Trends in Bioinformatics

ISSN 1994-7941

science alert

ANSInet
an open access publisher
http://ansinet.com

OPEN ACCESS Trends in Bioinformatics

ISSN 1994-7941
DOI: 10.3923/tb.2018.33.43

Research Article
Signal Peptide Sequence Analysis of Selected Protein Sequences from *Cryptosporidium parvum*

[1]Mohd Aiman Barudin, [2,3]Muhammad Lokman Md Isa and [2,3]Afzan Mat Yusof

[1]Department of Biomedical Science, Kulliyyah of Allied Health Sciences, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia
[2]Department of Basic Medical Sciences, Kulliyyah of Nursing, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia
[3]Integrated Cellular and Molecular Biology Cluster (iMolec), International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia

Abstract

Background and Objective: Studies on signal peptides remain little and lacking to *Cryptosporidium*. This study represents the first attempt to investigate selected putative protein sequences of *Cryptosporidium parvum* in its established genome database for signal peptide. **Methodology:** Prediction analysis of protein sequences and identification of signal peptides on this parasite were analysed by SignalP 4.1 server. A total of 100 protein sequences were randomly selected and analysed, of which 6 (6%) corresponded to be predicted as signal peptide sequences. **Results:** Based on the comparison of sequences with database protein sequences in GenBank NCBI, almost all protein sequences were highly conserved with other protein sequences from different species of *Cryptosporidium*. Considering the secretory proteins in *Cryptosporidium*, several virulence protein sequences of this parasite may also show presence of N-terminal signal peptide or not. **Conclusion:** Hence, further studies need to be carried out for identification of N-terminal signal peptide function in *Cryptosporidium* for facilitating its virulence and pathogenesis pathway.

Key words: *Cryptosporidium*, N-terminal, peptide, protein, signal

Citation: Mohd Aiman Barudin, Muhammad Lokman Md Isa and Afzan Mat Yusof, 2018. Signal peptide sequence analysis of selected protein sequences from *Cryptosporidium parvum*. Trends Bioinform., 11: 33-43.

Corresponding Author: Afzan Mat Yusof, Department of Basic Medical Sciences, Kulliyyah of Nursing, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia Tel: +60129656253

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Cryptosporidium is a protozoan parasite causing cryptosporidiosis. It is regarded as minimally invasive and has made intestinal epithelial cells (IECs) as a host of interest for completing its life cycle[1]. Despite advances in the immunogenetics and genomics studies of *Cryptosporidium*, hiatus of knowledge has been present until now[2]. One of the challenges that remains undiscovered to this parasite is signal peptide proteins. In fact, the study on signal peptides remains little and lacking to *Cryptosporidium*. To date, only one study has been reported on signal peptide related to filarial nematode parasite, *Brugia Malayi*[3].

With regard to this matter, excretory-secretory (ES) products are a subject of this study for revealing the potential signal peptides from *Cryptosporidium* proteins. These were probably to inherit from independent progenitors to be useful for parasitic adaptation[4,5]. They are considered to involve an important function in parasitism and modulation of host immunity[6,7]. These ES proteins generally consist of cytokines, proteinases and antioxidant enzymes, respectively[8-10]. Most of those proteins are vital for immune response within the host, invading of host tissue and acting as reactive oxygen species-protective agent on the infection-stimulated host phagocytes[11-13].

With the advancement of genomic applications, a number of software for secretory proteins prediction are available[14,15]. It is noted that SignalP version 4.1 is regarded as the most accurate program that analyses protein sequence at N termini as determined by the application of Neural Networks (NN) and Hidden Markov Models (HMM) to predict co-translational and translocation of proteins[16]. This parasite is widely known as a neglected parasite worldwide. With regard to urgent interest of drug development against *Cryptosporidium*, it is noted that ES proteins have gained insight due to their important function of parasitism in cryptosporidiosis. In fact, *C. parvum* genome has been established more than a decade ago[17]. Hence, it provides a great platform to run the data mining analysis of secretory protein sequences in this parasite, including the retrieval of those protein sequences for the matter of signal peptide analysis as reported in this article.

MATERIALS AND METHODS

This present study was carried out in January, 2018 to April, 2018. Selected protein sequences of *C. parvum* were retrieved indiscriminately from NCBI GenBank in FASTA format. Subsequently, these protein sequences were subjected to BLASTP analyses[18]. SignalP 4.1 server was applied to those protein sequences for prediction of secretory signal peptides[16].

