Journal of
Inflammatory Diseases
(Previous Title: The Journal of Qazvin University of Medical Sciences)

Spring 2022. Vol 26. Num 1

Research Paper

Prediction of an Efficient Signal Peptide for Optimized Expression of Mycobacterium Tuberculosis Heparin-binding Hemagglutinin Gene in Periplasmic Compartment of *Escherichia Coli*: A Bioinformatics Investigation

Ali Forouharmehr[1*], Nemat Shams[2], Narges Nazifi[2], Amin Jaydari[2], Ehsan Rashidian[2]

1. Department of Animal Sciences, Faculty of Agriculture, Lorestan University, Khorramabad, Iran.
2. Department of Pathobiology, Faculty of Veterinary Medicine, Lorestan University, Khorramabad, Iran.

Use your device to scan and read the article online

Citation Forouharmehr A, Shams N, Nazifi N, Jaydari A, Rashidian E. Prediction of an Efficient Signal Peptide for Optimized Expression of Mycobacterium Tuberculosis Heparin-binding Hemagglutinin Gene in Periplasmic Compartment of *Escherichia Coli*: A Bioinformatics Investigation. Journal of Inflammatory Diseases. 2022; 26(1):9-18. http://dx.doi.org/10.32598/JID.26.1.4

http://dx.doi.org/10.32598/JID.26.1.4

Article info:
Received: 23 Dec 2021
Accepted: 22 Jan 2022
Publish: 01 Apr 2022

Keywords:
Bioinformatics, Heparin-binding hemagglutinin protein, Signal peptide, Prokaryotic system, Mycobacterium tuberculosis

ABSTRACT

Background: The heparin-binding hemagglutinin (HBHA) protein belonging to Mycobacterium tuberculosis is known as a molecular adjuvant.

Objective: Hence, the expression of this protein in the prokaryotic system is essential.

Methods: To predict an appropriate signal peptide for the expression of the HBHA protein, 50 signal peptides were selected from the signal peptide database. Then, the crucial parameters of signal peptides, including the probability of signal peptide, different regions of signal peptides, physicochemical features, sorting of signal peptides, and sub-cellular location were completely investigated by reliable tools. After the best-predicted signal peptide was identified, it was linked to the HBHA protein, and its secondary structure, tertiary structure, and in silico cloning in pET21a (+) was assessed.

Findings: The results of different evaluations confirmed that only 13 signal peptides passed all features, including clearance of N, H, and C regions, D-score >0.7, instability index >40, and periplasmic localization. Finally, based on D-scores, among these 13 signal peptides, the asr (acid shock RNA) peptide with D-score=90 was selected as the best-predicted signal peptide to apply. Moreover, the results showed that the secondary structure of the adjuvant linked to asr peptide contained 88.18% alpha helix and 9.5% random coil. Also, the results of in silico cloning showed that the nucleotide sequences of the adjuvant linked to the asr peptide were successfully cloned between BamHI and XhoI enzymes in the multiple cloning site of pET21a (+).

Conclusion: The results of this study confirmed that the asr peptide can be used as an appropriate signal peptide for the expression of the HBHA protein in the prokaryotic system.

*** Corresponding Author:**
Ali Forouharmehr
Address: *Department of Animal Sciences, Faculty of Agriculture, Lorestan University, Khorramabad, Iran.*
Phone: *+98 (912) 7240710*
E-mail: *forouharmehr.a@lu.ac.ir*

Journal of
Inflammatory Diseases
(Previous Title: The Journal of Qazvin University of Medical Sciences)

1. Introduction

The heparin-binding hemagglutinin (HBHA) is a lectin-like factor that can be extracted from Mycobacterium tuberculosis (M. tuberculosis) [1]. It has been reported that HBHA can not only agglutinate erythrocytes but also can fortify attachment and invasion of M. tuberculosis to epithelial cells [2-4]. Also, many researchers have shown that this protein has a vital role in dendritic cell maturation, overexpression of surface molecules (e.g. CD40, CD80 and CD86), and stimulation of proinflammatory cytokines, such as IL6, IL12 [5]. Consequently, HBHA protein can be applied as a molecular adjuvant to reinforce immune responses. In many studies, HBHA protein has been employed as a fusion protein in a new generation of vaccines against different infections [6-8]. Hence, the expression of HBHA as a recombinant protein in an efficient condition should be considered. Nowadays, eukaryotic and prokaryotic systems are known as the vital hosts to express recombinant proteins [9]. The eukaryotic system is extensively applied to express recombinant proteins which have post-translational modification (PTM) and complex structure, while the prokaryotic system is widely employed to express simple recombinant proteins without PTM [10]. Given that the prokaryotic system is more cost-effective and user-friendly than the eukaryotic system, consequently, this system is strongly recommended to express simple proteins, such as HBHA which do not need PTM [11]. Although the prokaryotic system has advantages, success in purification of a recombinant protein in this system completely depends on the existence of an appropriate signal peptide at the N-terminal of the nascent protein to be expressed [10, 12].

