Molecular Biology Research Communications 2020;9(1):1-10 DOI: 10.22099/mbrc.2019.35429.1457 MBRC
Original Article Open Access

Analyzing Signal Peptides for Secretory Production of Recombinant Diagnostic Antigen B8/1 from *Echinococcus granulosus*: An *In silico* Approach

Seyyed Hossein Khatami[1], Mortaza Taheri-Anganeh[2], Farzane Arianfar[3], Amir Savardashtaki[4, 5], Bahador Sarkari[6,7], Younes Ghasemi[5,8], Zohreh Mostafavi-Pour[1,9] *

1) Department of Biochemistry, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
2) Cellular and Molecular Research Center, Research Institute on Cellular and Molecular Medicine, Urmia University of Medical Sciences, Urmia, Iran
3) Department of Medical Biotechnology, Faculty of Allied Medicine, Iran University of Medical Sciences, Tehran, Iran
4) Department of Medical Biotechnology, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran
5) Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
6) Department of Parasitology and Mycology School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
7) Basic Sciences in Infectious Diseases Research Center, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
8) Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran
9) Maternal-Fetal Medicine Research Center, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran

ABSTRACT

Recombinant AgB8/1 as the most evaluated antigen for serological diagnosis of Cystic Echinococcosis (CE) can provide early and accurate diagnosis for proper management and treatment of the disease. Thus, the secretory production of this recombinant protein is the main goal and the application of signal peptides at the N terminus of the desired protein can help to achieve this goal. The present study applied few bioinformatics tools to evaluate several signal peptides to offer the best candidate for extracellular production of AgB8/1 of *Echinococcus granulosus* in *Escherichia coli*. The sequences related to signal peptides were obtained from "Signal Peptide Website" and were checked by "UniProt". In addition, UniProt was employed to retrieve the sequence of AgB8/1. Then, the probable signal peptide sequences and their cleavage site locations were determined by SignalP 4.1 followed by evaluation of their physicochemical features, using ProtParam. The solubility of the target recombinant proteins was accessed by SOLpro. Finally, PRED-TAT and ProtCompB were implemented to predict protein secretion pathways and final destinations. Among the 39 candidate signal peptides, ENTC2_STAAU and ENTC1_STAAU are the best ones which are stable and soluble in connection with AgB8/1 and can secrete target protein through Sec pathway. The signal

*Corresponding Author: Professor of Biochemistry, Department of Biochemistry, Shiraz University of Medical Sciences, Shiraz, IR Iran.
Tel: +98 713 2303029; Fax:+98 713 2303029
E. mail:zmostafavipour88@yahoo.co.uk

pISSN 2322-181X eISSN 2345-2005

peptides recommended in this investigation are valuable for rational designing of secretory stable and soluble AgB8/1. Such information is useful for future experimental production of the mentioned antigen.

Keywords: Antigen B; *Echinococcus granulosus*; Signal Peptide; *In Silico*

INTRODUCTION

Antigen B (AgB) is a major antigen of *Echinococcus granulosus* cyst fluid, which has been extensively evaluated for the diagnosis of Cystic Echinococcosis (CE) or hydatid cyst [1]. Hydatid cyst affects human health and welfare, consisting of both direct and indirect costs, calculated around 3 billion USD in the endemic areas [2]. On the other hand, since most of the patients at the early stages of the disease are asymptomatic, the physical imaging methods cannot be used for routine screening of CE infection. This calls for creation of an easy to use and cost-effective methods such as serological tests [3, 4]. A serological test based on AgB subunits can be used as an effective diagnostic tool for patient's follow-up after surgical or pharmacological treatment. AgB as a highly immunogenic antigen has shown high specificity and sensitivity in the serological diagnosis of CE. The antigen is a multimeric protein consisting of 8 kDa subunits, including 8-12, 16 and 24 kDa antigens [5-7]. It was confirmed that the 8 kDa subunit is the most suitable antigen for the serological diagnosis of CE. 8 kDa a subunit of AgB, called antigen B8/1, has exhibited the highest diagnostic sensitivity and specificity in comparison with other antigens that were applied for the serodiagnosis of CE [8]. In this regard, another investigation found specific antibodies against these two antigens, using western blotting [9]. The production of recombinant antigen B8/1 expressed in a heterologous system have several advantages such as ease of purification and reduction of cross-reactivity [10, 11]. The recombinant antigen could be produced in prokaryotic systems such as different strains of *E. coli*, recognized as the most common and alluring prokaryotic host to obtain recombinant proteins [12, 13].

E.coli as a desirable host for recombinant protein expression has several advantages including short generation time, an engineered genome and low-cost maintenance [14]. With the advent of recombinant DNA technology, increasing the solubility of a heterologous protein can be considered for large-scale bio-manufacturing, which can lead to serious problems such as misfolding and accumulation of protein that results in the formation of inclusion bodies. Inclusion body is a misfolded, insoluble aggregation of denatured proteins that reduce the yield of correctly folded proteins [15-17]. To overcome this difficulty, the target protein should be transferred to the oxidizing situation that exist in the sub-compartment of *E.coli* called periplasm [18]. Production of the secretory type of recombinant proteins using *E coli* offers a solution to this problem. Proteins transportation to the periplasm has some advantages, since the process of protein purification can be facilitated if the desired protein exist in this area that has fewer proteins, and its content can be selectively secreted by osmotic pressure or other approaches This approach provides several benefits including ease of purification, prevention of protease attack and N-terminal Met extension as well as having a more properly folded protein. An optimal secretion procedure consists of different stages depending on several factors.