RESULTS AND DISCUSSION

About 100 protein sequences of *Cryptosporidium parvum* were randomly retrieved and then performed BLASTP analyses in comparison to non-redundant protein in NCBI GenBank database (Table 1). Similarity search is important to provide new information about nucleotide or protein sequences using the most widely used method like BLAST[19]. The sequences were paired with existing protein database sequences based on the high similarity of sequence coverage and low E-value. Most sequences either nucleotide or protein form are necessarily to compare with databases based on sequence similarity searching[20]. Most of the protein sequences were similar to database sequences of the same genus, *Cryptosporidium* but were slightly different in species like *C. hominis* and *C. ubiquitum*. Only one hypothetical protein of *C. parvum* (accession number: XP_001388429.1) had no significant hit after BLASTP search on GenBank database. It possibly means that this protein sequence is potentially a novel and different to any available protein entry sequences. However, no significant hit or no significant similarity can be simply described as low complexity regions (LCRs) with unusual or biased amino acid composition containing little diversity whereby causing artifactual hits[21]. SignalP 4.1 was then applied to analyse the protein dataset, which anticipated that 6% (6 of 100) of the protein sequences was secretory proteins which definitely possessed N-terminal signal peptides (Table 1, Fig. 1-6). Signal sequence analysis of *Cryptosporidium* hypothetical protein (Fig. 1, accession number: XP_001388414.1) has D-score of 0.789 with cutoff value of 0.45 at the position of amino acid residue starting from 1-26. D-score is important to be noted in this study because it is used for discriminating signal peptides from non-signal peptides[22]. Recent studies revealed that signal peptides that were analysed based on D-score value using signal peptide prediction tool such as SignalP 4.1 were investigated from *Escherichia coli*, *Pseudomonas syringae*, *Anabaena* sp., *Pichia pastoris*, *Neurospora crassa*, *Ralstonia solanacearum*, *Caenorthaditis elegans*, *Saccharomyces cerevisiae*, *Schistosoma mansoni* and *Agrobacterium tumefaciens*, respectively[23-33]. Other than that, S-score is signal peptide score that corresponds to the output of this server analysis. Signal sequence analysis of *Cryptosporidium* S1/P1nuclease (Fig. 2, accession number: XP_001388393.1) has 0.743 D-score value at the position of 1-22 amino acid

Table 1: SignalP 4.1 result of selected protein sequences and with homologs in other species of *Cryptosporidium* from GenBank NCBI database