A signal peptide is an amino acid sequence with 15-25 residues in length that, from N-terminal to the C terminal, contains three obvious parts, including N, H, and C regions. N and H regions of a signal peptide participate in the translocation of nascent protein in the periplasmic compartment, while the C region is known as a cleavable site [13]. It is essential to note that without an appropriate signal peptide, nascent protein cannot be directed to the periplasmic compartment. As a result, the nascent protein will be sediment in the cytoplasm of its host, and it can negatively impact folding and purification processes [10]. Despite the importance of signal peptide role in the expression of a recombinant protein, identification of an appropriate signal peptide using experimental analysis is too expensive and time-consuming. Hence, an alternative strategy should be developed to identify an appropriate signal peptide for each protein. Bioinformatics is known as an affordable strategy that can precisely analyze a large amount of data in a short time using a combination of biology, statistics, and computer sciences. Moreover, along with the extension of bioinformatics, many databases have been developed that can provide researchers with reliable data [10, 13]. Therefore, it seems that applying both bioinformatics and information databases can introduce a vital strategy to find an appropriate signal peptide. The current study was designed to predict an appropriate signal peptide for the expression of HBHA protein in a prokaryotic system. This study was conducted to investigate some signal peptides retrieved from the database and introduce the best signal peptide for the proper folding of HBHA protein in the *Escherichia coli* expression system using some reliable online tools.

2. Material and Methods

Amino acid sequences retrieving

In this bioinformatics study, amino acid sequences of 50 initial signal peptides (Table 1) were collected from the signal peptide database. Also, an amino acid sequence of HBHA protein with the accession number NC_000962.3 was extracted from the national center for biotechnology information database.

Investigation of signal peptide probability

To assess signal peptide probabilities, SignalP 4.1 server was applied. In this case, each signal peptide was linked to the N-terminal of the HBHA amino acid sequence and pasted to the server. Filtration of signal peptides was performed based on D-score (discrimination score) with cut-off=0.7 [10]. The signal peptides with a D-score less than 0.7 were deleted from subsequent analysis.

Investigation of different regions of signal peptides and physicochemical features

To determine different parts of remained signal peptides, including N, H, and C regions. Signal P 3.0 server was employed. In this case, each signal peptide was embedded at the N-terminal of the HBHA protein and the whole amino acid sequence was pasted in the mentioned server. In this step, the signal peptides with unknown cleavable sites were eliminated from the study. Also, different physicochemical features of the remained signal peptides, including pI, instability index (instability index<40), aliphatic index, and grand average of hydropathicity (GRAVY) were assessed by ProtParam server and unstable signal peptide were deleted from the list.

Identification of signal peptide type and sub-cellular location

To identify the type of remained signal peptides, the PRED-TAT server was used. Also, different sub-cellular locations of protein, including cytoplasmic, membrane, secreted, and periplasmic compartments were predicted by ProtcompB server version 9. It should be mentioned that the signal peptides that were not able to locate a protein in the periplasmic compartment were deleted from the study.

Selection of the best signal peptide

To introduce the best signal peptide for optimized expression of HBHA protein, the remained signal peptides which were able to pass all steps of the analysis were ranked based on D-score. The biggest D-score belonged to the best signal peptide.

Secondary and tertiary structures

To investigate the secondary structure of HBHA protein in the presence of the best signal peptide, the SOPMA server was applied. Also, to model the tertiary structure of the HBHA protein associated with the best signal peptide, I-TASSER was employed. In both cases, the amino acid sequence of the best-predicted signal peptide was embedded in the N-terminal of the HBHA protein.

Codon optimization and in silico cloning

In the current study, the nucleotide sequence of the HBHA protein associated with the best-predicted signal peptide was codon-optimized by JCat server expression of the prokaryotic system. Then, restriction sites of BamHI and XhoI enzymes were added to the 5 and 3 of the codon-optimized sequences, respectively. After double digestion with the NcoI and XhoI enzymes, the digested sequence was pasted in multiple cloning sites of the pET21a (+) vector.

3. Results

Investigation of signal peptide probability

As mentioned, in the current study, the signal peptide probability of 50 signal peptides was evaluated based on D-score with cut-off=0.7 According to the results, the highest D-score belonged to asr (0.90) and the lowest value belonged to traV (0.25). Moreover, the results revealed that 36 signal peptides of the initial list had D-score more than 0.7, while the D-score of 14 signal peptides, including sbp, surA, yaeT, phoA, cusF, pldA, rlpB, nlpE, slp, lpp, ompT, dacB, nlpI, and traV were less 0.7 (Table 1). Therefore, these signal peptides were deleted from the study, and subsequent analysis was performed by the remaining signal peptides (36 signal peptides) [14].

Investigation of different regions of the signal peptides

The results of these analyses demonstrated that all signal peptides had clear N, H, and C regions with a length of 3-6, 8-11, and 6-7 residues, respectively. Also, the results showed that all signal peptides except tortB, thiB, pspE, mepA, dsbD, ampC, cpdB, ygiW, ompL, ynfD, trbC, lolA, yfdX, ybcL, phoE, erfK, tolB, and hdeA had known cleavable sites with Alanine-X-Alanine (AXA) (Table 1). In this step, the signal peptides with unknown cleavable sites were deleted and subsequent analysis continued the remaining 18 signal peptides.