Signal peptide is one of the most significant elements affecting different stages of secretion process and the yield of protein. Therefore, the bioinformatics assessment of the signal peptide sequence is recommended to determine the potential signal peptide A signal peptide is a short sequence consisting of 15-30 specific amino acids added to the N-terminus of proteins to permit their exportation to the outside of cytoplasm [17, 19-21]. Signal peptides structure consists of three parts including a positively-charged amino-terminal (n-region), a central hydrophobic core (h-region) and a polar carboxyl-terminal domain (c-region) [22]. The selection and connection of a signal peptide, appropriate to the target protein, is a critical step [23]. To this end, some

bioinformatics tools are available to predict the suitable signal peptides based on their specific characteristics [24].

To the best of our knowledge, there is no published data with respect to the secretory production of AgB8/1, using appropriate signal peptides. Consequently, in the present study an *in silico* approach assessing different signal peptides was used in order to suggest the best choice for secretory production of AgB8/1 in *E.coli*.

MATERIALS AND METHODS

Data Collection: The amino acid sequence related to AgB8/1 was obtained using UniProt server at http://www.uniprot.org/. The signal peptide sequences and their different parts including n, h and c-regions were achieved from Signal Peptide Website at http://www.signalpeptide.de/. All signal peptides were confirmed experimentally. The selected signal peptides were fused to AgB8/1 for further analysis. The total procedure of study is shown in Figure 1.

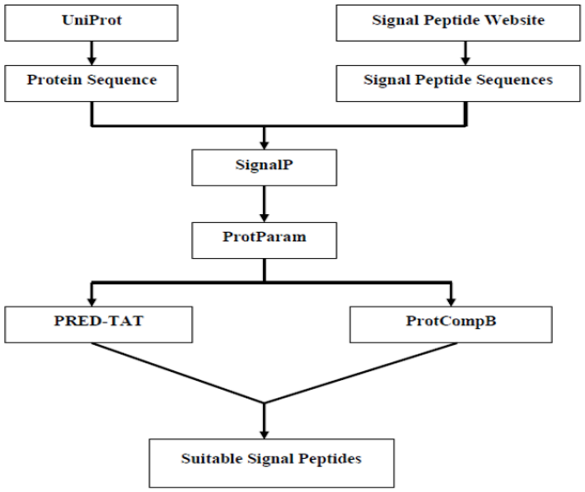

Figure 1: Flowchart of the procedure

Prediction of the existence of signal peptide and cleavage site position: Several bioinformatics tools are employed to predict the presence of signal sequences and location of their cleavage sites. The SignalP has the most accuracy and reliability among signal peptide identifying tools. Therefore, for the mentioned purposes the SignalP 4.1 web server (http://www.cbs.dtu.dk/services/SignalP/) which is based on a hidden Markov model (HMM) was applied [25]. SignalP predicts signal peptides probability for target protein and defines cleavage sites.

Evaluation of Physicochemical Parameters and Solubility: The ProtParam software related to the ExPASy server at http://web.expasy.org/protparam/ was implemented for assessment of physicochemical properties of the signal peptides, including aliphatic index, GRAVY (grand average of hydropathicity), instability index, positive charge and theoretical pI [26]. SOLpro, an online server at http://scratch.proteomics.ics.uci.edu/ was utilized to predict the solubility of the recombinant protein expressed in *E.coli*. The submitted protein sequences are categorized as soluble or insoluble determined by a probability score. This server employs a sequence-based technique for predicting protein solubility in *E. coli.* Finally, this server emails the results and their related probability score [27].

Prediction of Secretion Pathway and Sub-cellular location: "PRED-TAT" online server (http://www.compgen.org/tools/PRED-TAT) was employed to predict secretion pathway of B8/1 fused signal peptides. PRED-TAT defines secretion pathway based on Hidden Markov

Models (HMMs). The Sec pathway is the most desirable pathway for protein secretion in *E. coli*. Then, "ProtCompB" online server (http://www.softberry.com) was implemented to predict localization of signal peptide fused B8/1 dependent on neural networks. The most favorable result for secretion destination is "secreted" state [28, 29].

RESULTS

The AgB8/1 sequence was retrieved from UniProt (UniProt ID: Q2EN83). Next, the information related to 39 signal peptides of different organisms are shown in Table 1. The collected data consist of signal peptides separated from secretory proteins which are specific for eubacteria.-The signal peptide related scores (C, S, Y, S-mean and D), various regions of signal peptides consist of n, h and, c regions and their cleavage site positions are shown in Table 2.

The signal peptides named as TSH_ECOLX, PAG_BACAN, FIM1C_ECOLX, CEXE_ECOLX, FANH_ECOLX, HBP_ECOLX, PET_ECOLX, AIDA_ECOLX, PAPG_ECOLX and ESPP_ECOLX were reported with D-score values under cut-off; hence, they are not considered as appropriate signal peptides for the secretion of AgB8/1 and discarded for next steps of the study.

