A comprehensive in silico characterization of bacterial signal peptides for the excretory production of *Anabaena variabilis* phenylalanine ammonia lyase in *Escherichia coli*

Hajar Owji^{1,2} · Shiva Hemmati^{1,2}

Received: 3 June 2018 / Accepted: 13 November 2018 / Published online: 16 November 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Anabaena variabilis double mutant (C503S/C565S) phenylalanine ammonia-lyase (PAL) is an appealing enzyme in the enzyme-replacement therapy of phenylketonuria. Yet abundant production of this enzyme has been of concern for industrial production. In this study, we have characterized 1175 bacterial signal peptides (SPs) and identified the most efficient ones for the excretory production of mutant AvPAL. Analysis by SignalP 4.1 revealed that more than 61% of SPs had a *D*-score greater than 0.7, denoting to be highly efficient. The optimum length of a bacterial SP was 25–30. The preferable net positive charge and the second residue of N-region were +2 and Lys/Arg, respectively. Highly efficient SPs possessed 3–5 Leus in their H-region and A/L/VXA-FF cleavage site. Highly efficient SPs were from *Escherichia coli*, corroborating the necessity of an agreement between SPs and the host. Physicochemical characterization of mutant AvPAL conjugates via ProtParam and PROSOII, revealed that ~99.5% of proteins would not be entrapped in inclusion bodies. Secretory pathways were identified by EffectiveDB and more than 98% of SPs were cleavable. Chimeras were modeled using the I-TASSER program, being evaluated by the Ramachandran plots. The mRNA secondary structure of mutant AvPAL upon linkage to SPs was assessed using the mfold program. It was shown that the linkage of a SP does not affect mutant AvPAL's stability at the protein or mRNA level. AllergenFP tool demonstrated that chimeras were not allergen. SPs, including FMF4_ECOLX, E2BB_ECOLX, and LPTA_ECOLI exhibited the highest propensity for secretion and appropriate features in all analyses.

Keywords Phenylalanine ammonia lyase · Signal peptide · *Anabaena variabilis* · Phenylketonuria · Secretory

Abbreviations

Phe	Phenylalanine
PAL	Phenylalanine Ammonia Lyase
PKU	Phenylketonuria
<i>t</i> -CA	<i>Trans</i> -cinnamic acid
LNAA	Large neutral amino acid
SP	Signal peptide
HMM	Hidden Markov model

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-018-1517-3>) contains supplementary material, which is available to authorized users.

✉ Shiva Hemmati
hemmatish@sums.ac.ir

¹ Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran

² Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

Introduction

Phenylalanine ammonia-lyase (PAL, EC 4.3.1.24) is the first fundamental enzyme of the phenylpropanoid pathway, by which a number of important secondary metabolites are produced in higher plants. PAL is found ubiquitously in plants and abundantly in yeasts, particularly red yeast, such as *Rhodotorula* species. While animals lack this enzyme, a few numbers of bacterial species including *cyanobacteria* contain it (Aydaş et al. 2013). PAL is responsible for the oxidative deamination of L-phenylalanine (L-Phe) by producing *trans*-cinnamic acid (*t*-CA) and ammonia (NH₃). PAL has been of special interest as a therapeutic agent, and also in industrial and biotechnological processes (Sarkissian and Gámez 2005; Kong 2015). It has been considered as the most promising approach for the enzyme-replacement therapy of a metabolic disorder called phenylketonuria (PKU) (Longo et al. 2014; Blau and Longo 2015; Sumaily and Mujamammi 2017). PAL has also been considered as a potential anti-cancer agent in the same manner as L-asparaginase (Babich

et al. 2013; Shrivastava et al. 2016). It has been investigated as a biocatalyst in the production of enantiomerically pure α - or β - amino acids, such as L-Phe. These reactions are of significance in the food industry for provision of L-Phe as the key precursor of artificial sweetener aspartame or low-Phe foods (Turner 2011; Cui et al. 2012). PAL has demonstrated tyrosine ammonia lyase (TAL) activity in some plant species; thereby, it has been used in the quantitative analysis of L-Tyr as well as L-Phe in the body fluids (Watanabe et al. 1992). Moreover, it has been used to alter the plasma amino acid concentrations in sleep disorders in animal models (MacDonald and D'Cunha 2007; Cui et al. 2014).

Comprehensive analyses among PALs of different origins have introduced five PAL isoforms as viable options for commercial production and therapeutic purposes. These enzymes have been derived from *Petroselinum crispum*, *Rhodospiridium toruloides*, *Rhodotorula glutinis*, *Nostoc punctiforme*, and *Anabaena variabilis* (Sarkissian et al. 2008; Cui et al. 2014). The double mutant (C503S/C565S) PAL of *Anabaena variabilis* (AvPAL) is the preferred target due to less propensity toward aggregation in comparison to that of the wild-type (Wang et al. 2008; Bell et al. 2017). Nevertheless, up to now, the abundant commercial production of this enzyme has been the subject of concern.

To scale up the production of recombinant proteins, such as PAL, cytoplasmic expression in different eukaryotic and prokaryotic hosts, including *Escherichia coli* has become widely accepted. *E. coli* provides a number of advantages, namely being cheap and simple, having rapid cell growth, being genetically well characterized, and availability of a number of biotechnological tools for its manipulation (Fakruddin et al. 2012). However, there have been several hindrances in the large-scale production of recombinant proteins. They might not fold correctly due to the paucity of essential post-translational modifications or interference of the host endogenous proteins, leading to the formation of nonfunctional aggregates, called inclusion bodies. At times, the produced recombinant proteins get degraded by the host proteases or entrapped in the periplasmic compartment of Gram-negative bacteria (Jonet et al. 2012). Thereby, the protein yield will be reduced. On the other hand, the expression level is not the only determining factor in the protein's yield, but also the translocation efficiency plays a part. Henceforth, excretory production of recombinant proteins, the process of protein secretion into the extracellular medium, provides several advantages, including correct protein folding, increasing translocation efficiency, omission of the cell lysis step, getting rid of the endogenous host proteins, and removal of N-terminal methionine (Zhou et al. 2018).

The efficiency of excretory production depends on two main factors, namely the secretory machinery of the host cell and signal peptides (SPs). Several secretory pathways have been evolved in *E. coli*, from which one can take advantage

to determine protein localization along with the folding status of protein (Green and Meccas 2016). Moreover, SPs address the protein secretion pathway. SPs have been considered as promising tools by which the recombinant protein production has increased to commercially significant levels (Mergulhao et al. 2005; Low et al. 2013; Ohmuro-Matsuyama and Yamaji 2017). A typical signal peptide is a short peptide of 25–30 residues, mostly located in the N-terminal of secretory proteins. It is generally composed of three main parts: (i) N-region, the positive-charged domain; (ii) H-region, the hydrophobic core; and (iii) C-region, the cleavage site (von Heijne 1990). The positively charged N-region confers an apt orientation to SP favoring translocation (Nilsson et al. 2015). The hydrophobic H-region has been deemed as an indispensable part of SP for an efficient secretion (Hegde and Bernstein 2006). The Cleavage site harboring particular motifs, especially the AXA motif has been regarded as the rate-limiting determinant of protein secretion (Geukens et al. 2004). It has been contended that variations in the composition of SPs highly affect translocation efficiency, cleavage sites, and even the post-cleavage events (Chou 2002). Hence, the identification of an appropriate SP for each protein is of special importance.

One might plump for the most appropriate SP by delving into the literature, using the commercially available SPs or personal preference. Several methods have been implemented to identify the best performing SP, including computational methods, constructing a SP library, and screening proteome profile. The two latter methods are somewhat laborious, incurring high costs. Hence, employing computational methods, at least in initial stages, seems more rational (Heng 2009). In this study, we have collated a pool of reported bacterial SPs to identify the best ones for the excretory production of the mutant AvPAL in *E. coli*. Thereafter, we have implemented several analyses by means of the newest bioinformatics tools with high accuracy and precision over their counterparts. Our results significantly cut down the number of trial-and-error experiments, and provide a well-built basis for future experimental studies.