Physicochemical features of signal peptides

The results of the physicochemical analysis showed that the highest grand average of hydropathicity (GRAVY) belonged to csgA (1.725) and the lowest belonged to skp (0.860). Moreover, the range of the aliphatic index was between 78.50 (dsbC) and 171.90 (ompC). Also, the results revealed that pI was between 8.50 (nanM and eco) and 11 (sodC and ytfQ), while the instability index was between -7.48 (csgA) and 56.98 (ytfQ). It is essential to note that when the instability index of a protein is less than 40, it is considered a stable protein. On the contrary, when the instability index of a protein is more than 40, it is known as an unstable protein [14]. Based on this definition, the unstable signal peptides, including ytfJ (43.30), ytfQ (56.98), and ompG (53.39) were eliminated from the current study (Table 2). Hence, the project continued using the remained signal peptides (15 signal peptides).

Identification of signal peptide type and sub-cellular location

The results of signal peptide sorting demonstrated that all remaining signal peptides belonged to the secretory (Sec) pathway. Also, the results of the sub-cellular location showed that all remaining signal peptides except art I (secreted) and btuB (membrane) could translocate a nascent protein to the periplasmic compartment. Consequently, artI and btuB were filtered from the study, and other signal peptides were kept to introduce the best signal peptide (Table 2).

Journal of
Inflammatory Diseases
(Previous Title: The Journal of Qazvin University of Medical Sciences)

Table 1. Primary list of the signal peptides and investigation of different regions (N, H, and C) and cleavable sites of the signal peptides

List of the Signal Peptides					Evaluation of Different Regions (N, H, and C) and Cleavable Sites					
Row	Gene	Accession No.	Organism	Signal Sequence	Length	D-score	N Region	H Region	C Region	Cleavable Site
1	rlpB	P0ADC1	*Escherichia coli*	MRYLATLLLSLAVLITAG	18	0.56	-	-	-	-
2	slp	P37194	*Escherichia coli*	MNMTKGALILSLSFLLAA	18	0.54	-	-	-	-
3	thiB	P31550	*Escherichia coli*	MLKKCLPLLLLCTAPVFA	18	0.73	1-4 (4)	5-12 (8)	13-18 (6)	unknown
4	traV	P41069	*Escherichia coli*	MKQTSFFIPLLGTLLLYG	18	0.25	-	-	-	-
5	torT	P38683	*Escherichia coli*	MRVLLFLLLSLFMLPAFS	18	0.81	1-3 (3)	4-12 (9)	13-18 (6)	unknown
6	nlpI	P0AFB1	*Escherichia coli*	MKPFLRWCFVATALTLAG	18	0.40	-	-	-	-
7	artJ	P30860	*Escherichia coli*	MKKLVLAALLASFTF-GASA	19	0.82	1-3 (3)	4-12 (9)	13-19 (7)	ASA
8	sbp	P0AG78	*Escherichia coli*	MNKWGVGLTFLLAATS-VMA	19	0.67	-	-	-	-
9	dsbD	P36655	*Escherichia coli*	MAQRIFTLILLLCSTSVFA	19	0.80	1-4 (4)	5-13 (9)	14-19 (6)	unknown
10	dsbA	P0AEG4	*Escherichia coli*	MKKIWLALAGLVLAF-SASA	19	0.82	1-4 (4)	5-13 (9)	14-19 (6)	ASA
11	pspE	P23857	*Escherichia coli*	MFKKGLLALALVFSLPVFA	19	0.85	1-4 (4)	5-13 (9)	14-19 (6)	unknown
12	sodC	P0AGD1	*Escherichia coli*	MKRFSLAILALVVAT-GAQA	19	0.77	1-4 (4)	5-13 (9)	14-19 (6)	AQA
13	nanM	P39371	*Escherichia coli*	MNKTITALAIMMAS-FAANA	19	0.74	1-4 (4)	5-13 (9)	14-19 (6)	ANA
14	mepA	P0C0T5	*Escherichia coli*	MNKTAIALLALLASSASLA	19	0.84	1-4 (4)	5-13 (9)	14-19 (6)	unknown
15	cpdB	P08131	*Escherichia coli*	MIKFSATLLATLIAASVNA	19	0.71	1-4 (4)	5-13 (9)	14-19 (6)	unknown
16	ampC	P00811	*Escherichia coli*	MFKTTLCALLITASCSTFA	19	0.76	1-4 (4)	5-13 (9)	14-19 (6)	unknown
17	artI	P30859	*Escherichia coli*	MKKVLIAALIAGFSLSATA	19	0.82	1-3 (3)	4-12 (9)	13-19 (7)	ATA
18	ytfJ	P39187	*Escherichia coli*	MTLRKILALTCLLLPM-MASA	20	0.88	1-5 (5)	6-14 (9)	15-20 (6)	ASA
19	dsbC	P0AEG6	*Escherichia coli*	MKKGFMLFTLLAAFS-GFAQA	20	0.79	1-4 (4)	5-13 (9)	14-20 (7)	AQA
20	btuB	P06129	*Escherichia coli*	MIKKASLLTACSVTAFSA-WA	20	0.71	1-4 (4)	5-13 (9)	14-20 (7)	AWA
21	nlpE	P40710	*Escherichia coli*	MVKKAIVTAMAVIS-LFTLMG	20	0.55	-	-	-	-
22	ompL	P76773	*Escherichia coli*	MKKINAIILLSSLTSASVFA	20	0.77	1-5 (5)	6-14 (9)	15-20 (6)	unknown
23	surA	P0ABZ	*Escherichia coli*	MKNWKTLLLGIAMI-ANTSFA	20	0.67	-	-	-	-
24	yaeT	P0A940	*Escherichia coli*	MAMKKLLIASLLFS-SATVYG	20	0.63	-	-	-	-