Table 1: Collected amino acid sequences dataset

No.	Accession	Signal Peptide	Source	Amino Acid Sequence
1	P0A910	OMPA_ECOLI	*E. coli (strain K12)*	**MKKT**AIAIAVALAGFATVAQA
2	P00634	PPB_ECOLI	*E. coli (strain K12)*	**MKQS**TIALALLPLLFTPVTKA
3	P06996	OMPC_ECOLI	*E. coli (strain K12)*	**MKVK**VLSLLVPALLVAGAANA
4	P09169	OMPT_ECOLI	*E. coli (strain K12)*	**MRAK**LLGIVLTTPIAISSFA
5	P02931	OMPF_ECOLI	*E. coli (strain K12)*	**MMKR**NILAVIVPALLVAGTANA
6	P0C1C1	PEL2_ERWCA	*Erwinia carotovora*	**MKY**LLPTAAAGLLLLAAQPAMA
7	P22542	HSTI_ECOLX	*Escherichia coli*	**MKK**NIAFLLASMFVFSIATNAYA
8	P02932	PHOE_ECOLI	*E. coli (strain K12)*	**MKKS**TLALVVMGIVASASVQA
9	P0AEX9	MALE_ECOLI	*E. coli (strain K12)*	**MKIKT**GARILALSALTTMMFSASALA
10	P69776	LPP_ECOLI	*E. coli (strain K12)*	**MKATK**LVLGAVILGSTLLAG
11	P02943	LAMB_ECOLI	*E. coli (strain K12)*	**MMITLRK**LPLAVAVAAGVMSAQAMA
12	P32890	ELBP_ECOLX	*Escherichia coli*	**MNKVK**CYVLFTALLSSLYAHG
13	P31746	CDGT_BACS2	*Bacillus sp. (strain 1-1)*	**MNDLNDFLKT**ILLSFIFFLLLSLPTVAEA
14	P0A618	MPT53_MYCTU	*Mycobacterium tuberculosis*	**MSLRLVSPIKAFADG**IVAVAIAVVLMFGLANTPRAVAA
15	Q50906	APA_MYCTU	*Mycobacterium tuberculosis*	**MHQVDPNLTRRKGR**LAALAIAAMASASLVTVAVPATANA
16	Q9XD84	TIBA_ECOLX	*Escherichia coli*	**MNKVYNTVWNESTGTWVVTSELTRKGGLRPRQIKRT**VLAGLI AGLLMPSM PALA
17	P06717	ELAP_ECOLX	*Escherichia coli*	**MK**NITFIFFILLASPLYA
18	Q8FDW4	SAT_ECOL6	*Escherichia coli O6*	**MNKIYSLKYSAATGGLIAVSELAKRVSGKTNRK**LVATMLSLAV AGTVNA
19	P06608	ASPG_ERWCH	*Erwinia chrysanthemi*	**MERWFK**SLFVLVLFFVFTASA
20	Q05044	SLAP_LACBR	*Lactobacillus brevis*	**MQSSLKK**SLYLGLAALSFAGVAAVSTTASA
21	P34071	ENTC2_STAAU	*Staphylococcus aureus*	**MNKSRFISC**VILIFALILVLFTPNVLA
22	Q47692	TSH_ECOLX	*Escherichia coli*	**MNRIYSLRYSAVARGFIAVSEFARKCVHKSVRR**LCFPVLLLIPVL FSAGS LA
23	P07965	HST3_ECOLX	*Escherichia coli*	**MKK**SILFIFLSVLSFSPFA
24	P39180	AG43_ECOLI	*E. coli (strain K12))*	**MKRHLNTCYRLVWNHMTGAFVVASELARARGKRGG**VAVALS LAAVTSLPV LA
25	P13811	ELBH_ECOLX	*Escherichia coli*	**MNKVKFY**VLFTALLSSLCAHG
26	P13423	PAG_BACAN	*Bacillus anthracis*	**MKKRK**VLIPLMALSTILVSSTGNLEVIQA
27	Q0Z8B6	HJM79_ENTHR	*Enterococcus hirae*	**MKKKVLKHC**VILGILGTCLAGIGTGIKVDA
28	P01553	ENTC1_STAAU	*Staphylococcus aureus*	**MNKSR**FISCVILIFALILVLFTPNVLA
29	P15917	LEF_BACAN	*Bacillus anthracis*	**MNIKKEFIK**VISMSCLVTAITLSGPVFIPLVQG
30	P24093	DRAA_ECOLX	*Escherichia coli*	**MKK**LAIMAAASMVFAVSSAHA
31	P62605	FIM1C_ECOLX	*Escherichia coli*	**MKLK**FISMAVFSALTLGVATNAS
32	A2TJI4	CEXE_ECOLX	*Escherichia coli*	**MKK**YILGVILAMGSLSAIA
33	P20862	FANH_ECOLX	*Escherichia coli*	**MIKK**VPVLLFFMASISITHA
34	O88093	HBP_ECOLX	*Escherichia coli*	**MNRIYSLR**YSAVARGFIAVSEFARKCVHKSVRRLCFPVLLLIPVLF SAGSLA
35	O68900	PET_ECOLX	*Escherichia coli*	**MNKIYSIK**YSAATGGLIAVSELAKKVICKTNRKISAALLSLAVISY TNIIYA
36	Q03155	AIDA_ECOLX	*Escherichia coli*	**MNK**AYSIIWSHSRQAWIVASELARGHGFVLAKNTLLVLAVVSTIG NAFA
37	P25394	FMF7_ECOLX	*Escherichia coli*	**MK**RLVFISFVALSMTAGSAMA
38	P13720	PAPG_ECOLX	*Escherichia coli*	**MKK**WFPAFLFLSLSGGNDALA
39	O32591	ESPP_ECOLX	*Escherichia coli*	**MNKIYSLK**YSHITGGLIAVSELSGRVSSRATGKKKHKRILALCFLG LLQSSYSFA