Materials and methods

Retrieval of SPs and the mutant AvPAL sequences

The primary structure of *Anabaena variabilis* PAL (AvPAL) (UniProt ID: Q3M5Z3) was retrieved from the UniProt server (<http://www.uniprot.org/>). Since the wild-type AvPAL has been shown to aggregate during production, the AvPAL C503S/C565S has been preferred in experiments as well as clinical trials (Bell et al. 2017). Henceforth, through this analysis, the secretion of mutant AvPAL via different SPs has been studied. The retrieved mutant AvPAL was overexpressed as a

PAL sequence through alignment against PALs of different origins, including *Petroselinum crispum* PAL (PcPAL uniprot ID: P24481), *Rhodospiridium toruloides* PAL (RtPAL uniprot ID: GOSVG1), *Rhodotorula glutinis* PAL (RgPAL uniprot ID: V5TFQ0), and *Nostoc punctiforme* PAL (NpPAL uniprot ID: B2J528) using the T-coffee program (<http://tcoffee.crg.cat>) (Notredame et al. 2000) and edited via Jalview (<http://www.jalview.org>) (Waterhouse et al. 2009). The relevant secondary structure was predicted by JNet available in Jalview server. Bacterial SPs, indicated as confirmed through experimental analyses, were retrieved from two databases, signalpeptide.de (<http://www.signalpeptide.de>) and SPdb (<http://proline.bic.nus.edu.sg/spdb/>) (Choo et al. 2005). To evaluate the secretory features of the engineered protein, retrieved SPs were linked with the whole sequence of mutant AvPAL.

Cleavage site prediction of SPs linked to the mutant AvAPL

The presence of cleavage sites and their locations in SPs were predicted using the SignalP 4.1 server (<http://www.cbs.dtu.dk/services/SignalP/>) (Nielsen 2017). This server calculates the probability of a particular protein sequence as SP using a discriminating score, called *D*-score, with a standard default cut-off of 0.57. *D*-scores above cut-off designate sequences as SPs and those higher than 0.7 represent a great possibility of being an efficient SP. The used settings were Gram-negative bacteria, default *D*-cutoff value of 0.57, standard graphics output, and the “Input sequences do not include TM regions” option. The sequences of the N-regions and H-regions of each SP were identified using the SignalP 3.0 server through Hidden Markov model (HMM) (Bendtsen et al. 2004). Other settings were Gram-negative bacteria and truncation to 70 residues.

Physiochemical properties investigation of the mutant AvAPL linked to SPs

The solubility of mutant AvPAL linked to different SPs upon expression in *E. coli* was predicted via PROSO II software (<http://mips.helmholtz-muenchen.de/prosoII>) (Smialowski et al. 2012). Other physiochemical properties of SPs linked to mutant AvPAL, including aliphatic index, GRAVY (grand average of hydropathicity), instability index, and amino acid composition were derived using ProtParam tool (<http://web.expasy.org/protparam/>) (Gasteiger et al. 2005).

Allergenicity evaluation of the mutant AvAPL linked to SPs

SPs that were linked to mutant AvPAL were assessed for allergenicity via the AllergenFP v.1.0 (<http://ddg-pharmfac.net/AllergenFP/>) (Dimitrov et al. 2013).

mRNA Secondary Structure prediction of SPs and the mutant AvAPL linked to pET-28a

The secondary structure and Gibbs free energy (ΔG) of pertinent mRNAs were predicted and calculated using the mfold online server at (<http://unafold.rna.albany.edu/?q=DINAMelt/Quickfold>) (Zuker 2003). In this analysis, pET-28a was selected since it has been regarded as the most promising vector in the cloning and expression of PAL enzyme (Gao et al. 2012). Thereafter, 35 nucleotides in the upstream of initiation codon in the vector were linked to the 36 initial nucleotides of SPs or the mutant AvPAL. Energy rule was set to RNA (3.0), and other default settings were applied. The retrieved mRNAs were compared upon the amount of ΔG and the location of initiation codon in the secondary structure of mRNA.

Secretory pathways prediction of the mutant AvAPL linked to SPs in *E. coli*

Different tools available at EffectiveDB (<http://www.effective-tors.org>) were implemented to identify Type-III and IV secretion proteins, as well as chaperone-binding sites and subcellular localization. EffectiveDB tools recognize eukaryotic-like protein domain and signal peptides in the primary sequence (Eichinger et al. 2015).

Prediction and evaluation of 3D structures of the mutant AvPAL linked to SPs

The mutant AvPALs linked to SPs, representing a *D*-score higher than 0.9 upon analysis by SignalP 4.1 tool, were the subject of conformational analysis. The FASTA sequences of mutant AvPALs linked to SPs and mutant AvPAL lacking SP were submitted to the I-TASSER (Iterative Threading ASSEMBLY Refinement) server (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>). The most confident PDB models derived from I-TASSER were uploaded to the PDBsum Generate EMBL-EBI (<https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html>) (Laskowski 2007). Thereafter, PDBsum (<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum>) generated a Ramachandran plot for each PDB model upon PROCHECK analysis (Laskowski et al. 2018). The Ramachandran plots were scanned for the most favored and the additional allowed conformational regions.

Result and discussion

Prediction of SPs and their cleavage sites

Identification of potential SPs for the excretory production of PAL

About 1175 non-redundant bacterial SPs, previously deposited in databases namely signalpeptide.de and SPdb were retrieved (Online Resource 1). Hereinafter, results of various analyses would be reported using the name of SPs in Online Resource 1. Retrieved SPs from databases had a length between 11 and 59; however, the downstream analysis excluded the long SPs. SignalP 4.1 server was implemented to identify whether SPs linked to the mutant AvPAL could act as a SP or not. The retrieved mutant AvPAL sequence was previously proved as a PAL sequence in multiple alignment analysis with other PAL

sequences. Two conserved sites as indicator of PALs as well as the relevant secondary structure are designated in Fig. 1. Schematic representations of the mutant AvPAL secretion as well as different parts of a SP are depicted in Fig. 2. The potential cleavage site of SPs in the mutant AvPAL-SP conjugate was identified by SignalP 4.1. Up to now, SignalP 4.1 with an accuracy of 87% has been considered as the most promising server for the identification of cleavage sites (Ivankov et al. 2013; Nielsen 2017). Its prediction is based on artificial neural network (ANN) and HMM. Four scores, namely *C*-, *Y*-, *S*-, and *D*-score were calculated for each entry. *C*-score (raw cleavage site score) distinguishes the possible cleavage sites from other sites. Residues located immediately after the cleavage site represent higher *C*-score. *S*-score (SP score) discriminates proteins with a SP from proteins lacking an SP. It also distinguishes sites as SP from other sites in the mature protein, not being an SP. It is reported as the maximum amount and the mean amount. *Y*-score (combined cleavage

Fig. 1 Multiple sequence alignment analysis of PALs. Sequences are aligned by T-coffee and edited by Jalview. The color scheme is an emulation of the default scheme used for alignments in ClustalX

program. Residues are shaded by degree of conservation. Secondary structure is predicted by JNet pred. Red bars and green arrows represent α-helix and β-sheet, respectively

Fig. 2 Schematic representation of the mutant AvPAL secretion

site score) employs both *C*- and *S*-score to identify the true unique cleavage site from multiple cleavage sites predicted by *C*-score. Finally, *D*-score (discrimination score) reports a weighted average of the mean *S*- and the max *Y*-scores by which SPs from non-SPs are discriminated. All scores in the non-secretory proteins are very low and close to the value of 0.1. In this study, 941 and 711 SPs had a *D*-score more than 0.5 and 0.7, respectively (Online Resource 2). SPs with a *D*-score higher than 0.7 are considered as highly efficient SPs. Hereinafter, the use of “highly efficient” phrase in this study refers to the pointed classification. Sequences with *D*-scores higher than 0.9 are shown in Table 1. Other analyses in this study have been carried out on the SPs representing a *D*-score greater than 0.5. Seven sequences out of 29 representing a *D*-score higher than 0.9, had promising features upon downstream analyses and can be introduced as highly efficient SPs for excretory production of the mutant AvPAL. These sequences are ALDC_BREBE, FLAD_CAUCR, FM12_KLEPN, FMF4_ECOLX, E2BB_ECOLX, LPTA_ECOLI, and CDGT1_PAEMA. Furthermore, several commercially available SPs or those indicated as efficient SPs in the previous experimental analyses of other recombinant proteins (Ivankov et al. 2013; Low et al. 2013) were also categorized as highly efficient SPs for the excretory production of mutant AvPAL. Some of the mentioned SPs are OMPE_ECOLI, OMPC_ECOLI, DSDA_VIBCH, SKP_SALTY, LAMP_SALTY, PEL2_ERWCT, OMPC_GLUDA, OMPA_THEMEA, SURA_ECOLI, OSMY_ECOLI, MALE_ENTAE, OSBA_ECOLI, PELB_ERWCH, PHOE_CITFR, TOLB_ECOLI, YEBE_ECOLI, and TORT_ECOLI. SPs, with a *D*-score higher than 0.7, had

an average length of 25 amino acids. However, SPs representing a *D*-score between 0.5 and 0.7 were longer and had an average length of 27 amino acids (Fig. 3). These results were consistent with the previous studies, stating a length of 25–30 amino acid for an optimum bacterial SP. In other words, long SPs were mainly efficient in the eukaryotic hosts rather than the prokaryotic ones and such SPs mostly remained uncleaved after protein translocation (Hiss et al. 2009).