List of the Signal Peptides					Evaluation of Different Regions (N, H, and C) and Cleavable Sites					
Row	Gene	Accession No.	Organism	Signal Sequence	Length	D-score	N Region	H Region	C Region	Cleavable Site
25	lpp	P69776	*Escherichia coli*	MKATKLVLGAVILGSTL-LAG	20	0.53	-	-	-	-
26	ygiW	P0ADU5	*Escherichia coli*	MKKFAAVIAVMALC-SAPVMA	20	0.85	1-4 (4)	5-14 (10)	15-20 (6)	unknown
27	csgA	P28307	*Escherichia coli*	MKLLKVAAIAAIVF-SGSALA	20	0.83	1-5 (5)	6-14 (9)	15-20 (6)	ALA
28	ompT	P09169	*Escherichia coli*	MRAKLLGIVLTTPIAISSFA	20	0.52	-	-	-	-
29	dacB	P24228	*Escherichia coli*	MRFSRFIIGLTSCIAFS-VQA	20	0.48	-	-	-	-
30	skp	P0AEU7	*Escherichia coli*	MKKWLLAAGLGLALAT-SAQA	20	0.82	1-4 (4)	5-14 (10)	15-20 (6)	AQA
31	pldA	P0A921	*Escherichia coli*	MRTLQGWLLPVFMLP-MAVYA	20	0.58	-	-	-	-
32	eco	P23827	*Escherichia coli*	MKTILPAVLFAAFATT-SAWA	20	0.79	1-4 (4)	5-14 (10)	15-20 (6)	AWA
33	yfdX	P76520	*Escherichia coli*	MKRLIMATMVTAILAS-STVWA	21	0.82	1-5 (5)	6-15 (10)	16-21 (6)	unknown
34	ytfQ	P39325	*Escherichia coli*	MWKRLLIVSAVSAAMS-SMALA	21	0.84	1-4 (4)	5-14 (10)	15-21 (7)	ALA
35	phoE	P02932	*Escherichia coli*	MKKSTLALVVMGIVA-SASVQA	21	0.77	1-5 (5)	6-15 (10)	16-21 (6)	unknown
36	ybcL	P77368	*Escherichia coli*	MKTLIVSTVLAFITF-SAQAAA	21	0.81	1-4 (4)	5-13 (9)	14-21 (8)	unknown
37	ompW	P0A915	*Escherichia coli*	MKKLTVAALAVT-TLLSGSAFA	21	0.85	1-5 (5)	6-15 (10)	16-21 (6)	AFA
38	asr	P36560	*Escherichia coli*	MKKVLALV-VAAAMGLSSAAFA	21	0.90	1-4 (4)	5-15 (11)	16-21 (6)	AFA
39	ompG	P76045	*Escherichia coli*	MKKLLPCTALVMCAG-MACAQA	21	0.74	1-4 (4)	5-14 (10)	15-21 (7)	AQA
40	ompA	P0A910	*Escherichia coli*	MKKTAIAIAVALAGFAT-VAQA	21	0.82	1-4 (4)	5-14 (10)	15-21 (7)	AQA
41	tolB	P0A855	*Escherichia coli*	MKQALRVAFGFLILWAS-VLHA	21	0.75	1-6 (6)	7-15 (9)	16-21 (6)	unknown
42	ompC	P06996	*Escherichia coli*	MKVKVLSLLVPALL-VAGAANA	21	0.85	1-4 (4)	5-14 (10)	15-21 (7)	ANA
43	phoA	P00634	*Escherichia coli*	MKQSTIALALLPLLFT-PVTKA	21	0.61	-	-	-	-
44	ompN	P77747	*Escherichia coli*	MKSKVLALLIPAL-LAAGAAHA	21	0.86	1-5 (5)	6-14 (9)	15-21 (7)	AHA
45	lolA	P61316	*Escherichia coli*	MKKIAITCALLSSLVASS-VWA	21	0.83	1-4 (4)	5-15 (11)	16-21 (6)	unknown
46	erfK	P39176	*Escherichia coli*	MRRVNILCSFALLFASHT-SLA	21	0.77	1-5 (5)	6-15 (10)	16-21 (6)	unknown
47	ynfD	P76172	*Escherichia coli*	MKLSTCCAALLLALAS-PAVLA	21	0.89	1-5 (5)	6-15 (10)	16-21 (6)	unknown
48	cusF	P77214	*Escherichia coli*	MKKALQVAMFSLFTVIG-FNAQ	21	0.60	-	-	-	-
49	hdeA	P0AES9	*Escherichia coli*	MKKVLGVILGGLLLLPV-VSNA	21	0.73	1-4 (4)	5-15 (11)	16-21 (6)	unknown
50	trbC	P18473	*Escherichia coli*	MKLSMKSLAALLMMLN-GAVMA	21	0.86	1-6 (6)	7-15 (9)	16-21 (6)	unknown

Signal peptides with a D-score less than 0.7 (36 signal peptides) were deleted from further evaluation of different regions (N, H, and C) and cleavable sites which have been shown with " - ".