Table 2: *In silico* evaluation of signal peptides for AgB8/1

Protein Name	n-region	h-region	c-region	Cleavage Site	C-score	Y-score	S-score	S-mean	D-score
OMPA_ECOLI	1-4(4)	5-13(9)	14-21(8)	AQA	0.710	0.838	0.997	0.963	0.897
PPB_ECOLI	1-4(4)	5-16(12)	17-21(5)	TKA	0.386	0.594	0.992	0.928	0.751
OMPC_ECOLI	1-4(4)	5-16(12)	17-21(5)	ANA	0.754	0.865	0.997	0.976	0.917
OMPT_ECOLI	1-4(4)	5-12(8)	13-20(8)	AMT	0.436	0.590	0.994	0.887	0.730
OMPF_ECOLI	1-4(4)	5-17(13)	18-22(5)	ANA	0.776	0.877	0.996	0.969	0.920
PEL2_ERWCA	1-3(3)	4-18(15)	19-22(4)	AMA	0.835	0.910	0.996	0.966	0.936
HSTI_ECOLX	1-3(3)	4-18(15)	19-23(5)	AYA	0.803	0.892	0.998	0.972	0.930
PHOE_ECOLI	1-4(4)	5-15(11)	16-21(6)	VQA	0.656	0.805	0.998	0.950	0.873
MALE_ECOLI	1-5(5)	6-20(15)	21-26(6)	ALA	0.593	0.768	0.999	0.968	0.862
LPP_ECOLI	1-5(5)	6-16(11)	17-20(4)	TQA	0.452	0.592	0.996	0.891	0.733
LAMB_ECOLI	1-7(7)	8-16(9)	17-25(9)	AMA	0.691	0.826	0.998	0.965	0.891
ELBP_ECOLX	1-5(5)	6-15(10)	16-21(6)	AHG	0.731	0.849	0.998	0.945	0.894
CDGT_BACS2	1-10(10)	11-21(11)	22-29(8)	AEA	0.616	0.762	0.990	0.821	0.789
MPT53_MYCTU	1-15(15)	15-25(11)	25-38(14)	AVA	0.611	0.711	0.926	0.726	0.718
APA_MYCTU	1-14(14)	14-24(11)	25-39(15)	ANA	0.293	0.504	0.995	0.845	0.664
TIBA_ECOLX	1-36(36)	37-46(10)	47-54(8)	ALA	0.640	0.776	0.989	0.463	0.629
ELAP_ECOLX	1-2(2)	2-13(12)	14-18(5)	LYA	0.596	0.767	0.998	0.957	0.856
SAT_ECOL6	1-33(33)	34-44(11)	45-49(5)	VNA	0.567	0.713	0.960	0.474	0.601
ASPG_ERWCH	1-6(6)	7-17(11)	18-21(5)	ASA	0.746	0.860	0.999	0.964	0.909
SLAP_LACBR	1-7(7)	8-24(17)	24-30(7)	ASA	0.555	0.715	0.990	0.934	0.818
ENTC2_STAAU	1-9(9)	10-20(11)	21-27(7)	VLA	0.793	0.888	0.998	0.973	0.928
TSH_ECOLX	*				0.235	0.474	0.995	0.598	0.532
HST3_ECOLX	1-3(3)	4-13(10)	14-19(6)	PFA	0.526	0.723	0.998	0.970	0.839
AG43_ECOLI	1-35(35)	36-45(10)	46-52(7)	VLA	0.502	0.660	0.946	0.516	0.592
ELBH_ECOLX	1-7(7)	8-14(7)	15-21(7)	AHG	0.504	0.706	0.998	0.964	0.828
PAG_BACAN	*				0.310	0.424	0.934	0.700	0.554
HJM79_ENTHR	1-9(9)	10-20(11)	21-30(10)	CLA	0.322	0.509	0.957	0.815	0.653
ENTC1_STAAU	1-5(5)	6-21(16)	22-32(11)	VLA	0.793	0.888	0.998	0.973	0.928
LEF_BACAN	1-9(9)	10-22(13)	23-33(11)	TQA	0.324	0.459	0.860	0.705	0.575
DRAA_ECOLX	1-3(3)	4-16(13)	17-21(5)	AHA	0.482	0.622	0.891	0.827	0.698
FIM1C_ECOLX	*				0.258	0.295	0.867	0.541	0.386
CEXE_ECOLX	*				0.454	0.422	0.564	0.437	0.428
FANH_ECOLX	*				0.407	0.436	0.695	0.511	0.464
HBP_ECOLX	*				0.147	0.156	0.217	0.168	0.160
PET_ECOLX	*				0.158	0.161	0.349	0.259	0.197
AIDA_ECOLX	*				0.240	0.165	0.214	0.121	0.149
FMF7_ECOLX	1-2(2)	3-14(12)	15-21(7)	AMA	0.497	0.569	0.852	0.719	0.625
PAPG_ECOLX	*				0.335	0.372	0.762	0.504	0.421
ESPP_ECOLX	*				0.473	0.233	0.262	0.170	0.210

The physicochemical parameters are shown in Table 3. As expected, the results obtained from ProtParam showed that the net positive charges of all the target signal peptides were between +1 to +8, since these sequences were related to native signal peptides of *E. coli* or other living hosts.

Based on the results, OMPC_ECOLI, ENTC1_STAAU, ENTC2_STAAU, AGAR_ALTAT and, LPP_ECOLI had the highest aliphatic indexes. Additionally, the information indicated that ENTC1_STAAU, OMPC_ECOLI and ELAP_ECOLX had the highest GRAVYs, sequentially. The least instability index belonged to HJM79_ENTHR, PHOE_ECOLI, OMPT_ECOLI and MALE_ECOLI, respectively. PPB_ECOLI, OMPF_ECOLI, PEL2_ERWCA, LAMB_ECOLI, APA_MYCTU, TIBA_ECOLX, ELAP_ECOLX, ENTC2_STAAU, HST3_ECOLX, ENTC1_STAAU and, LEF_BACAN had instability index over 40, which meant that they are unstable; hence, were excluded from the study in the next step. Based on the results of SOLpro, the AgB8/1 connected to HSTI_ECOLX, ELBP_ECOLX, CDGT_BACS2, MPT53_MYCTU, ASPG_ERWCH, SLAP_LACBR, TSH_ECOLX, AGAR_ALTAT and, FMF7_ECOLX will play a role as an insoluble protein. Additionally, the AgB8/1 linked to MALE_ECOLI had the maximum solubility.