Accession number of *Cryptosporidium* proteins	Name of *Cryptosporidium* protein	SignalP 4.1 result	Cleavage site	BLASTP hit search (accession number)	E-value
XP_001388452.1	Hypothetical protein	No	No	XP_666686.1	0
XP_001388451.1	Protein kinase NPK2	No	No	XP_666680.1	0
XP_001388450.1	Hypothetical protein, partial	No	No	OII75255.1	7^{e-97}
XP_001388449.1	Hypothetical protein	No	No	OLQ19452.1	7e-123
XP_001388448.1	Clathrin assembly protein	No	No	OII75239.1	8e-100
XP_001388447.1	tRNA synthetase class II, partial	No	No	XP_665011.1	1e-151
XP_001388446.1	Hypothetical protein, partial	No	No	OLQ16367.1	4e-105
XP_001388445.1	Ribosomal processing protein	No	No	CUV08004.1	0
XP_001388444.1	Hypothetical protein	No	No	CUV07998.1	0
XP_001388443.1	Cell differentiation protein rcd1	No	No	XP_667107.1	0
XP_001388442.1	Hypothetical protein	No	No	CUV07982.1	6e-139
XP_001388441.1	Hypothetical protein	No	No	CUV07976.1	5e-165
XP_001388440.1	Hypothetical protein, partial	No	No	OLQ17422.1	0
XP_001388439.1	RIKEN cDNA 9430077D24 gene	No	No	CUV07948.1	0
XP_001388438.1	Hypothetical protein, partial	No	No	CUV07915.1	1e-108
XP_001388437.1	Hypothetical protein	No	No	XP_668533.1	0
XP_001388436.1	Hypothetical protein, partial	No	No	CUV07893.1	5e-67
XP_001388435.1	Phosphatidylinositol-4-phosphate 5-kinase	No	No	OLQ17342.1	0
XP_001388434.1	Hypothetical protein	No	No	OLQ17330.1	2e-67
XP_001388433.1	Serine/threonine protein kinase KKIALRE	No	No	XP_667043.1	0
XP_001388432.1	Hypothetical protein, partial	No	No	OLQ19366.1	2e-122
XP_001388431.1	Hypothetical protein, partial	No	No	OLQ19365.1	2e-102
XP_001388430.1	Cyclin dependent kinase regulatory subunit, partial	No	No	CUV07800.1	2e-54
XP_001388429.1	Hypothetical protein, partial	No	No	No significant hit	No
XP_001388428.1	Cyclophilin-RNA interacting protein	No	No	XP_666493.1	0
XP_001388427.1	Hypothetical protein	No	No	XP_667371.1	0
XP_001388426.1	40S ribosomal subunit protein S9	No	No	OII73614.1	3e-69
XP_001388425.1	ATP-dependent RNA helicase	No	No	OII73612.1	0
XP_001388424.1	Protein kinase	No	No	XP_666886.1	0
XP_001388423.1	Hypothetical protein	No	No	OII72591.1	1e-154
XP_001388422.1	Coatomer protein complex subunit alpha, partial	No	No	OLQ18342.1	0
XP_001388421.1	Glutamate--tRNA ligase, partial	No	No	XP_665881.1	0
XP_001388420.1	Ubiquitin-conjugating enzyme E2, partial	No	No	OII70985.1	1e-97
XP_001388419.1	CDP-diacylglycerol--inositol 3-phosphatidyltransferase isoform 1; Phosphatidylinositol synthase; PtdIns synthase; PI synthase	No	No	XP_667153.1	1e-105
XP_001388418.1	DNA repair protein	No	No	XP_667158.1	0
XP_001388417.1	Exoribonuclease PH	No	No	OII70959.1	1e-170
XP_001388416.1	Ubiquitin-conjugating enzyme e2	No	No	XP_667527.1	2e-112
XP_001388415.1	NIMA-related kinase 5	No	No	XP_666371.1	0
XP_001388414.1	Hypothetical protein	Yes	Between 26 and 27	OLQ18156.1	0
XP_001388413.1	Hypothetical protein, partial	No	No	OII75050.1	2e-80
XP_001388412.1	Hypothetical protein	No	No	CUV07474.1	0
XP_001388411.1	Ribosomal protein S4, partial	No	No	OII75039.1	0
XP_001388410.1	Hypothetical protein, partial	No	No	XP_666552.1	5 e - 8 3
XP_001388409.1	Ubiquitin-like protein	No	No	OII75016.1	1e-48
XP_001388408.1	ATP-binding cassette, sub-family C (CFTR/MRP), member 2; Canalicular multispecific organic anion transporter; multidrug resist, partial	No	No	CUV07439.1	0
XP_001388407.1	Rablib, partial	No	No	XP_667916.1	6e-146
XP_001388406.1	RNA polymerase II	No	No	OII74999.1	5e-67
XP_001388405.1	Hypothetical protein	No	No	CUV07424.1	9e-100
XP_001388404.1	Sec61-gamma subunit of protein translocation complex, partial	No	No	OII74983.1	9e-52
XP_001388403.1	Hypothetical protein	No	No	CUV07406.1	6e-123
XP_001388402.1	Arsenical pump-driving ATPase	No	No	OLQ18072.1	0
XP_001388401.1	Hypothetical protein	No	No	XP_665976.1	0
XP_001388400.1	Hypothetical protein, partial	No	No	OLQ18064.1	4e-80
XP_001388399.1	Small nuclear ribonucleoprotein, partial	No	No	CUV07385.1	1e-33
XP_001388398.1	Pleiotropic regulator 1	No	No	XP_668039.1	0