Spring 2022. Vol 26. Num 1

Journal of Inflammatory Diseases
(Previous Title: The Journal of Qazvin University of Medical Sciences)

Table 2. Physicochemical features of and types and sub-cellular location of the signal peptides

Gene	Signal Sequence	Length	pI	Instability Index	Protein Classification	Aliphatic index	GRAVY	Type	Cytoplasmic	Membrane	Secreted	Periplas-mic	Final Loca-tion
artJ	MKKLVLAALLASFTFGASA	19	10	7.03	Stable	123.68	1.337	Sec	N	N	N	Y	Periplasmic
dsbA	MKKIWLALAGLVLAFSASA	19	10	11.50	Stable	144.21	1.416	Sec	N	N	N	Y	Periplasmic
artI	MKKVLIAALIAGFSLSATA	19	10	5.01	Stable	144.21	1.463	Sec	N	N	Y	N	Secreted
sodC	MKRFSLAILALVVATGAQA	19	11	27.83	Stable	138.95	1.274	Sec	N	N	N	Y	Periplasmic
nanM	MNKTITALAIMMASFAANA	19	8.50	26.92	Stable	93.16	1.000	Sec	N	N	N	Y	Periplasmic
ytfJ	MTLRKILALTCLLLPMMASA	20	9.50	43.30	Unstable	151.50	1.435	-	-	-	-	-	-
csgA	MKLLKVAAIAAIVFSGSALA	20	10	-7.48	Stable	156.50	1.725	Sec	N	N	N	Y	Periplasmic
skp	MKKWLLAAGLGLALATSAQA	20	10	15.67	Stable	127.50	0.860	Sec	N	N	N	Y	Periplasmic
dsbC	MKKGFMLFTLLAAFSGFAQA	20	10	5.25	Stable	78.50	1.000	Sec	N	N	N	Y	Periplasmic
eco	MKTILPAVLFAAFATTSAWA	20	8.50	30.88	Stable	103.00	1.265	Sec	N	N	N	Y	Periplasmic
btuB	MIKKASLLTACSVTAFSAWA	20	9.31	15.46	Stable	98.00	1.000	Sec	N	Y	N	N	Membran
asr	MKKVLALVVAAAMGLSSAAFA	21	10	14.65	Stable	130.48	1.590	Sec	N	N	N	Y	Periplasmic
ompN	MKSKVLALLIPALLAAGAAHA	21	10	5.20	Stable	158.57	1.352	Sec	N	N	N	Y	Periplasmic
ompW	MKKLTVAALAVTTLLSGSAFA	21	10	1.44	Stable	125.71	1.210	Sec	N	N	N	Y	Periplasmic
ompC	MKVKVLSLLVPALLVAGAANA	21	10	14.37	Stable	171.90	1.552	Sec	N	N	N	Y	Periplasmic
ytfQ	MWKRLLIVSAVSAAMSSMALA	21	11	56.98	Unstable	125.71	1.262	-	-	-	-	-	-
ompA	MKKTAIAIAVALAGFATVAQA	21	10	9.52	Stable	121.43	1.295	Sec	N	N	N	Y	Periplasmic
ompG	MKKLLPCTALVMCAGMACAQA	21	8.66	53.39	Unstable	93.33	1.133	Sec	-	-	-	-	-

Journal of
Inflammatory Diseases

GRAVY: Grand average of hydropathicity; N: No; Y: Yes; Sec: Secretory; Rows containing red information were removed from further sub-cellular location investigations because they were identified as unstable in physicochemical features.

Table 3. Ranking the best predicted signal peptides

Rank*	Gene	Signal Sequence	Location	Protein Classification	Cleavable Site	D-score
1	asr	MKKVLALVVAAAMGLSSAAFA	Periplasmic	Stable	AFA	0.90
2	ompN	MKSKVLALLIPALLAAGAAHA	Periplasmic	Stable	AHA	0.86
3	ompW	MKKLTVAALAVTTLLSGSAFA	Periplasmic	Stable	AFA	0.85
4	ompC	MKVKVLSLLVPALLVAGAANA	Periplasmic	Stable	ANA	0.85
5	csgA	MKLLKVAAIAAIVFSGSALA	Periplasmic	Stable	ALA	0.83
6	artJ	MKKLVLAALLASFTFGASA	Periplasmic	Stable	ASA	0.82
7	dsbA	MKKIWLALAGLVLAFSASA	Periplasmic	Stable	ASA	0.82
8	skp	MKKWLLAAGLGLALATSAQA	Periplasmic	Stable	AQA	0.82
9	ompA	MKKTAIAIAVALAGFATVAQA	Periplasmic	Stable	AQA	0.82
10	dsbC	MKKGFMLFTLLAAFSGFAQA	Periplasmic	Stable	AQA	0.79
11	eco	MKTILPAVLFAAFATTSAWA	Periplasmic	Stable	AWA	0.79
12	sodC	MKRFSLAILALVVATGAQA	Periplasmic	Stable	AQA	0.77
13	nanM	MNKTITALAIMMASFAANA	Periplasmic	Stable	ANA	0.74

Journal of
Inflammatory Diseases

*The signal peptides have been ranked based on their D-score; rank 1 has been devoted to the highest D-score.