According to PRED-TAT results, all fused proteins can be secreted *via* Sec pathway, except APA_MYCTU, TIBA_ECOLX and AG43_ECOLI. Also, ProtCompB results showed that only PPB_ECOLI, APA_MYCTU, ENTC2_STAAU and ENTC1_STAAU could target soluble B8/1

out of the cytoplasm. Therefore, it can be predicted that TIBA_ECOLX and AG43_ECOLI direct the protein into transmembrane segments.

Table 3: Prediction of signal peptides physico-chemical properties and solubility

Signal Peptides	Amino Acid Length	Net Positive Charge	Aliphatic Index	GRAVY	Instability Index (alone)	Instability index (fused to B8/1)	Solubility
OMPA_ECOLI	21	2	121.43	1.295	Stable (9.52)	Stable (25.67)	Soluble (0.58)
PPB_ECOLI	21	2	139.52	0.971	Unstable (56.02)	Stable (35.24)	Soluble (0.60)
OMPC_ECOLI	21	2	171.90	1.552	Stable (14.37)	Stable (26.66)	Soluble (0.52)
OMPT_ECOLI	20	2	146.50	1.290	Stable (2.62)	Stable (24.46)	Soluble (0.55)
OMPF_ECOLI	22	2	150.91	1.259	Unstable (67.18)	Stable (37.83)	Soluble (0.55)
PEL2_ERWCA	22	1	138.18	1.191	Unstable (41.42)	Stable (32.32)	Soluble (0.56)
HSTI_ECOLX	23	2	102.17	1.026	Stable (32.43)	Stable (30.42)	Insoluble (0.75)
PHOE_ECOLI	21	2	130.00	1.195	Stable (1.44)	Stable (24.00)	Soluble (0.66)
MALE_ECOLI	26	3	113.08	1.012	Stable (2.85)	Stable (23.29)	Soluble (0.75)
LPP_ECOLI	20	2	161.00	1.400	Stable (10.64)	Stable (26.05)	Soluble (0.63)
LAMB_ECOLI	25	2	125.20	1.332	Unstable (42.97)	Stable (32.95)	Soluble (0.61)
ELBP_ECOLX	21	2	111.43	0.695	Stable (26.85)	Stable (29.24)	Soluble (0.54)
CDGT_BACS2	29	1	151.38	1.183	Stable (17.41)	Stable (26.57)	Insoluble (0.52)
MPT53_MYCTU	38	3	141.32	1.403	Stable (24.78)	Stable (28.23)	Insoluble (0.55)
APA_MYCTU	39	4	107.95	0.467	Unstable (42.02)	Stable (33.81)	Soluble (0.78)
TIBA_ECOLX	54	7	99.26	0.043	Unstable (47.89)	Stable (37.07)	Soluble (0.71)
ELAP_ECOLX	18	1	141.11	1.500	Unstable (88.98)	Stable (40.60)	Insoluble (0.65)
SAT_ECOL6	49	7	109.59	0.357	Stable (14.27)	Stable (23.98)	Soluble (0.50)
ASPG_ERWCH	21	2	106.67	1.352	Stable (29.64)	Stable (29.81)	Insoluble (0.76)
SLAP_LACBR	30	2	107.67	0.837	Stable (25.39)	Stable (28.65)	Insoluble (0.60)
ENTC2_STAAU	27	2	169.63	1.730	Unstable (49.08)	Stable (34.66)	Soluble (0.53)
HST3_ECOLX	19	2	123.16	1.416	Unstable (52.87)	Stable (34.23)	Insoluble (0.82)
AG43_ECOLI	52	7	108.85	0.465	Stable (26.67)	Stable (28.61)	Soluble (0.65)
AGAR_ALTAT	23	1	165.22	1.361	Stable (13.84)	Stable (26.31)	Insoluble (0.69)
ELBH_ECOLX	21	2	111.43	0.890	Stable (31.10)	Stable (30.11)	Soluble (0.54)
HJM79_ENTHR	30	5	139.67	0.890	Stable (-6.90)	Stable (19.92)	Soluble (0.64)
ENTC1_STAAU	27	2	169.63	1.730	Unstable (49.08)	Stable (34.66)	Soluble (0.53)
LEF_BACAN	33	3	132.73	1.042	Unstable (46.72)	Stable (34.73)	Insoluble (0.81)
DRAA_ECOLX	21	2	98.10	1.162	Stable (16.49)	Stable (27.10)	Soluble (0.76)
FMF7_ECOLX	21	2	102.38	1.290	Stable (29.55)	Stable (29.79)	Insoluble (0.63)

DISCUSSION

In our recent investigation we reported a successful expression of AgB8/1 in *E.coli* but easy purification can be accelerated by secretory production of a protein [30]. Since there are no suggested signal peptides for secretory production of AgB8/1 in *E.coli*; hence, in the present study several bioinformatics tools were used to suggest appropriate signal peptides to achieve secretory production of *Echinococcus granulosus* B8/1 antigen. For this purpose, 39 prokaryotic signal peptides were assessed computationally.