Table 1: Continue

Accession number of *Cryptosporidium* proteins	Name of *Cryptosporidium* protein	SignalP 4.1 result	Cleavage site	BLASTP hit search (accession number)	E-value
XP_001388397.1	Calcium-dependent protein kinase 7 (CDPK)(CPK7)	No	No	XP_668032.1	0
XP_001388396.1	Hypothetical protein, partial	No	No	OLQ18040.1	3e-180
XP_001388395.1	Helicase	No	No	CUV07337.1	0
XP_001388394.1	Transcription activator, partial	No	No	CUV07310.1	3e-28
XP_001388393.1	S1/P1nuclease	Yes	Between 22 and 23	XP_668268.1	0
XP_001388392.1	Hypothetical protein	No	No	CUV07294.1	0
XP_001388391.1	Ubiquitin conjugating enzyme, partial	No	No	OII74155.1	4e-115
XP_001388390.1	Ubiquitin-conjugating enzyme E2, partial	No	No	CUV07266.1	2e-89
XP_001388389.1	FALZ protein	No	No	XP_666108.1	0
XP_001388388.1	Serine/threonine protein phosphatase	No	No	OII75651.1	0
XP_001388387.1	Hypothetical protein, partial	No	No	CUV07241.1	1e-160
XP_001388386.1	Skp1 family protein	No	No	CUV07236.1	2e-111
XP_001388385.1	Hypothetical protein, partial	No	No	CUV07214.1	2e-109
XP_001388384.1	Transmembrane protein, partial	Yes	Between 17 and 18	CUV07199.1	5e-144
XP_001388383.1	Flap endonuclease 1	No	No	XP_667077.1	0
XP_001388382.1	Ruv DNA-helicase-related protein	No	No	XP_668483.1	0
XP_001388381.1	Step II splicing factor SLU7	No	No	XP_666864.1	0
XP_001388380.1	Endonuclease III, partial	No	No	OLQ17844.1	3e-138
XP_001388379.1	Rab7 GTPase	No	No	OII74041.1	3e-159
XP_001388378.1	Uracil-DNA glycosylase	No	No	XP_668581.1	3e-141
XP_001388377.1	Aspartate--tRNA ligase	No	No	OII74024.1	0
XP_001388376.1	Peptidylprolyl isomerase, partial	No	No	OLQ17821.1	6e-47
XP_001388375.1	Hypothetical protein	No	No	XP_668595.1	0
XP_001388374.1	Casein kinase II, alpha subunit	Yes	Between 21 and 22	XP_668610.1	0
XP_001388373.1	Hypothetical protein, partial	No	No	CUV07092.1	0
XP_001388372.1	Hypothetical protein	No	No	CUV07083.1	0
XP_001388371.1	Hypothetical protein	No	No	CUV07081.1	0
XP_001388370.1	Patched family protein, partial	No	No	OII73957.1	0
XP_001388369.1	N-acetylglucosaminyl-phosphatidylinositol de-N-acetylase	No	No	OLQ17743.1	2e-122
XP_001388368.1	Hypothetical protein	No	No	CUV07032.1	0
XP_001388367.1	Cyclin-dependent kinase-related kinase	No	No	XP_665899.1	0
XP_001388366.1	Hypothetical protein	No	No	CUV07017.1	0
XP_001388365.1	Cell division control protein 28	No	No	OII73900.1	0
XP_001388364.1	GTP-binding nuclear protein ran/tc4	No	No	OII73894.1	2e-156
XP_001388363.1	Methyltransferase-related, partial	No	No	OII73888.1	2e-166
XP_001388362.1	cAMP-dependent protein kinase regulatory subunit, partial	No	No	OLQ17684.1	0
XP_001388361.1	High affinity sulfate transporter-related	No	No	XP_667980.1	0
XP_001388360.1	Hypothetical protein	No	No	CUV06978.1	2e-154
XP_001388359.1	Peptidase'insulinase like peptidase'	Yes	Between 21 and 22	CAD98358.1	0
XP_001388358.1	ABC transporter protein, partial	Yes	Between 20 and 21	OLQ16548.1	0
XP_001388357.1	Multi-pass transmembrane protein	No	No	CUV06936.1	0
XP_001388356.1	Transcription factor	No	No	OLQ16521.1	0
XP_001388355.1	Serine/threonine-protein kinase	No	No	CUV06923.1	0
XP_001388354.1	Acylphosphatase, partial	No	No	CAD98307.1	5e-73
XP_001388353.1	Hypothetical protein	No	No	XP_666696.1	0

In this table, accession number and name of *Cryptosporidium* proteins, SignalP 4.1 result and its cleavage site, BLASTP hit search (accession number) and last but not least E-value. If the result is "yes" from SignalP 4.1 server, there will be a cleavage site detected in the protein sequence. E value is better to be referenced if the value is below than 0

residues. D-score of 0.647 from signal sequence analysis of *Cryptosporidium* trans membrane protein (Fig. 3, accession number: XP_001388384.1) located at the first amino acid residue to 17th position. Signal sequence analysis of *Cryptosporidium* casein kinase II, alpha subunit (Fig. 4, accession number: XP_001388374.1) possess 0.769 value corresponding to D-score encompassing of 1-21 amino acid residues. Signal sequence analysis of *Cryptosporidium* peptidase' insulinase like peptidase' (Fig. 5, accession number: XP_001388359.1) has 0.608 value of D-score at the first to