Characterization of SP structure

The sequences of N-regions, H-regions, and the cleavage sites are shown in Table 2 and Online Resource 3. It has been claimed that a rise in the net positive charge of N-region results in the higher translocation efficiency as well as secretion (Puziss et al. 1992; Nesmeyanova et al. 1997; Ismail et al. 2011). However, at times, an augmentation of the net positive charge contributes to protein aggregation, jamming of secretion translocase, and activation of stress-related genes, thereby a reduction in secretion efficiency (Kvam et al. 2010; Joshi et al. 2012; Low et al. 2013). Comprehensive analyses of the SPs of various origins have revealed different values of positive charge depending on the expression host. Nonetheless, it has been asserted that at least a net positive charge of +1 is necessary for an efficient secretion (Brockmeier et al. 2006; Choo et al. 2008; Degering et al. 2010; Low et al. 2013). In this study, the net positive charge of N-regions was between 0 and 8 from which 18%, 17%, and 12% had a net positive charge of +3, +1, and +2, respectively. However, considering the net positive charge of the highly efficient SPs for secretion

Table 1 Results of SP's analyses using the signalP 4.1 server with a *D*-score higher than 0.9 (Continued in Online resource 2)

Signal peptide	<i>C</i> -score	<i>Y</i> -score	<i>S</i> -score	<i>S</i> -mean	<i>D</i> -score
GUNE_CLOTM	0.836	0.899	0.989	0.967	0.931
ASR_ECOLI	0.84	0.898	0.985	0.962	0.928
CWBA_BACSU	0.821	0.893	0.99	0.964	0.926
AMPC_ENTCL	0.824	0.89	0.973	0.961	0.924
FRD2_SHEFN	0.807	0.883	0.981	0.963	0.921
ETXH_STAAU	0.81	0.887	0.98	0.955	0.919
ALDC_BREBE	0.813	0.885	0.987	0.956	0.919
PSBO_SYNEN	0.791	0.873	0.995	0.97	0.918
XYNA_BACPU	0.817	0.883	0.989	0.957	0.918
FLAD_CAUCR	0.788	0.872	0.988	0.968	0.917
CYC3_DESVM	0.772	0.867	0.995	0.969	0.915
Q8VLR1_9ACTO	0.765	0.866	0.991	0.97	0.915
FM12_KLEPN	0.775	0.866	0.991	0.968	0.914
NOSD_PSEST	0.861	0.902	0.968	0.927	0.914
FBPA_NEIMA	0.798	0.872	0.989	0.957	0.912
FMF4_ECOLX	0.822	0.88	0.983	0.947	0.912
FBP_NEIGO	0.798	0.872	0.989	0.957	0.912
PSPE_ECOLI	0.818	0.881	0.966	0.945	0.911
E2BB_ECOLX	0.835	0.89	0.973	0.934	0.911
CBPM_STRAL	0.815	0.877	0.993	0.951	0.911
LACE_AGRRD	0.807	0.88	0.986	0.941	0.909
LPTA_ECOLI	0.779	0.863	0.991	0.958	0.907
CDGT1_PAEMA	0.767	0.859	0.988	0.961	0.907
GUNG_CLOTH	0.732	0.847	0.99	0.975	0.907
NCASE_PSEAE	0.758	0.855	0.982	0.962	0.905
Q93L40_CAUCR	0.775	0.858	0.988	0.956	0.904
Q8KKG1_CAUCR	0.775	0.858	0.988	0.956	0.904
GUNS_CLOTM	0.755	0.852	0.992	0.959	0.902
OMP32_DELAC	0.792	0.864	0.97	0.943	0.901

of the mutant AvPAL, the story was completely different. About 57% of SPs had a net charge of +2, denoting an apt charge for the SP's N-region of the mutant AvPAL would be +2 (Fig. 3). For instance, only one SP among those having a *D*-score greater than 0.9, called CDGT_PAEMA, had a high net charge of +4. Interestingly, this observation would be a

compelling reason to do further analysis regarding the effect of higher charge on secretion efficiency.

Apart from the positive charge of N-region, the amino acid composition of this region is of special importance. The presence of basic residues, particularly Lys or Arg as the second position has been considered essential to an efficient secretion (Puziss et al. 1992; Sletta et al. 2007; Futatsumori-Sugai and Tsumoto 2010). In this analysis, ~55% of the highly efficient SPs for mutant AvPAL secretion had Lys or Arg at their second positions. This finding suggests the necessity for incorporating any of these two residues to devise a highly efficient SP for mutant AvPAL secretion.

A hydrophobic H-region has been considered as the cornerstone of a SP, on which the rate and efficiency of secretion are depended (Hegde et al. 2006). Nonetheless, not all kinds of hydrophobic residues constitute an efficient SP. It has been suggested that the presence of at least 3–5 Leu amino acids highly augments translocation (Rusch et al. 1994; Chen et al. 1996). Herein, we scanned for the number of Leu amino acids in the studied SPs. About 61% of the studied SPs had 3–5 Leu amino acids in their H-region. On the other hand, H-region has been shown to be more than a set of hydrophobic residues, but also consists of unique motifs, called H-motifs. In other words, SPs with the same hydrophobicity are not functionally equal. H-motifs have been conserved among species (Hegde et al. 2006; Duffy et al. 2010). The *E. coli* H-motifs has been shown to be rich in Ala besides Leu (Palazzo et al. 2007; Duffy et al. 2010; Nilsson et al. 2015). Considering the aforementioned evidence, the majority of other H-regions lacking 3–5 Leu amino acids in this study also could be considered as efficient SPs for the excretory production of mutant AvPAL in *E. coli* since they were rich in Ala. For instance, FLAD_CAUCR, which represented a *D*-score of 0.917 for the mutant AvPAL secretion was rich in Ala instead of Leu.

Characterization of the cleavage sites of mutant AvPAL linked to SPs

Cleavage is considered as the rate-limiting step of protein secretion, which highly relies on the amino acids at position

Fig. 3 A comparison between the *D*-score value with length and positive charge **(a)** Distribution of SPs' length according to *D*-score **(b)** Distribution of SPs' charge according to *D*-score

Table 2 Characterization of the regions of SPs with a *D*-score higher than 0.9 (Continued in Online resource 3)

Signal peptide	Cleavage site	Length			N-region sequence	N-reg charge	H-region sequence
		N-reg	H-reg	C-reg			
GUNE_CLOTM	VVA-AS	6	12	7	MKKIVS	2	LVCVLVMLVSIL
ASR_ECOLI	AFA-FF	4	11	6	MKKV	2	LALVVAAAMGL
CWBA_BACSU	SFA-FF	6	12	7	MKSCCKQ	2	LIVCSLAAILL
AMPC_ENTCL	ALA-FF	4	10	6	MMRK	2	SLCCALLLGI
FRD2_SHEFN	AMA-FF	5	10	7	MKLKY	2	LVSAMALVVL
ETXH_STAAU	SYA-KA	6	10	6	MINKIK	2	ILFSFLALLL
ALDC_BREBE	AFA-AF	7	11	6	MKKNIIT	2	SITSLALVAGL
PSBO_SYNEN	AFA-FF	5	12	9	MKYRI	2	LMATLLAVCLGI
XYNA_BACPU	AHA-FF	8	12	7	MNLRKLRL	3	LFVMCIGLTLIL
FLAD_CAUCR	AFA-GT	5	12	6	MKAFL	1	FAAAATLVITAL
CYC3_DESVM	ALA-AF	4	12	6	MKKM	2	FLTGVLALAVAI
Q8VLR1_9ACTO	ALA-AG	6	11	7	MSRSRT	2	AMLAALTAAAG
FM12_KLEPN	ALA-FF	5	11	7	MKIKT	2	LAMIVVSALAL
NOSD_PSEST	AQA-FF	11	9	7	MFKAQATFSRY	2	SAAVSLLLL
FBPA_NEIMA	ALA-FF	7	9	6	MKTSIRY	2	TSIRYATPA
FMF4_ECOLX	VMA-FF	4	12	6	MKKT	2	LIALAVAASAAV
FBP_NEIGO	ALA-FF	7	9	6	MKTSIRY	2	ALLAAALTA
PSPE_ECOLI	VFA-FF	4	9	6	MFKK	2	GLLALALVF
E2BB_ECOLX	AHA-FF	8	9	6	MSFKKIHK	3	AFVIMAALV
CBPM_STRAL	AVA-AP	8	12	7	MRPRIRL	3	LLTALVGAGLAF
LACE_AGRRD	AFA-FF	10	12	6	MDYSRLKRS	2	VSAALTAALLC
LPTA_ECOLI	AFA-FF	11	10	6	MKFKTNKLNLN	3	LVLASSLLAA
CDGT1_PAEMA	AWA-FF	9	11	7	MKSRYKRLT	4	SLALSLSMALG
GUNG_CLOTH	CFA-QN	5	12	6	MKKAK	3	AIFSLVVALMVL
NCASE_PSEAE	ARA-FF	6	12	6	MSRSAF	1	TALLSCVLLAL
Q93L40_CAUCR	AHA-FF	5	10	6	MKRLI	2	LAAAASLLAL
Q8KKG1_CAUCR	AHA-FF	5	10	6	MKRLI	2	LAAAASLLAL
GUNS_CLOTM	AFA-FF	8	12	7	MVKSARKIS	3	ILLAVAMLVSIM
OMP32_DELAC	AMA-FF	4	9	6	MKKS	2	LIALAVLAA