SignalP (version 4.1) was implemented to predict the presence of signal peptides and cleavage site locations. This server has two capabilities including differentiation between signal peptides and other sequences and also it can determine cleavage site locations. The SignalP uses an artificial neural network algorithm to calculate some scores, such as C, S, Y, and D score. The C-score (raw cleavage site score) plays a role in discriminating signal peptide cleavage sites from every other position. The S-score (signal peptide score) separates the signal peptide sequences from the mature area of proteins by defining the presence or absence of signal peptide. The Y-score (combined cleavage site score) is defined as a combination (geometric average) of the C-score and the slope of the S-score, which can lead to a better cleavage site estimation in comparison with the raw C-score alone. The Y-score is used to differentiate between C-score peaks through the selection of signal peptide where the slope of the S-score is sharp. The mean S is the average S-score, belonging to a possible signal peptide. D-score (discrimination score) is a weighted average of the mean S and the max Y scores, which can be

used to distinguish signal peptides from non-signal peptide sequences. The cut-off for all scores was set on 0.570 [25].

The combination of several signal peptides including OMPA_ECOLI, OMPC_ECOLI, OMPF_ECOLI, PEL2_ERWCA, HSTI_ECOLX, PHOE_ECOLI, MALE_ECOLI, LAMB_ECOLI, ELBP_ECOLX, ELAP_ECOLX, ASPG_ERWCH, SLAP_LACBR, ENTC2_STAAU, HST3_ECOLX, ELBH_ ECOLX, ENTC1_STAAU and AgB8/1 exhibited high D-scores using SignalP 4.0. Therefore, they can be regarded as appropriate signal peptides for AgB8/1.

The results of some other *in silico* investigations were in accordance with our results because they reported that the signal peptides called OMPC_ECOLI, OMPA_ECOLI, OMPF_ECOLI, PHOE_ECOLI, MALE_ECOLI and PEL2_ERWCA had the highest D-scores, too [6, 7].

The physico-chemical characteristics of the signal peptides play a significant role in the protein secretion. After evaluation by the SignalP server, the signal peptides that were reported with a D-score lower than the cut-off were discarded and the remaining were analyzed using ProtParam software to investigate their physico-chemical characteristics and stability. Proteins with the instability index<40 were regarded as stable and the instability index>40 meant that the signal peptide might be unstable [26]. Among the fusion proteins evaluated in this study, those connected to OMPT_ECOLI, MALE_ECOLI and HJM79_ENTHR were the most stable.

If the positive net charge of the n-region turns to zero or to a negative value, the transportation rate of the desired protein decreases significantly. These positive charges facilitate the interaction between signal peptide, the phospholipids and the translocation machinery located in the membrane. Therefore, the existence of one or more basic amino acids in the n-region permits the evolution of a useful signal peptide [19]. In this study, the net positive charge was 2 for most of the signal peptides. The CDGT_BACS2 with a net charge of -2 was the lowest one, and in contrast, AG43_ECOLI had the highest net positive charge of 7.

The reduction of hydrophobicity of the h-region has an inhibitory effect on the protein processing and translocation, which requires a minimal length and a minimum hydrophobic density of the h-region. Hence, disturbing this region by polar or charged amino acid residues can reduce or even completely terminate membrane transportation [31].

The hydrophobicity levels of the signal peptides were estimated by considering the aliphatic index and GRAVY (Table 3). The aliphatic index of a protein shows the relative volume filled by aliphatic side chains (alanine, valine, isoleucine, and leucine). The grand average of hydropathy (GRAVY) value for a peptide or protein is introduced as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. A lesser hydrophobicity results in a higher solubility [26]. Based on the obtained GRAVY and aliphatic index amount, some signal peptides including LPP_ECOLI, MPT53_MYCTU, ELAP_ECOLX, ENTC2_STAAU, AGAR_ALTAT and ENTC1_STAAU showed the highest hydrophobicity levels among all the remaining signal peptides.

The feature of cleavage efficiency has a great influence on the protein secretion level since the cleavage step is the rate-limiting factor in the protein secretion process. The determinative positions of C-region are considered as 1 and 3 prior to the cleavage site displayed as the (-3,-1) rule or AXA motif [19]. These positions are usually occupied by alanine, constructing the so-called Ala-X-Ala box, which is identified and cut by signal peptidase. Almost half of the signal peptides in this study had AXA motif in their cleavage sites as shown in Table 1.

The solubility of the passenger proteins identified by amino acid sequences can be considered as a key factor for secretion [32]. Therefore, the above mentioned stable signal peptides were accessed by SOLpro to define their solubility. SOLpro was used to predict the susceptibility of a protein to be soluble during overexpression in *E.coli*. The total accuracy of the SOLpro is 74.15% with a threshold of 0.5. SOLpro accurately labels 68.1% of the soluble proteins and 80.3% of the insoluble proteins [28]. In the midst of various signal peptides, HSTI_ECOLX, CDGT_BACS2, ASPG_ERWCH, SLAP_LACBR, and AGAR_ALTAT were supposed to construct insoluble proteins while it was reported that the target protein with

HSTI_ECOLX was insoluble, but the combination of the other four signal peptides (CDGT_BACS2, ASPG_ERWCH, SLAP_LACBR, and AGAR_ALTAT) and the desired protein was soluble [20]. On the other hand, the fusion of OMPA_ECOLI, OMPC_ECOLI, PHOE_ECOLI and MALE_ECOLI with our desired proteins were suggested to be soluble whereas they were considered as insoluble in Zamani *et al*., study [33].