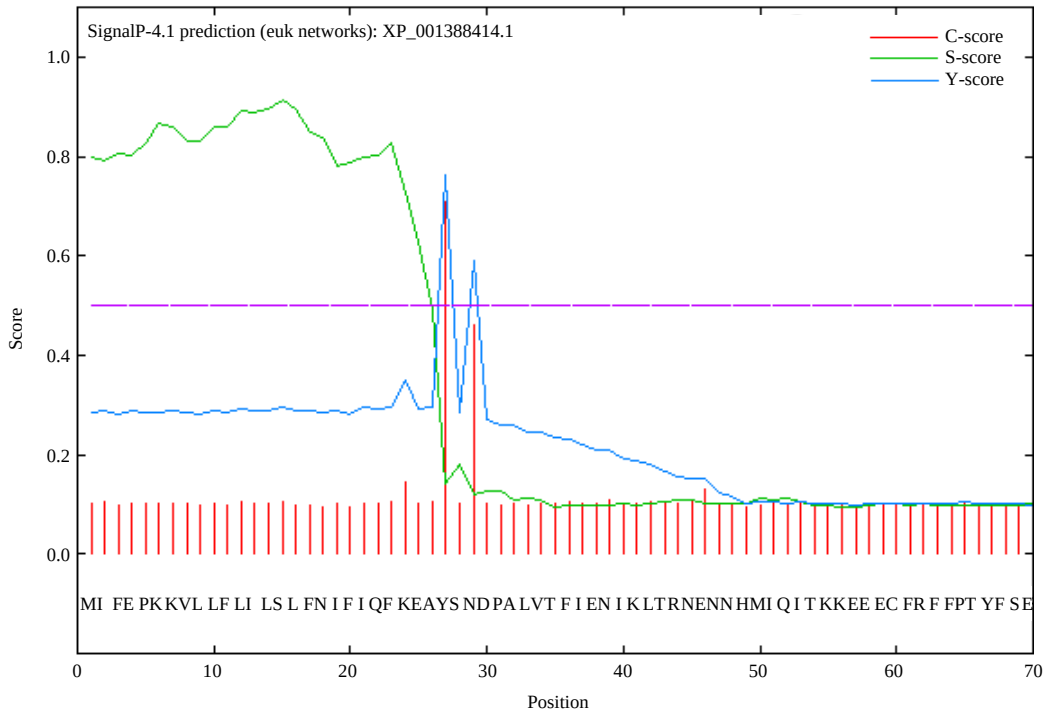

Fig. 1: Signal sequence analysis of *Cryptosporidium* hypothetical protein (accession number: XP_001388414.1). The C-score is a raw cleavage site score and S-score is a signal peptide score, Y-score is a combination of C-score and S-score. The cutoff value of more than 0.45 shows the presence of signal peptide in protein sequence