−3 and −1 prior to the cleavage site (Fig. 2). AXA motif is the most prevalent motif at the mentioned site (Geukens et al. 2004). Indeed, amino acids larger than Ala, including Val, Leu, and Ile are allowed at position −3 (Choo et al. 2005). Regarding the abovementioned facts, Ala was highly conserved at position −1 of all the identified cleavage sites, except ~12% of them. The identified cleavage sites lacking Ala at position −1 represented low *D*-scores, denoting not being highly efficient SPs for the excretory production of mutant AvPAL. In contrast to position −1, the amino acid at position −3 was variable; though, the preference of Ala, Leu, and Val was observed. The region immediately after the cleavage site is called pro-region, which is a determinant factor in the cleavage efficiency, particularly at position +1 and +2. Neutral or acidic amino acids in the Pro-region are preferred (von Heijne 1990; Zalucki et al. 2017). In ~77% of the identified pro-regions, Phe as a neutral amino acid was present either at one of the positions or at both of them.

It was shown that the presence of acidic residues at the second position of the pro-region enhances secretion in *E. coli* (Zalucki 2017). In agreement with the recent finding, Glu or Asp was present in the second position of ~0.05% of efficient SPs through this analysis. Surprisingly, the cleavage sites harboring acidic residues at position +2 did not have the typical AXA motif. For instance, the consensus sites of AFF-SE and SFF-SE were found in the mentioned cleavage sites. Highly efficient SPs with the consensus site of AFF-SE in conjugate with the mutant AvPAL were DSBA_ECOLI, Q70UQ6_STRUB, NANH_VIBC3, ESTE_VIBMI, ESTE_VIBMI, HEP1_PEDHE, AZUP_PARN, CY553_DESVH, LYS4_ECOLX, OSMB_ECOLI, PRLB_ACHLY, LPPL_PSEAE, CY552_HYDTH, ARGT_ECOLI, TA34_TREPA, and Q6ZXZ8_WOLSU. In contrast to AFF-SE site, the consensus site of SFF-SE was less prevalent among highly efficient SPs and it was only present in one sequence, namely, OM24_PASMU. It can be deduced that the presence of an

acidic residue at position +2 might lead to an optimal interaction between the cleavage enzyme and SP, thereby increasing cleavage efficiency and secretion of mutant AvPAL. In other words, the acidic residues at position +2 might compensate the lack of Ala at position −1.

SP distribution according to the organism

Earlier studies assumed a high degree of degeneracy for SPs, which makes them viable candidates for increasing translocation efficiency in heterologous hosts (Rapoport et al. 1996; Kober et al. 2013). Nevertheless, this concept has become the subject of debate thereafter from several aspects. For instance, it has been suggested that there should be an optimal agreement between the SP and the passenger protein. In technical terms, not all combinations of SPs and a specific protein will result in an increase in the translocation efficiency, though the SP might be considered as a promising SP for other proteins. This issue is corroborated by several site-directed mutagenesis studies or exchanging SPs leading to the decimation of SP functionality (Pfeiffer et al. 2006; Weltman et al. 2007). In this study, in accordance with the abovementioned results, the SPs originated from *A. variabilis*, called PSBO_ANASP and PSAF_ANAVT, represented a *D*-score of 0.790 and 0.848, respectively. High *D*-scores for the AvSPs suggests a greater tendency of AvPAL secretion with consistent SPs. Moreover, SPs possess substrate-specific properties that make them suitable for particular translocation machinery. These substrate-specific properties have been evolutionary conserved and differ conspicuously among Gram-positive and Gram-negative bacteria or even a specific superfamily (Kim et al. 2002; Wang and Chi 2004). In our study, there was a great inclination toward secretion of the mutant AvPAL in *E. coli* by Gram-negative-originated SPs (Fig. 4). For instance, about 82% of SPs, which represented a *D*-score greater than 0.7, were originated from Gram-negative bacteria. To be specific, SPs, including FMF4_ECOLX, E2BB_ECOLX, and LPTA_ECOLI had a *D*-score higher than 0.9 and showed appropriate properties in downstream analyses for the secretion of mutant AvPAL. This observation delineates that SPs originated from *E. coli*

itself would be highly commensurate with the heterologous expression of mutant AvPAL in *E. coli*.

Investigation of the physicochemical properties of SPs linked to mutant AvPAL

Various physicochemical properties of the mutant AvPAL linked to SPs were calculated via ProtParam. The results are shown in Table 3 and Online Resource 4. The identified SPs ranged from 14 to 49 amino acids. The average molecular weight of the mutant AvPAL linked to SPs was 6 kDa and the average aromaticity was 0.07. All proteins had an instability index lower than 40, which denote all of them can be stable in the test tube. The pI ranged from 5.6 to 8.1 with an average of 6.4. The GRAVY value varied from −0.1 to 0.3 suggesting that ~99.5% of the studied proteins were soluble and only ~0.5% of proteins might form inclusion bodies. Proteins with negative GRAVY values were hydrophilic and soluble, whereas proteins with positive GRAVY values were hydrophobic. Since GRAVY has been considered as a determinant factor in predicting the expression efficacy (Chan et al. 2010), it seems that GRAVY would not be a hindrance upon secretory production of mutant AvPAL.

Prediction of the mutant AvPAL-SP solubility upon expression in *E. coli*

PROSO II online server was used to predict the solubility of mutant AvPAL linked to SPs. PROSO II is a sequence-based protein solubility evaluator, which recruits differences between soluble proteins from TargetDB, the PDB, as well as literature mining. The sequences are categorized as soluble or insoluble on the basis of the solubility score between 0 and 1 with a default threshold of 0.6 (Smialowski et al. 2012). PROSO II has the highest prediction accuracy percentage (71%) compared to several other similar servers, such as CCSOL (54.20%), SOLpro (59.95%), PROSO (57.85%), and recombinant protein solubility (51.4%) (Chang et al. 2013). Concerning the results of PROSO II, all mutant AvPAL linked to SPs were insoluble upon expression in *E. coli* with a solubility index ranging from 0.359 to

Fig. 4 A comparison between the *D*-score value with the organism distribution

(a) G-/G+ distribution of efficient SPs **(b) G-/G+ distribution of highly efficient SPs**

Table 3 Physiochemical characterization of SPs. (Continued in Online resource 4). Values are rounded up to two decimal places

Signal peptide	Molecular weight	Aromaticity	Instability index	Isoelectric Point	GRAVY
GUNE_CLOTM	65240.83	0.07	29.10	6.43	0.01
ASR_ECOLI	63739.94	0.07	29.04	6.31	- 0.03
CWBA_BACSU	64339.72	0.07	30.14	6.31	- 0.02
AMPC_ENTCL	63775.05	0.07	29.12	6.31	- 0.04
FRD2_SHEFN	63975.24	0.07	29.14	6.31	- 0.04
ETXH_STAAU	64423.75	0.07	29.45	6.43	- 0.04
ALDC_BREBE	64225.45	0.07	29.47	6.31	- 0.04
PSBO_SYNEN	64491.87	0.07	29.77	6.31	- 0.02
XYNA_BACPU	64699.19	0.07	30.00	6.46	- 0.03
FLAD_CAUCR	64555.83	0.07	29.29	6.20	- 0.01
CYC3_DESVM	64022.38	0.07	30.27	6.31	- 0.02
Q8VLR1_9ACTO	64553.78	0.08	30.96	6.31	- 0.04
FM12_KLEPN	63994.26	0.07	28.68	6.31	- 0.03
NOSD_PSEST	64540.69	0.07	28.67	6.31	- 0.05
FBPA_NEIMA	63909.05	0.07	29.42	6.31	- 0.05
FMF4_ECOLX	63780.95	0.07	29.03	6.3	- 0.04
FBP_NEIGO	63909.05	0.07	29.42	6.31	- 0.05
PSPE_ECOLI	63756.00	0.07	29.17	6.31	- 0.03
E2BB_ECOLX	64194.50	0.07	29.06	6.46	- 0.04
CBPM_STRAL	65736.15	0.07	28.84	6.31	- 0.02
LACE_AGRRD	64592.79	0.07	30.10	6.31	- 0.05
LPTA_ECOLI	64539.83	0.07	29.01	6.43	- 0.05
CDGT1_PAEMA	64655.97	0.07	29.87	6.57	- 0.07
GUNG_CLOTH	64865.28	0.07	28.29	6.43	- 0.03
NCASE_PSEAE	64213.49	0.07	32.06	6.31	- 0.04
Q93L40_CAUCR	63753.91	0.07	29.59	6.35	- 0.04
Q8KKG1_CAUCR	63753.91	0.07	29.59	6.35	- 0.04
GUNS_CLOTM	64596.07	0.07	29.07	6.43	- 0.02
OMP32_DELAC	63507.66	0.07	28.98	6.31	- 0.04

0.541 out of 1 (Online resource 5 and Table 4). The solubility of passenger proteins has been regarded as a fundamental criterion for secretion. Otherwise, the insoluble proteins will aggregate in the inclusion bodies (Chang et al. 2013). Although physiochemical characterizations announced all proteins as soluble, results derived from analysis by PROSO II were not promising. Hence, deeper analyses are necessary to approve whether insolubility is a hindrance in our study or not.