Selection of the best signal peptide

To introduce the best-predicted signal peptide among the top 13 signal peptides, D-score was considered. These results showed that the asr peptide with D-score=90 can be an appropriate candidate for the expression of HBHA protein in *E. coli* (Table 3).

Secondary and tertiary structures

To investigate the secondary structure of the HPHA protein related to the asr peptide, the SOPMA server was used. These results revealed that this protein contained 88.18 % alpha helix, 1.82% extended strand, 0.45 % beta-turn, and 9.55 % random coil (Figure 1A). Also, the result of the I-TASSER server showed that the HBHA protein related to the asr peptide was well modeled with a C-sore of -2.97 (Figure 1B).

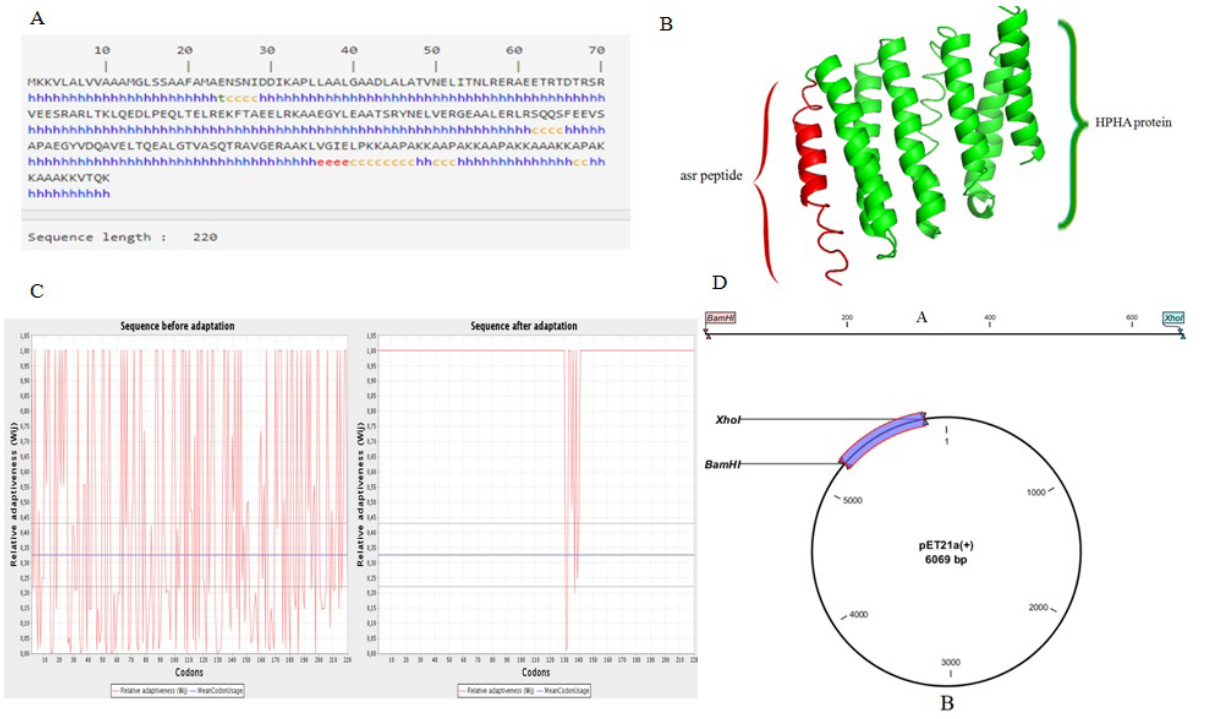
Codon optimization and in silico cloning

As mentioned, in this study JCat server was applied to perform codon optimization. The results of codon optimization showed that the Codon Adaptation Index (CAI) of HBHA protein associated with asr peptide before and after optimization were 0.18 and 0.94 respectively (Figure 1C). Moreover, the results of in silico cloning revealed that the nucleotide sequence of HPHA protein related to asr peptide with 660 base pairs in length could be successfully cloned in multiple cloning sites of pET21a (+) (Figure 1D).

4. Discussion

Nowadays, *E. coli* as a prokaryotic system is widely applied to express simple proteins worldwide. A Sec protein can be purified in both cytoplasmic and periplasmic compartments of *E. coli*. It has been reported that the cytoplasm of *E. coli* is more crowded than the periplasm, thus it is difficult to purify a protein from the cytoplasm. Also, in contrast to the periplasm, the cytoplasm is a reductive environment that can inhibit protein folding with disulfide bonds [15-17]. Therefore, a strategy should be considered to translocate a nascent protein to the periplasmic compartment. Applying an appropriate signal peptide can lead to the translocation of a Sec protein to the periplasmic compartment. Based on these important issues, the current study was conducted to find an appropriate signal peptide for the expression of HBHA protein in the periplasm environment of *E. coli*. To do this, 50 initial signal peptides extracted from *E. coli* were analyzed based on their D-scores. D-score is a parameter applied to distin-

Journal of
Inflammatory Diseases
(Previous Title: The Journal of Qazvin University of Medical Sciences)



Journal of
Inflammatory Diseases

Figure 1. Structure of Heparin-Binding Hemagglutinin (HBHA) protein

A: Secondary structure of Heparin-Binding Hemagglutinin (HBHA) protein related to asr peptide, different regions, including alpha helix, extended strand, beta-turn and random coil have been colored blue, red, green, and orange, respectively.