As shown in Table 4, most signal peptides can secret B8/1 through Sec pathway. According to PRED-TAT, TIBA_ECOLX and AG43_ECOLI direct target protein into transmembrane section and the results were confirmed with ProtCompB. *E.coli* excretes 90% of its secretory proteins through Sec system, which can secret unfolded proteins while Tat system secrets fully folded proteins. Protein folding in cytoplasm is time-consuming and might result in protein accumulation and aggregation in cytoplasm. Hence, Sec pathway is more desirable to avoid inclusion bodies formation [34-36]. APA_MYCTU, ENTC2_STAAU and ENTC1_STAAU can secrete B8/1 to medium which APA_MYCTU can secrete through Tat pathway while ENTC2_STAAU and ENTC1_STAAU use Sec pathway to excrete the protein out of bacteria.

The variation in the results of the aforementioned investigations in comparison with ours might be due to differences in the targeted proteins, since the combination of different proteins with the same signal peptides can lead to the different solubility of the proteins.

Table 4: Secretion sorting and sub-cellular location of SPs

	Secretion pathway		Sub-cellular Localization				
Signal peptides	**Type of SP**	**Reliability Score (%)**	**Cytoplasmic**	**Membrane**	**Secreted (extracellular)**	**Periplasmic**	**Final prediction site**
OMPA_ECOLI	Sec	0.973	0.85	5.65	1.38	2.12	Inner Membrane
PPB_ECOLI	Sec	0.949	0.13	4.11	0.13	5.64	Periplasmic
OMPC_ECOLI	Sec	0.949	0.18	7.75	0.08	1.99	Inner Membrane
OMPT_ECOLI	Sec	0.913	0.21	9.79	0.00	0.00	Inner Membrane
OMPF_ECOLI	Sec	0.908	0.08	9.92	0.00	0.00	Inner Membrane
PEL2_ERWCA	Sec	0.988	0.35	3.79	0.00	5.86	membrane bound Periplasmic
PHOE_ECOLI	Sec	0.966	0.00	9.86	0.14	0.00	Inner Membrane
MALE_ECOLI	Sec	0.979	0.00	0.00	0.00	10.00	membrane bound Periplasmic
LPP_ECOLI	Sec	0.926	0.00	7.25	2.75	0.00	Inner Membrane
LAMB_ECOLI	Sec	0.974	0.00	6.65	0.02	3.33	Inner Membrane
ELBP_ECOLX	Sec	0.899	0.35	6.81	1.72	1.72	Inner Membrane
APA_MYCTU	Tat	0.829	0.00	0.42	9.53	0.05	Secreted
TIBA_ECOLX	TM segment	0.841	0.00	9.34	0.50	0.15	Inner Membrane
SAT_ECOL6	Sec	0.897	0.00	7.54	2.46	0.00	Inner Membrane
ENTC2_STAAU	Sec	0.973	0.00	2.25	7.73	0.02	Secreted
AG43_ECOLI	TM segment	0.928	0.00	8.39	1.38	0.22	Inner Membrane
ELBH_ECOLX	Sec	0.917	0.51	5.31	1.29	2.89	Inner Membrane
HJM79_ENTHR	Sec	0.926	0.00	10.00	0.00	0.00	Inner Membrane
ENTC1_STAAU	Sec	0.973	0.00	2.25	7.73	0.02	Secreted
DRAA_ECOLX	Sec	0.987	0.08	2.99	0.01	6.92	membrane bound Periplasmic

Acknowledgements: This study was financially supported by the office of vice-chancellor for research of Shiraz University of Medical Sciences (Grant No. 1396-01-01-15941). The results described in this paper were part of MSc student thesis of Seyyed Hossein Khatami. The authors wish to thank Mr.H.Argasi at the Research Consultation Center (RCC) of Shiraz University of Medical Sciences for his invaluable assistance in editing this manuscript.

Conflict of Interest: There is no conflict of interests.