Prediction of the allergenicity of mutant AvPAL linked to SPs

The mutant AvPAL linked to SPs should be assessed for allergenicity if it is designed as a biotherapeutic. Allergenicity was predicted by means of the AllergenFP v.1.0. The AllergenFP v.1.0 recruits an alignment-free descriptor-based fingerprint approach, which identifies allergens and non-allergens on the basis of their physicochemical properties. Its accuracy has shown to be about 88%

(Dimitrov et al. 2013). Having a glance at the results reported by AllergenFP (Online Resource 5 and Table 4), it can be deduced that 84% of the studied mutant AvPAL linked to SPs would not be an allergen. For instance, FMF4_ECOLX, E2BB_ECOLX, and LPTA_ECOLI, with a *D*-score higher than 0.9 and appropriate physicochemical properties might not cause allergenicity in conjugation with the mutant AvPAL. Allergenicity has been of concern in the administration of intravenous or subcutaneous formulations of PAL (Ikeda et al. 2005). Hypersensitivity-type reactions were the most prevalent types of drug-related adverse effects after multiple dose prescription of the injectable PAL. At times, the allergic reactions have led to the patient dropout and discontinuation of therapy (Blau and Longo 2015). In the current study, the linkage of SPs to AvPAL did not affect allergenicity significantly. Thus, secretory production of AvPAL could be suggested as a promising method for commercial production of injectable formulations.

Table 4 Solubility scores, allergenicity, and the mRNA's DG. Predicted by PROSO II, AllergenFP v1.0, and mfold tools (Continued in Online Resource 5)

Signal peptide	Solubility score	Allergenicity	ΔG of mRNA
GUNE_CLOTM	0.523	PROBABLE NON-ALLERGEN	– 5.8
ASR_ECOLI	0.515	PROBABLE NON-ALLERGEN	– 14.5
CWBA_BACSU	0.524	PROBABLE NON-ALLERGEN	– 12.1
AMPC_ENTCL	0.517	PROBABLE ALLERGEN	– 9.7
FRD2_SHEFN	0.557	PROBABLE NON-ALLERGEN	– 11.9
ETXH_STAAU	0.493	PROBABLE NON-ALLERGEN	– 4.7
ALDC_BREBE	0.531	PROBABLE NON-ALLERGEN	– 3
PSBO_SYNEN	0.513	PROBABLE NON-ALLERGEN	– 8.4
XYNA_BACPU	0.519	PROBABLE NON-ALLERGEN	– 11.2
FLAD_CAUCR	0.52	PROBABLE NON-ALLERGEN	– 7.4
CYC3_DESVM	0.517	PROBABLE NON-ALLERGEN	– 10.7
Q8VLR1_9ACTO	0.528	PROBABLE NON-ALLERGEN	– 15.6
FM12_KLEPN	0.512	PROBABLE NON-ALLERGEN	– 6.1
NOSD_PSEST	0.507	PROBABLE NON-ALLERGEN	– 10.9
FBPA_NEIMA	0.511	PROBABLE ALLERGEN	– 11.5
FMF4_ECOLX	0.525	PROBABLE NON-ALLERGEN	– 5.7
FBP_NEIGO	0.511	PROBABLE ALLERGEN	– 11.5
PSPE_ECOLI	0.519	PROBABLE ALLERGEN	– 11.3
E2BB_ECOLX	0.548	PROBABLE NON-ALLERGEN	– 7.3
CBPM_STRAL	0.487	PROBABLE NON-ALLERGEN	– 7.9
LACE_AGRRD	0.508	PROBABLE NON-ALLERGEN	– 11.4
LPTA_ECOLI	0.545	PROBABLE NON-ALLERGEN	– 1.2
CDGT1_PAEMA	0.534	PROBABLE NON-ALLERGEN	– 5.4
GUNG_CLOTH	0.537	PROBABLE NON-ALLERGEN	– 12
NCASE_PSEAE	0.522	PROBABLE NON-ALLERGEN	– 20.7
Q93L40_CAUCR	0.521	PROBABLE NON-ALLERGEN	– 11.8
Q8KKG1_CAUCR	0.521	PROBABLE NON-ALLERGEN	– 11.8
GUNS_CLOTM	0.485	PROBABLE NON-ALLERGEN	– 8.4
OMP32_DELAC	0.536	PROBABLE NON-ALLERGEN	– 8.9

Prediction of mRNA secondary structure of SPs and the mutant AvAPL linked to pET-28a

Several factors other than SP works hand in hand to lead in an efficient secretion. For example, the fusion of SP to protein might affect the mRNA stability; therefore, different levels of mRNAs and protein precursors are produced. In this study, the ΔG and the secondary structure of pET28a linked to mRNA of SP-AvPAL or AvPAL were retrieved. The higher ΔG (the more positive ΔG) indicates the higher expression probability (Zhu et al. 2016). According to the results retrieved from the Quickfold server, addition of SP did not always lead to an increase in expression. About 12% of SP-AvPAL-pEt28a represented lower ΔG in comparison to that of the control (AvPAL-pET-28a). A mere 1.4% of the studied mRNA had a ΔG equal to the ΔG of AvPAL-pET28a construct, which was 7.6 kcal/mol. About 86.6% of SP-AvPAL-pET28a mRNA secondary structures showed higher ΔG in comparison to that of the

control, supporting the idea of an increase in expression upon attachment of a SP (Online resource 5 and Table 4). The mRNA structures with an exposed start codon (a start codon located on the loop instead of the stem) tend to be easily expressed (Zhang et al. 2006). The aforementioned sequences, representing high expression probability, were checked for the position of the start codon, AUG. In this analysis, ~55% of mRNAs out of 86.6% of sequences with high expression probability had an exposed start codon. Examples of a start codon, either on the loop or on the stem, are demonstrated in Fig. 5. For instance, the mRNA of mutant AvPAL conjugates with FMF4_ECOLX, E2BB_ECOLX, and LPTA_ECOLI, which represented promising features in upstream and downstream analyses, had higher ΔG in comparison to that of the AvPAL-pET28a construct, denoting higher inclination toward expression. Moreover, the initiation codon in all three mentioned constructs was located in the loop which can be easily accessed by translation machinery.

Fig. 5 Predicted secondary structure of the mRNAs. 36 bases upstream of the start codon and 35 bases downstream of the start codon were selected. The start codon AUG is labeled in a yellow arrow