B: Tertiary structure of Heparin-Binding Hemagglutinin (HPHA) protein (white colore) related to asr peptide (red colore), the structure was visualized by PyMol software.

C: Codon optimization plots of Heparin-Binding Hemagglutinin (HBHA) protein related to asr peptide before (left-hand) and after (right hand) optimization.

D: In silico cloning of nucleotide sequence of Heparin-Binding Hemagglutinin (HBHA) protein related to asr peptide in pET21a (+) using CLC main workbench 5 software, (A) digestion product of bamhi and XhoI enzymes; (B) pET21a (+) which contains insert fragment.

guish between a signal and non-signal peptides [13], this parameter is usually considered with cutoff=0.57, but in the current study to increase the precision of the project, cutoff=0.7 was exerted. Using this threshold, 14 signal peptides were omitted and the remaining 36 signal peptides were used for further study. In the next step, the N, H, and C regions of the remaining signal peptides were evaluated. It has been confirmed that the N region of a signal peptide has a vital role in the interaction between the signal peptide and phospholipid membrane [10, 13, 18]. The results revealed that all signal peptides had clear N regions with a length of 3-6 residues. A direct relationship is observed between the length of the H region and the transmission rate of a signal peptide from the phospholipid membrane. A length of 9-12 residues has been reported to be acceptable for the H region [19]. As shown in Table 2, all 36 signal peptides had a clear H region with a length of 8-11 residues. One of the crucial features of an appropriate signal peptide is its cleavable site which contains three residues with an AXA box. In this pattern, a big amino acid, such as phenylalanine and tryptophan is embedded between two alanines [20].

As shown in Table 2, 18 signal peptides followed the AXA pattern. Therefore, other signal peptides which did not have the mentioned pattern were removed from further evaluations. In subsequent analysis, the physicochemical feature of the remaining 18 signal peptides was evaluated. The aliphatic index and GRVY are two parameters related to the hydrophobicity of a signal peptide [6]. As shown in Table 3, all remained signal peptides had an acceptable aliphatic index and GRAVY. The instability index is an index that is considered a valuable factor in determining the stability or instability of proteins. A protein with an instability index >40 is known as an unstable protein while a protein with an instability index <40 is recognized as a stable protein [12, 14, 21]. As reported in Table 3, three signal peptides, including ytfJ, ytfQ and ompG are unstable; then

they were excluded from the current study. In the next analysis, the type of remaining 15 signal peptides was determined, the results of this analysis confirmed that all remained 15 signal peptides belonged to the Sec pathway. Three ways, including twin-arginine translocation (TAT), Sec, and signal recognition particle (SRP) pathways have been detected for the translocation of a nascent protein to the periplasmic compartment. It has been reported that the Sec and signal recognition particle (SRP) pathways can translocate an unfolded protein to the periplasmic compartment, while the TAT pathway translocates a folded protein to this area [22, 23]. Moreover, sub-cellular location analysis confirmed that, except artI and btuB signal peptides, the remaining 13 signal peptides can transmit the nascent protein to the periplasmic compartment. Hence, artI and btuB were deleted and the rest were the candidate to choose the best signal peptide for the expression of HBHA protein in *E. coli*. Given that the final 13 signal peptides were able to successfully pass all analysis steps; they were ranked based on D-score as crucial features. The results of this ranking showed that the asr peptide was the best-predicted signal peptide to express HBHA protein in the periplasmic compartment of *E. coli*. After selecting the asr peptide as the best signal peptide, it was connected to the N-terminal of the HBHA protein and the secondary and tertiary structures of the protein were assessed. The result of the secondary structure revealed that the alpha helix (88.18%) was the main region of the HPHA protein associated with the asr peptide (Figure 1A).

The tertiary structure results showed that HPHA protein in the presence of the asr peptide can be successfully folded with C-score=-2.97 (Figure 1B). Finally, in silico cloning results confirmed that the nucleotide sequence of the HPHA protein associated with the asr peptide with 660 base pairs in length was successfully cloned in pET21a (+).

5. Conclusion

The results of the present study to identify the appropriate signal peptide for proper folding of HBHA protein confirmed that the asr signal peptide can be an appropriate candidate to express HBHA protein in the periplasmic compartment of *E. coli*. Therefore, by using this signal peptide to express HBHA protein, we can benefit from the biological properties of this protein (for example, using it as a molecular adjuvant in the design of recombinant vaccines).

Ethical Considerations

Compliance with ethical guidelines

There were no ethical considerations to be considered in this research.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Authors' contributions

Supervision and project administration: Ali Forouharmehr; Data mining: Ali Forouharmehr and Narges Nazifi; Collaborating on signal peptide analysis and protein structure evaluations: Ali Forouharmehr and Ehsan Rashidian; Original draft preparation, writing, review, and editing: Nemat Shams and Amin Jaydari.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

The author thanks the Faculty of Agriculture of Lorestan University for providing a calm environment for the research. The present project was done without funding support.