REFERENCES

1. Siracusano A, Margutti P, Delunardo F, Profumo E, Riganò R, Buttari B, Teggi A, Ortona E. Molecular cross-talk in host-parasite relationships: the intriguing immunomodulatory role of Echinococcus antigen B in cystic echinococcosis. Int J Parasitol 2008;38:1371-1376.
2. Torgerson P. Economic effects of echinococcosis. Acta Trop 2003;85:113-118.
3. Vola A, Tamarozzi F, Noordin R, Yunus MH, Khanbabaie S, De Silvestri A, Brunetti E, Mariconti M. Preliminary assessment of the diagnostic performances of a new rapid diagnostic test for the serodiagnosis of human cystic echinococcosis. Diagn Microbiol Infect Dis 2018;92:31-33.
4. Sarkari B, SFEDAN AF, Moshfe A, Khabisi SA, Savardashtaki A, Hosseini F, Shahbazi A. Clinical and molecular evaluation of a case of giant primary splenic hydatid cyst: A case report. Iran J Parasitol 2016;11:585.
5. Mohammadzadeh T, Sako Y, Sadjjadi SM, Sarkari B, Ito A. Comparison of the usefulness of hydatid cyst fluid, native antigen B and recombinant antigen B8/1 for serological diagnosis of cystic echinococcosis. Trans R Soc Trop Med Hyg 2012;106:371-375.
6. Siracusano A, Delunardo F, Teggi A, Ortona E. Cystic echinococcosis: aspects of immune response, immunopathogenesis and immune evasion from the human host. Endocr Metab Immune Disord Drug Targets 2012;12:16-23.
7. Manzano-Román R, Sánchez-Ovejero C, Hernández-González A, Casulli A, Siles-Lucas M. Serological diagnosis and follow-up of human cystic echinococcosis: a new hope for the future? Biomed Res Int 2015;2015.
8. Sarkari B, Rezaei Z. Immunodiagnosis of human hydatid disease: where do we stand? World J Methodol 2015;5:185.
9. Rott MB, Fernández V, Farias S, Ceni J, Ferreira HB, Haag KL, Zaha A. Comparative analysis of two different subunits of antigen B from Echinococcus granulosus: gene sequences, expression in Escherichia coli and serological evaluation. Acta Trop 2000;75: 331-340.
10. Virginio VG, Hernandez A, Rott MB, Monteiro KM, Zandonai AF, Nieto A, Ferreira HB. A set of recombinant antigens from Echinococcus granulosus with potential for use in the immunodiagnosis of human cystic hydatid disease. Clin Exp Immunol 2003;132:309-315.
11. Taheri-Anganeh M, Khatami SH, Jamali Z, Movahedpour A, Ghasemi Y, Savardashtaki A, Mostafavi-Pour Z. LytU-SH3b fusion protein as a novel and efficient enzybiotic against methicillin-resistant Staphylococcus aureus. Mol Biol Res Commun 2019:151-158.
12. Siles-Lucas M, Casulli A, Conraths F, Müller N. Laboratory diagnosis of Echinococcus spp. in human patients and infected animals. Adv parasitol 2017;96:159-257.
13. Taheri-Anganeh M, Khatami SH, Jamali Z, Savardashtaki A, Ghasemi Y, Mostafavipour Z. In silico analysis of suitable signal peptides for secretion of a recombinant alcohol dehydrogenase with a key role in atorvastatin enzymatic synthesis. Mol Biol Res Commun 2019;8:17-26.
14. Bell MR, Engleka MJ, Malik A, Strickler JE. To fuse or not to fuse: what is your purpose? Protein Sci 2013;22:1466-1477.
15. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol 2014;5:172.
16. Kaur J, Kumar A, Kaur J. Strategies for optimization of heterologous protein expression in E. coli: Roadblocks and reinforcements. Int J Biol Macromol 2018;106:803-822.
17. Asadi M, Gharibi S, Khatami SH, Shabaninejad Z, Kargar F, Yousefi F, Taheri-Anganeh M, Savardashtaki A. Analysis of suitable signal peptides for designing a secretory thermostable cyanide degrading nitrilase: An in silico approach. J Environ Treat Tech 2019;7:506-513.
18. Zhong C, Wei P, Zhang YP. Enhancing functional expression of codon-optimized heterologous enzymes in Escherichia coli BL21 (DE3) by selective introduction of synonymous rare codons. Biotechnol Bioeng 2017;114:1054-1064.

19. Owji H, Nezafat N, Negahdaripour M, Hajiebrahimi A, Ghasemi Y. A comprehensive review of signal peptides: Structure, roles, and applications. Eur J Cell Biol 2018;97:422-441.
20. Mousavi P, Mostafavi-Pour Z, Morowvat MH, Nezafat N, Zamani M, Berenjian A, Ghasemi Y. In silico analysis of several signal peptides for the excretory production of reteplase in Escherichia coli. Curr Proteomics 2017;14:326-335.
21. Chen L. Bioinformatics analysis of protein secretion in plants. Plant Protein Secretion: Springer; 2017. p.33-43.
22. Freudl R. Signal peptides for recombinant protein secretion in bacterial expression systems. Microb Cell Fact 2018;17:52.
23. Brockmeier U, Caspers M, Freudl R, Jockwer A, Noll T, Eggert T. Systematic screening of all signal peptides from Bacillus subtilis: a powerful strategy in optimizing heterologous protein secretion in Gram-positive bacteria. J Mol Biol 2006;362:393-402.
24. Chang CC, Song J, Tey BT, Ramanan RN. Bioinformatics approaches for improved recombinant protein production in Escherichia coli: protein solubility prediction. Brief Bioinform 2014;15:953-962.
25. Nielsen H. Predicting secretory proteins with SignalP. Protein Function Prediction: Methods Mol Biol 2017;1611:59-73.
26. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Marc R, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. The proteomics protocols Handbook: Springer; 2005.p.571-607.
27. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. Bioinformatics 2009;25:2200-2207.
28. Bagos PG, Nikolaou EP, Liakopoulos TD, Tsirigos KD. Combined prediction of Tat and Sec signal peptides with hidden Markov models. Bioinformatics 2010;26:2811-2817.
29. Zeng R, Gao S, Xu L, Liu X, Dai F. Prediction of pathogenesis-related secreted proteins from Stemphylium lycopersici. BMC Microbiol 2018;18:191.
30. Savardashtaki A, Sarkari B, Arianfar F, Mostafavi-Pour Z. Immunodiagnostic value of Echinococcus granulosus recombinant B8/1 subunit of antigen B. Iran J Immunol 2017;14: 111-122.
31. Negahdaripour M, Nezafat N, Hajighahramani N, Soheil Rahmatabadi S, Hossein Morowvat M, Ghasemi Y. In silico study of different signal peptides for secretory production of interleukin-11 in Escherichia coli. Curr Proteomics 2017;14:112-121.
32. Jia B, Jeon CO. High-throughput recombinant protein expression in Escherichia coli: current status and future perspectives. Open Biol 2016;6:160196.
33. 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;21:261-268.
34. Reed B, Chen R. Biotechnological applications of bacterial protein secretion: from therapeutics to biofuel production. Res Microbiol 2013;164:675-682.
35. Rusch SL, Kendall DA. Interactions that drive Sec-dependent bacterial protein transport. Biochemistry 2007;46:9665-9673.
36. Green ER, Mecsas J. Bacterial secretion systems:an overview. Microbiol Spectr 2016;4.