Prediction of secretory pathways of the mutant AvPAL linked to SPs in *E. coli*

The general secretory pathway (Sec-pathway) and the twin-arginine pathway (Tat-pathway) have been the most prevalent pathways in *E. coli*; however, they cannot direct protein into the extracellular medium. Therefore, it behooves to manipulate recombinant proteins in a way to be secreted via pathways dedicated to excretory secretion, such as T2, T3, and T5 types of secretion systems (Green and Mecsas 2016). In this study, the secretion possibility of different SPs linked to the mutant AvPAL by secretory pathways, including T3 and T4 in *E. coli* was investigated. Additionally, it has been determined whether the studied SPs could be considered as an organellar-targeting sequence. These analyses were performed using different tools available at EffectiveDB. Proteins being secreted via the T3 pathway carry uncleavable SPs and remain in an unfolded state. The T3 pathway is of particular significance since it is responsible for secretion of virulence factors (Büttner 2012). Discrimination of organellar-targeting sequences from typical SPs leads in an efficient selection of SPs according to the host and the passenger protein. Organellar-targeting sequences and typical SPs differ in structural and secretory characteristics (Garg and Gould 2016). Through this analysis, scores derived from EffectiveT3 ranged between 0 and 1. Higher scores indicate a higher confidence in the prediction. Only 2% of

the studied proteins had a score more than 0.8 denoting the possibility of secretion via the T3 pathway (Online Resource 6). Among the highly efficient SPs directing mutant AvPAL through the T3 pathway, OMPC_NEIGO, CY552_BRAJA, BP26_BRUAB, WPRA_BACSU, LIP1_PHOLU, BLAN_ENTCL, SAV_STRAV, CYAA_BACAN, CGKA_ALTCA, MPPA_ECOLI, PMPF_CHLTR, EMP_STAAU, and ALXM_PHOS4 can be mentioned. Moreover, EffectiveCCBD predicts proteins secreted via T3 pathway on the basis of chaperone-binding sites detection. According to the results produced by EffectiveCCBD, only three proteins had chaperone-binding sites in their N-terminals, including EDIN_STAAU, SURA_ECOLI, and LIP1_PHOLU. None of the studied proteins were secreted via the T4 pathway based on the T4SE pre tool's results. Hence, one could proclaim with a great certitude that the majority of the studied SPs are secreted with the conventional secretory pathways, such as the Sec-pathway. Thus, SPs are cleaved subsequent to the translocation, denoting the secretory production of the mutant AvPAL can be a feasible method. We have screened N-terminal signaling sequences and eukaryotic subcellular localization via Predator tool. Although all of the analyzed N-terminal sequences were previously proved to be a SP by SignalP 4.1, Predator tool did not deem about 3% of them as SP. Furthermore, about 87% of the studied SPs were predicted to be similar to endoplasmic reticulum (ER) SPs and the rest were affiliated to the mitochondria.

Prediction and evaluation of the mutant AvPAL 3D structures linked to SPs

In this study, we have propounded the question whether the linkage of SPs to the mutant AvPAL affect the conformational characteristics of the protein. The PDB models for SP-AvPAL were produced via the I-TASSER server. Top 5 models were predicted for each entry. Models with the highest C-score for each entry were recruited for further analysis. Then, Ramachandran plots were generated using the PDBsum tool. The percentages of residues located in each region of the plot were reported. The model representing a percentage of greater 90% related to either the most favored regions or the most favored plus the additional allowed regions was regarded as a sterically robust model sharing more conformational similarities with the native protein (Sakhteman et al. 2016) (Online Resource 7, Fig. 6). In about 74.9–82.1% of all models, either with or without SP, 90% or fewer residues were in the most favored region. It is of special notice that the addition of SPs did not affect the aforementioned percentages. While the percentage of residues in the most favored regions for the mutant AvPAL lacking SP was 81.6%, there were SPs linked to the mutant AvPAL that represented a higher percentage, namely GUNG_CLOTH, OMP32_DELAC, and Q8VLR1_9ACTO.

Although in all generated models less than 90% of residues were in the most favored regions the sum of residues located in the most favored and the additional allowed regions was greater than 90% in all of them. Three of these SPs in conjugate with the mutant AvPAL, which were also promising in upstream analyses, were FMF4_ECOLX, E2BB_ECOLX, and LPTA_ECOLI. However, according to the statistics, all of the studied AvPAL mutants linked to a SP had somewhat lower percentage of the sum compared to that of the AvPAL mutant lacking SP. Altogether, the linkage of SPs to the mutant AvPAL did not reduce the protein folding properties to a significant level and all of the studied proteins were considered as sterically robust models, thereby stable upon linkage of a SP.

The lower stability of mutant AvPAL linked to SPs in comparison to the protein lacking SP could be attributed to the potential interaction of the hydrophobic regions of the SP and the exposed hydrophobic regions of the protein. For instance, in one study the maltose binding protein (MBP) was destabilized in the presence of SP. In the mentioned study, the SP had enhanced the probability of remaining in an unfolded state in the pre-MBP (the MBP linked to SP) (Beena et al. 2004). This phenomenon on one side could be detrimental since it might affect the interaction of protein with translocation machinery, thereby

Fig. 6 Predicted 3D structures and Ramachandran plots of proteins. **a** Mutant AvPAL, **b** Mutant AvPAL-OMP32_DELAC. In the Ramachandran plot the red regions (A, B, and L) refer to the most favored regions; the brown regions (a, b, l, and p) refer to the additional allowed regions; the yellow regions (~a, ~b, ~l, and ~p) refer to the generously allowed regions. The three dihedral angles in a protein chain are represented as phi and psi degrees, shown in the horizontal and vertical axis, respectively

reducing the protein secretion. On the other side, it can be beneficial since the less folded protein might be less prone to aggregation. Several studies have disclosed the fact that a SP reduces or even increases the propensity of aggregation by modulating the folding state of the protein (Kulothungan et al. 2009; Singh et al. 2013). Henceforth, kinetic studies are necessary to scrutinize the effect of SPs on folding state and stability of AvPAL mutant, though the in silico results, provided in this study, were promising.

Conclusion

Consistent results, derived from various bioinformatics analyses in this study have introduced the potential highly efficient SPs for the heterologous expression of the mutant AvPAL. Furthermore, the characterization of a pool of SPs has explained the necessary criteria in devising an efficient SP for the mutant AvPAL. Physiochemical characterizations, as well as identification of the secretory pathways revealed that the mutant AvPAL linked to different SPs is liable to secrete. Allergenicity characterization ascertained that allergenicity would not be an issue of concern in the injectable formulation of the mutant AvPAL. Furthermore, it was shown that the linkage of a SP to the mutant AvPAL does affect the stability neither at the mRNA nor at the protein level to a great fault. Hence, SPs might not interfere with the expression or the folding process of the mutant AvPAL. Thus, secretory production of mutant AvPAL would be an encouraging solution to overcome the obstacles through commercial production of PAL. In a nutshell, SPs, including FMF4_ECOLX, E2BB_ECOLX, and LPTA_ECOLI can be introduced as the best SPs for the excretory production of mutant AvPAL since they had a *D*-score higher than 0.9 and appropriate features in all analyses. Among the most reviewed signal peptides in literature, SPs indigenous to *E. coli* such as OMPE_ECOLI, OMPC_ECOLI, SURA_ECOLI, OSMY_ECOLI, OSBA_ECOLI, TOLB_ECOLI, YEBE_ECOLI, and TORT_ECOLI can also be utilized as they had high *D*-scores and promising characteristics in downstream analyses. To the best of our knowledge, hitherto, there has been no study concerning the excretory production of PAL. However, further investigations especially in vitro and in vivo studies are required to elucidate the superiority of the introduced SPs in comparison to that of the conventional ones in depth.

Acknowledgements Authors would like to thank Shiraz University of Medical Sciences, Shiraz, IRAN for the grant number: 1396-01-36-16684.

Compliance with ethical standards

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

References

- Aydaş SB, Ozturk S, Aslım B (2013) Phenylalanine ammonia lyase (PAL) enzyme activity and antioxidant properties of some cyanobacteria isolates. *Food Chem* 136(1):164–169. <https://doi.org/10.1016/j.foodchem.2012.07.119>
- Babich OO, Pokrovsky VS, Anisimova NY, Sokolov NN, Prosekov AY (2013) Recombinant l-phenylalanine ammonia lyase from *Rhodospiridium toruloides* as a potential anticancer agent. *Biotechnol Appl Biochem* 60(3):316–322. <https://doi.org/10.1002/bab.1089>
- Beena K, Udgaonkar JB, Varadarajan R (2004) Effect of signal peptide on the stability and folding kinetics of maltose binding protein. *Biochemistry* 43(12):3608–3619. <https://doi.org/10.1021/bi0360509>
- Bell SM, Wendt DJ, Zhang Y, Taylor TW, Long S, Tsuruda L, Zhao B, Laipis P, Fitzpatrick PA (2017) Formulation and PEGylation optimization of the therapeutic PEGylated phenylalanine ammonia lyase for the treatment of phenylketonuria. *PloS one* 12(3):e0173269. <https://doi.org/10.1371/journal.pone.0173269>
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340(4):783–795. <https://doi.org/10.1016/j.jmb.2004.05.028>
- Blau N, Longo N (2015) Alternative therapies to address the unmet medical needs of patients with phenylketonuria. *Expert Opin Pharmacother* 6(6):791–800. <https://doi.org/10.1517/14656566.2015.1013030>
- Brockmeier U, Caspers M, Freudl R, Jockwer A, Noll T, Eggert T (2006) Systematic screening of all signal peptides from *Bacillus subtilis*: a powerful strategy in optimizing heterologous protein secretion in Gram-positive bacteria. *J Mol Biol* 362(3):393–402. <https://doi.org/10.1016/j.jmb.2006.07.034>
- Büttner D (2012) Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiol Mol Biol Rev* 76(2):262–310. <https://doi.org/10.1128/MMBR.05017-11>
- Chan W-C, Liang P-H, Shih Y-P, Yang U-C, Lin W-c, Hsu C-N (2010) Learning to predict expression efficacy of vectors in recombinant protein production. *BMC Bioinform* 11(1):S21. <https://doi.org/10.1186/1471-2105-11-S1-S21>
- Chang CCH, Song J, Tey BT, Ramanan RN (2013) Bioinformatics approaches for improved recombinant protein production in *Escherichia coli*: protein solubility prediction. *Brief Bioinform* 15(6):953–962. <https://doi.org/10.1093/bib/bbt057>
- Chen H, Kim J, Kendall DA (1996) Competition between functional signal peptides demonstrates variation in affinity for the secretion pathway. *J Bacteriol* 178(23):6658–6664. <https://doi.org/10.1128/jb.178.23.6658-6664.1996>
- Choo KH, Ranganathan S (2008) Flanking signal and mature peptide residues influence signal peptide cleavage. *BMC Bioinform* 9(12):S15. <https://doi.org/10.1186/1471-2105-9-S12-S15>
- Choo KH, Tan TW, Ranganathan S (2005) SPdb—a signal peptide database. *BMC Bioinform* 6(1):249. <https://doi.org/10.1186/1471-2105-6-249>
- Chou K-C (2002) Prediction of protein signal sequences. *Curr Prot Pept Sci* 3(6):615–622. <https://doi.org/10.2174/1389203023380468>
- Cui J-D, Zhang S, Sun L-M (2012) Cross-linked enzyme aggregates of phenylalanine ammonia lyase: novel biocatalysts for synthesis