References

[1] Parra M, Pickett T, Delogu G, Dheenadhayalan V, Debrie A-S, Locht C, et al. The mycobacterial heparin-binding hemagglutinin is a protective antigen in the mouse aerosol challenge model of tuberculosis. Infect Immun. 2004; 72(12):6799-805. [DOI:10.1128/IAI.72.12.6799-6805.2004] [PMID] [PMCID]

[2] Menozzi FD, Rouse JH, Alavi M, Laude-Sharp M, Muller J, Bischoff R, et al. Identification of a heparin-binding hemagglutinin present in mycobacteria. J Exp Med. 1996; 184(3):993-1001. [DOI:10.1084/jem.184.3.993] [PMID] [PMCID]

[3] Mueller-Ortiz SL, Wanger AR, Norris SJ. Mycobacterial protein HbhA binds human complement component C3. Infect Immun. 2001; 69(12):7501-11. [DOI:10.1128/IAI.69.12.7501-7511.2001] [PMID] [PMCID]

[4] Pethe K, Alonso S, Biet F, Delogu G, Brennan MJ, Locht C, et al. The heparin-binding haemagglutinin of M. tuberculosis is required for extrapulmonary dissemination. Nature. 2001; 412(6843):190-4. [DOI:10.1038/35084083] [PMID]

Journal of
Inflammatory Diseases
(Previous Title: The Journal of Qazvin University of Medical Sciences)

[5] Jung ID, Jeong SK, Lee C-M, Noh KT, Heo DR, Shin YK, et al. Enhanced efficacy of therapeutic cancer vaccines produced by co-treatment with Mycobacterium tuberculosis heparin-binding hemagglutinin, a novel TLR4 agonist. Cancer Research. 2011; 71(8):2858-70. [DOI:10.1158/0008-5472.CAN-10-3487] [PMID]

[6] Forouharmehr A. Engineering an efficient poly-epitope vaccine against Toxoplasma gondii infection: A computational vaccinology study. Microb Pathog. 2021; 152:104646. [DOI:10.1016/j.micpath.2020.104646] [PMID]

[7] Rashidian E, Forouharmehr A, Nazifi N, Jaydari A, Shams N. Computer-Aided Design of a Novel Poly-Epitope Protein in Fusion with an Adjuvant as a Vaccine Candidate Against Leptospirosis. Current Proteomics. 2021; 18(2):113-23. [DOI:10.2174/1570164617666200319144331]

[8] Validi M, Karkhah A, Prajapati VK, Nouri HR. Immuno-informatics based approaches to design a novel multi epitope-based vaccine for immune response reinforcement against Leptospirosis. Mol Immunol. 2018;104:128-38. [DOI:10.1016/j.molimm.2018.11.005] [PMID]

[9] Jia B, Jeon CO. High-throughput recombinant protein expression in *Escherichia coli*: Current status and future perspectives. Open Biology. 2016; 6(8):160196. [DOI:10.1098/rsob.160196] [PMID] [PMCID]

[10] 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. Current Proteomics. 2018; 15(1):24-33. [DOI:10.2174/1570164614666170725144424]

[11] Idicula-Thomas S, Balaji PV. Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in *Escherichia coli*. Protein Sci. 2005; 14(3):582-92. [DOI:10.1110/ps.041009005] [PMID] [PMCID]

[12] Ghovvati S, Pezeshkian Z, Mirhoseini SZ. In silico analysis of different signal peptides to discover a panel of appropriate signal peptides for secretory production of Interferon-beta 1b in *Escherichia coli*. Acta Biochim Pol. 2018; 65(4):521-34. [DOI:10.18388/abp.2018_2351] [PMID]

[13] 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*. InInt J Pept Res Ther. 2015; 21(3):261-8. [DOI:10.1007/s10989-015-9454-z]

[14] Jaydari A, Forouharmehr A, Nazifi N. Determination of immunodominant scaffolds of Com1 and OmpH antigens of Coxiella burnetii. Microb Pathog. 2019; 126:298-309. [DOI:10.1016/j.micpath.2018.11.012] [PMID]

[15] Berndt C, Lillig CH, Holmgren A. Thioredoxins and glutaredoxins as facilitators of protein folding. Biochim Biophys Acta Mol Cell Res. 2008; 1783(4):641-50. [DOI:10.1016/j.bbamcr.2008.02.003] [PMID]

[16] De Marco A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. Microb Cell Factories. 2009; 8(1):1-18. [DOI:10.1186/1475-2859-8-26] [PMID] [PMCID]

[17] Messens J, Collet J-F. Pathways of disulfide bond formation in *Escherichia coli*. Int J Biochem Cell Biol. 2006; 38(7):1050-62. [DOI:10.1016/j.biocel.2005.12.011] [PMID]

[18] 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] [PMID]

[19] 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] [PMID] [PMCID]

[20] Payne SH, Bonissone S, Wu S, Brown RN, Ivankov DN, Frishman D, et al. Unexpected diversity of signal peptides in prokaryotes. MBio. 2012; 3(6):e00339-12. [DOI:10.1128/mBio.00339-12] [PMID] [PMCID]

[21] 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(3):989-95. [DOI:10.1007/s10989-018-9746-1]

[22] Pugsley AP, Schwartz M. Export and secretion of proteins by bacteria. FEMS Microbiol Rev. 1985; 1(1):3-38. [DOI:10.1111/j.1574-6968.1985.tb01180.x]

[23] Talmadge K, Gilbert W. Cellular location affects protein stability in *Escherichia coli*. PNAS. 1982; 79(6):1830-3. [DOI:10.1073/pnas.79.6.1830] [PMID] [PMCID]