaffagh J, Javadmanesh A. Effects of adding poly-histidine tag on stability, antimicrobial activity and safety of recombinant buforin I expressed in periplasmic space of *Escherichia coli*. Sci. Rep. 2023;13(1):5508. DOI

19. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. 2022;40(7):1023-5. DOI

20. Choi JH, Lee S. Secretory and extracellular production of recombinant proteins using *Escherichia coli*. Appl. Microbiol. Biotechnol. 2004;64:625-35. DOI

21. Gasteiger E. Protein identification and analysis tools on the ExPASy server. In: The Proteomics Protocols Handbook. Humana Press. 2005. DOI

22. Byun H, Park J, Kim SC, Ahn JH. A lower isoelectric point increases signal sequence–mediated secretion of recombinant proteins through a bacterial ABC transporter. J. Biol. Chem. 2017;292(48):19782-91. DOI

23. 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;21:261-8. DOI

24. Auclair SM, Bhanu MK, Kendall DA. Signal peptidase I: Cleaving the way to mature proteins. Protein Sci. 2012;21(1):13–25. DOI