- of L-phenylalanine. *Appl Biochem Biotechnol* 167(4):835–844. <https://doi.org/10.1007/s12010-012-9738-0>
- Cui JD, Qiu JQ, Fan XW, Jia SR, Tan ZL (2014) Biotechnological production and applications of microbial phenylalanine ammonia lyase: a recent review. *Crit Rev Biotechnol* 34(3):258–268. <https://doi.org/10.3109/07388551.2013.791660>
- Degering C, Eggert T, Puls M, Bongaerts J, Evers S, Maurer K-H, Jaeger K-E (2010) Optimization of protease secretion in *Bacillus subtilis* and *Bacillus licheniformis* by screening of homologous and heterologous signal peptides. *Appl Environ Microbiol* 76(19):6370–6376. <https://doi.org/10.1128/AEM.01146-10>
- Dimitrov I, Naneva L, Doytchinova I, Bangov I (2013) AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* 30(6):846–851. <https://doi.org/10.1093/bioinformatics/btt619>
- Duffy J, Patham B, Mensa-Wilmot K (2010) Discovery of functional motifs in h-regions of trypanosome signal sequences. *Biochem J* 426(2):135–145. <https://doi.org/10.1042/BJ20091277>
- Eichinger V, Nussbaumer T, Platzer A, Jehl M-A, Arnold R, Rattei T (2015) EffectiveDB—updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Res* 44(D1):D669–D674. <https://doi.org/10.1093/nar/gkv1269>
- Fakruddin M, Mohammad Mazumdar R, Bin Mannan KS, Chowdhury A, Hossain MN (2012) Critical factors affecting the success of cloning, expression, and mass production of enzymes by recombinant *E. coli*. *ISRN Biotechnol*. <https://doi.org/10.5402/2013/590587>
- Futatsumori-Sugai M, Tsumoto K (2010) Signal peptide design for improving recombinant protein secretion in the baculovirus expression vector system. *Biochem Biophys Res Commun* 391(1):931–935. <https://doi.org/10.1016/j.bbrc.2009.11.167>
- Gao J, Zhang S, Cai F, Zheng X, Lin N, Qin X, Ou Y, Gu X, Zhu X, Xu Y (2012) Characterization, and expression profile of a phenylalanine ammonia lyase gene from *Jatropha curcas* L. *Mol Biol Rep* 39(4):3443–3452. <https://doi.org/10.1007/s11033-011-1116-4>
- Garg SG, Gould SB (2016) The role of charge in protein targeting evolution. *Trends Cell Biol* 26(12):894–905. <https://doi.org/10.1016/j.tcb.2016.07.001>
- Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook. Springer, New York 571–607
- Geukens N, Frederix F, Reekmans G, Lammertyn E, Van Mellaert L, Dehaen W, Maes G, Anné J (2004) Analysis of type I signal peptide affinity and specificity for preprotein substrates. *Biochem Biophys Res Commun* 314(2):459–467. <https://doi.org/10.1016/j.bbrc.2003.12.122>
- Green ER, Mecsas J (2016) Bacterial secretion systems—an overview. *Microbiol Spectr* 4(1). <https://doi.org/10.1128/microbiolspec.VMBF-0012-2015>
- Hegde RS, Bernstein HD (2006) The surprising complexity of signal sequences. *Trends Biochem Sci* 31(10):563–571. <https://doi.org/10.1016/j.tibs.2006.08.004>
- Heng CK (2009) Bioinformatic analysis of bacterial and eukaryotic amino-terminal signal peptides
- Hiss JA, Schneider G (2009) Architecture, function and prediction of long signal peptides. *Brief Bioinform* 10(5):569–578. <https://doi.org/10.1093/bib/bbp030>
- Ikeda K, Schiltz E, Fujii T, Takahashi M, Mitsui K, Kadera Y, Matsushima A, Inada Y, Schulz G, Nishimura H (2005) Phenylalanine ammonia-lyase modified with polyethylene glycol: potential therapeutic agent for phenylketonuria. *Amino Acids* 29(3):283–287. <https://doi.org/10.1007/s00726-005-0218-5>
- Ismail NF, Hamdan S, Mahadi NM, Murad AMA, Rabu A, Bakar FDA, Klappa P, Ilias RM (2011) A mutant L-asparaginase II signal peptide improves the secretion of recombinant cyclodextrin glucanotransferase and the viability of *Escherichia coli*. *Biotechnol Lett* 33(5):999–1005. <https://doi.org/10.1007/s10529-011-0517-8>
- Ivankov DN, Payne SH, Galperin MY, Bonissone S, Pevzner PA, Frishman D (2013) How many signal peptides are there in bacteria? *Environ Microbiol* 15(4):983–990. <https://doi.org/10.1111/1462-2920.12105>
- Jonet MA, Mahadi NM, Murad AMA, Rabu A, Bakar FDA, Rahim RA, Low KO, Ilias RM (2012) Optimization of a heterologous signal peptide by site-directed mutagenesis for improved secretion of recombinant proteins in *Escherichia coli*. *J Mol Microbiol Biotechnol* 22(1):48–58. <https://doi.org/10.1159/000336524>
- Joshi SN, Butler DC, Messer A (2012) Fusion to a highly charged proteasomal retargeting sequence increases soluble cytoplasmic expression and efficacy of diverse anti-synuclein intrabodies. *MAbs*, Taylor & Francis, Didcot
- Kim SJ, Mitra D, Salerno JR, Hegde RS (2002) Signal sequences control gating of the protein translocation channel in a substrate-specific manner. *Dev Cell* 2(2):207–217. [https://doi.org/10.1016/S1534-5807\(01\)00120-4](https://doi.org/10.1016/S1534-5807(01)00120-4)
- Kober L, Zehe C, Bode J (2013) Optimized signal peptides for the development of high expressing CHO cell lines. *Biotechnol Bioeng* 110(4):1164–1173. <https://doi.org/10.1002/bit.24776>
- Kong J-Q (2015) Phenylalanine ammonia-lyase, a key component used for phenylpropanoids production by metabolic engineering. *RSC Adv* 5(77):62587–62603. <https://doi.org/10.1039/c5ra08196c>
- Kulothungan SR, Das M, Johnson M, Ganesh C, Varadarajan R (2009) Effect of crowding agents, signal peptide, and chaperone SecB on the folding and aggregation of *E. coli* maltose binding protein. *Langmuir* 25(12):6637–6648. <https://doi.org/10.1021/la900198h>
- Kvam E, Sierks MR, Shoemaker CB, Messer A (2010) Physico-chemical determinants of soluble intrabody expression in mammalian cell cytoplasm. *Prot Eng Des Sel* 23(6):489–498. <https://doi.org/10.1093/protein/gzq022>
- Laskowski RA (2007) Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature. *Bioinformatics* 23(14): 1824–1827. <https://doi.org/10.1093/bioinformatics/btm085>
- Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM (2018) PDBsum: Structural summaries of PDB entries. *Prot Sci* 27(1):129–134. <https://doi.org/10.1002/pro.3289>
- Longo N, Harding CO, Burton BK, Grange DK, Vockley J, Wasserstein M, Rice GM, Dorenbaum A, Neuenburg JK, Musson DG (2014) Single-dose, subcutaneous recombinant phenylalanine ammonia lyase conjugated with polyethylene glycol in adult patients with phenylketonuria: an open-label, multicentre, phase 1 dose-escalation trial. *Lancet* 384(9937):37–44. [https://doi.org/10.1016/S0140-6736\(13\)61841-3](https://doi.org/10.1016/S0140-6736(13)61841-3)
- Low KO, Mahadi NM, Ilias RM (2013) Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Appl Microbiol Biotechnol* 97(9):3811–3826
- MacDonald MJ, D’Cunha GB (2007) A modern view of phenylalanine ammonia lyase. *Biochem Cell Biol* 85(3):273–282. <https://doi.org/10.1139/O07-018>
- Mergulhao F, Summers DK, Monteiro GA (2005) Recombinant protein secretion in *Escherichia coli*. *Biotechnol Adv* 23(3):177–202. <https://doi.org/10.1016/j.biotechadv.2004.11.003>
- Nesmeyanova MA, Karamyshev AL, Karamysheva ZN, Kalinin AE, Ksenzenko VN, Kajava AV (1997) Positively charged lysine at the N-terminus of the signal peptide of the *Escherichia coli* alkaline phosphatase provides the secretion efficiency and is involved in the interaction with anionic phospholipids. *FEBS Lett* 403(2):203–207. [https://doi.org/10.1016/S0014-5793\(97\)00052-5](https://doi.org/10.1016/S0014-5793(97)00052-5)
- Nielsen H (2017) Predicting secretory proteins with SignalP. *Protein Funct Predict Methods Protoc* 59–73

- Nilsson I, Lara P, Hessa T, Johnson AE, von Heijne G, Karamyshev AL (2015) The code for directing proteins for translocation across ER membrane: SRP cotranslationally recognizes specific features of a signal sequence. *J Mol Biol* 427(6):1191–1201. <https://doi.org/10.1016/j.jmb.2014.06.014>
- Notredame C, Higgins DG, Heringa J (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment1. *J Mol Biol* 302(1):205–217. <https://doi.org/10.1006/jmbi.2000.4042>
- Ohmuro-Matsuyama Y, Yamaji H (2017) Modifications of a signal sequence for antibody secretion from insect cells. *Cytotechnology* 70:1–8. <https://doi.org/10.1007/s10616-017-0109-0>
- Palazzo AF, Springer M, Shibata Y, Lee C-S, Dias AP, Rapoport TA (2007) The signal sequence coding region promotes nuclear export of mRNA. *PLoS Biol* 5(12):e322. <https://doi.org/10.1371/journal.pbio.0050322>
- Pfeiffer T, Pisch T, Devitt G, Holtkotte D, Bosch V (2006) Effects of signal peptide exchange on HIV-1 glycoprotein expression and viral infectivity in mammalian cells. *FEBS Lett* 580(15):3775–3778. <https://doi.org/10.1016/j.febslet.2006.05.070>
- Puziss JW, Harvey R, Bassford P (1992) Alterations in the hydrophilic segment of the maltose-binding protein (MBP) signal peptide that affect either export or translation of MBP. *J Bacteriol* 174(20):6488–6497. <https://doi.org/10.1128/jb.174.20.6488-6497.1992>
- Rapoport TA, Jungnickel B, Kutay U (1996) Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu Rev Biochem* 65(1):271–303. <https://doi.org/10.1146/annurev.bi.65.070196.001415>
- Rusch SL, Chen H, Izard JW, Kendall DA (1994) Signal peptide hydrophobicity is finely tailored for function. *J Cell Biochem* 55(2):209–217. <https://doi.org/10.1002/jcb.240550208>
- Sakhteman A, Khoddami M, Negahdaripour M, Mehdizadeh A, Tatar M, Ghasemi Y (2016) Exploring 3D structure of human gonadotropin hormone receptor at antagonist state using homology modeling, molecular dynamic simulation, and cross-docking studies. *J Mol Model* 22(9):225. <https://doi.org/10.1007/s00894-016-3091-0>
- Sarkissian CN, Gámez A (2005) Phenylalanine ammonia lyase, enzyme substitution therapy for phenylketonuria, where are we now?. *Mol Genet Metabol* 86:22–26. <https://doi.org/10.1016/j.ymgme.2005.06.016>
- Sarkissian CN, Gámez A, Wang L, Charbonneau M, Fitzpatrick P, Lemontt JF, Zhao B, Vellard M, Bell SM, Henschell C (2008) Preclinical evaluation of multiple species of PEGylated recombinant phenylalanine ammonia lyase for the treatment of phenylketonuria. *Proc Nat Acad Sci* 105(52):20894–20899. <https://doi.org/10.1073/pnas.0808421105>
- Shrivastava A, Khan AA, Khurshid M, Kalam MA, Jain SK, Singhal PK (2016) Recent developments in L-asparaginase discovery and its potential as anticancer agent. *Crit Rev Oncol Hematol* 100:1–10. <https://doi.org/10.1016/j.critrevonc.2015.01.002>
- Singh P, Sharma L, Kulothungan SR, Adkar BV, Prajapati RS, Ali PSS, Krishnan B, Varadarajan R (2013) Effect of signal peptide on stability and folding of *E. coli* thioredoxin. *PLoS one* 8(5):e63442. <https://doi.org/10.1371/journal.pone.0063442>
- Sletta H, Tøndervik A, Hakvåg S, Aune TV, Nedal A, Aune R, Evensten G, Valla S, Ellingsen T, Brautaset T (2007) The presence of N-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivations of *Escherichia coli*. *Appl Environ Microbiol* 73(3):906–912. <https://doi.org/10.1128/AEM.01804-06>
- Smialowski P, Doose G, Torkler P, Kaufmann S, Frishman D (2012) PROSO II—a new method for protein solubility prediction. *FEBS J* 279(12):2192–2200. <https://doi.org/10.1111/j.1742-4658.2012.08603.x>
- Sumaily KM, Mujammi AH (2017) Phenylketonuria: a new look at an old topic, advances in laboratory diagnosis, and therapeutic strategies. *Int J Health Sci* 11(5):63–70
- Turner NJ (2011) Ammonia lyases and aminomutases as biocatalysts for the synthesis of α -amino and β -amino acids. *Curr Opin Chem Biol* 15(2):234–240. <https://doi.org/10.1016/j.cbpa.2010.11.009>
- von Heijne G (1990) The signal peptide. *J Membr Biol* 115(3):195–201. <https://doi.org/10.1007/BF01868635>
- Wang C-Z, Chi C-W (2004) Conus peptides—a rich pharmaceutical treasure. *Acta Biochim Biophys Sin* 36(11):713–723. <https://doi.org/10.1093/abbs/36.11.713>
- Wang L, Gamez A, Archer H, Abola EE, Sarkissian CN, Fitzpatrick P, Wendt D, Zhang Y, Vellard M, Bliesath J (2008) Structural and biochemical characterization of the therapeutic *Anabaena variabilis* phenylalanine ammonia lyase. *J Mol Biol* 380(4):623–635. <https://doi.org/10.1016/j.jmb.2008.05.025>
- Watanabe SK, Hernandez-Velazco G, Iturbe-Chinas F, Lopez-Munguia A (1992) Phenylalanine ammonia lyase from *Sporidiobolus parvoseus* and *Rhodospiridium toruloides*: Application for phenylalanine and tyrosine deamination. *W J Microbiol Biotechnol* 8(4):406–410
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Weltman JK, Skowron G, Liorot GB (2007) Influenza A H5N1 hemagglutinin cleavable signal sequence substitutions. *Biochem Biophys Res Commun* 352(1):177–180. <https://doi.org/10.1016/j.bbrc.2006.10.184>
- Zalucki YM, Jennings MP (2017) Signal peptidase I processed secretory signal sequences: selection for and against specific amino acids at the second position of mature protein. *Biochem Biophys Res Commun* 483(3):972–977. <https://doi.org/10.1016/j.bbrc.2017.01.044>
- Zhang W, Xiao W, Wei H, Zhang J, Tian Z (2006) mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in *Escherichia coli*. *Biochem Biophys Res Commun* 349(1):69–78. <https://doi.org/10.1016/j.bbrc.2006.07.209>
- Zhou Y, Lu Z, Wang X, Selvaraj JN, Zhang G (2018) Genetic engineering modification and fermentation optimization for extracellular production of recombinant proteins using *Escherichia coli*. *Appl Microbiol Biotechnol* 102(4):1545–1556. <https://doi.org/10.1007/s00253-017-8700-z>
- Zhu F, Liu F, Wu B, He B (2016) Efficient extracellular expression of metalloprotease for Z-aspartame synthesis. *J Agric Food Chem* 64(51):9631–9638. <https://doi.org/10.1021/acs.jafc.6b04164>
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406–3415. <https://doi.org/10.1093/nar/gkg595>