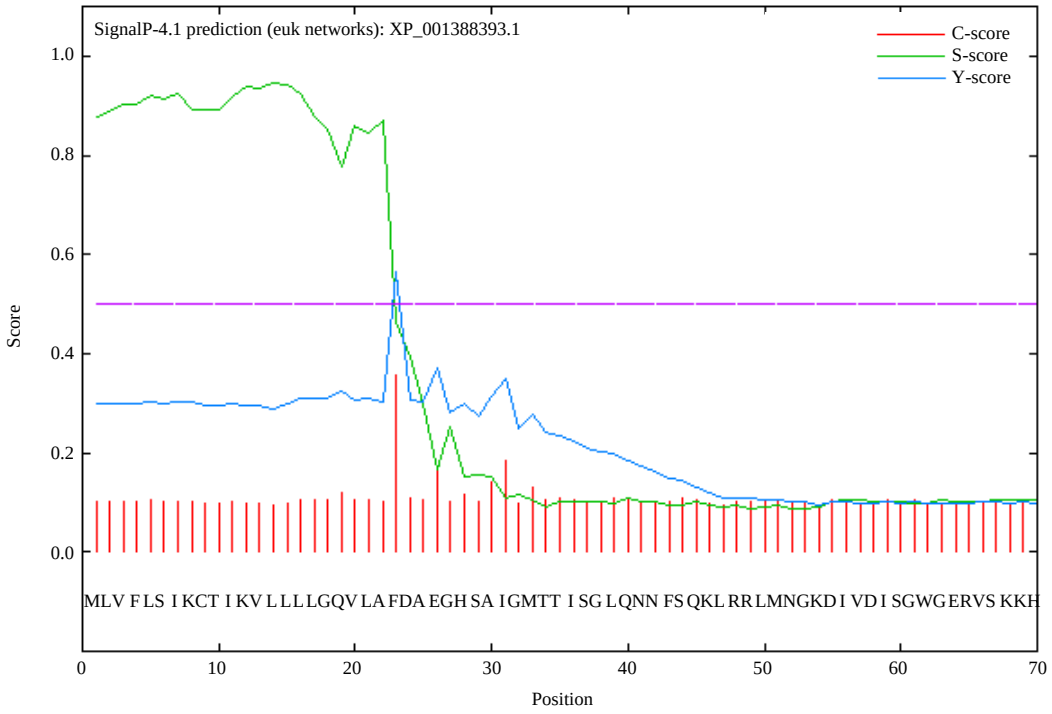

Fig. 2: Signal sequence analysis of *Cryptosporidium* S1/P1nuclease (accession number: XP_001388393.1). The C-score is a raw cleavage site score and S-score is a signal peptide score, Y-score is a combination of C-score and S-score. The cutoff value of more than 0.45 shows the presence of signal peptide in protein sequence

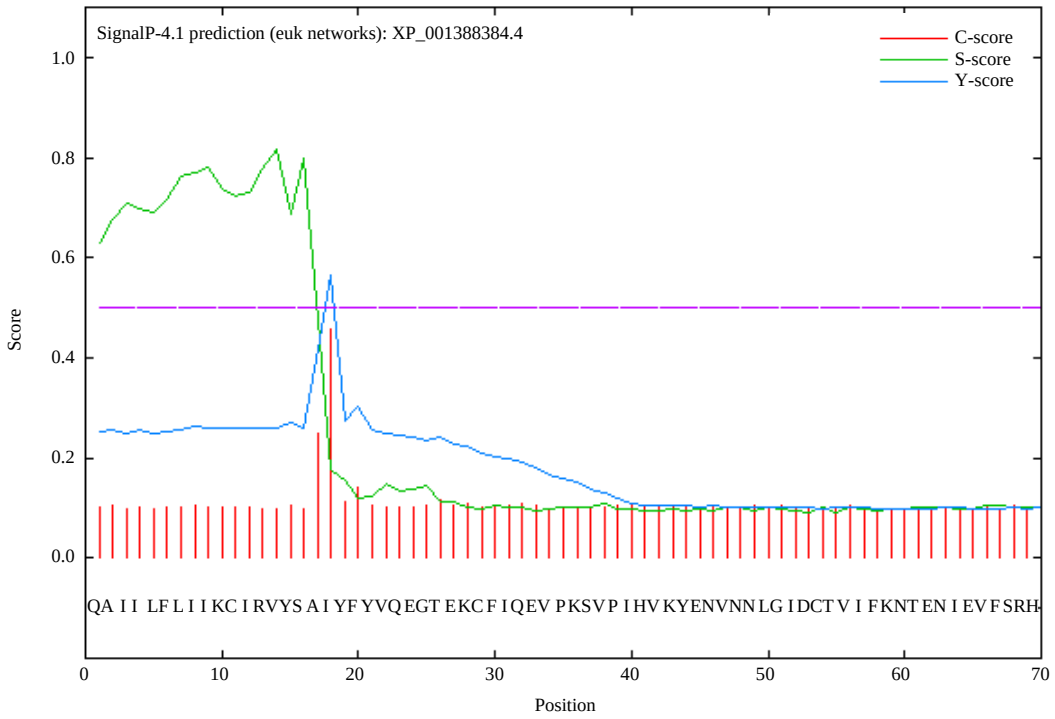

Fig. 3: Signal sequence analysis of *Cryptosporidium* trans membrane protein (accession number: XP_001388384.1). The C-score is a raw cleavage site score and S-score is a signal peptide score, Y-score is a combination of C-score and S-score. The cutoff value of more than 0.45 shows the presence of signal peptide in protein sequence

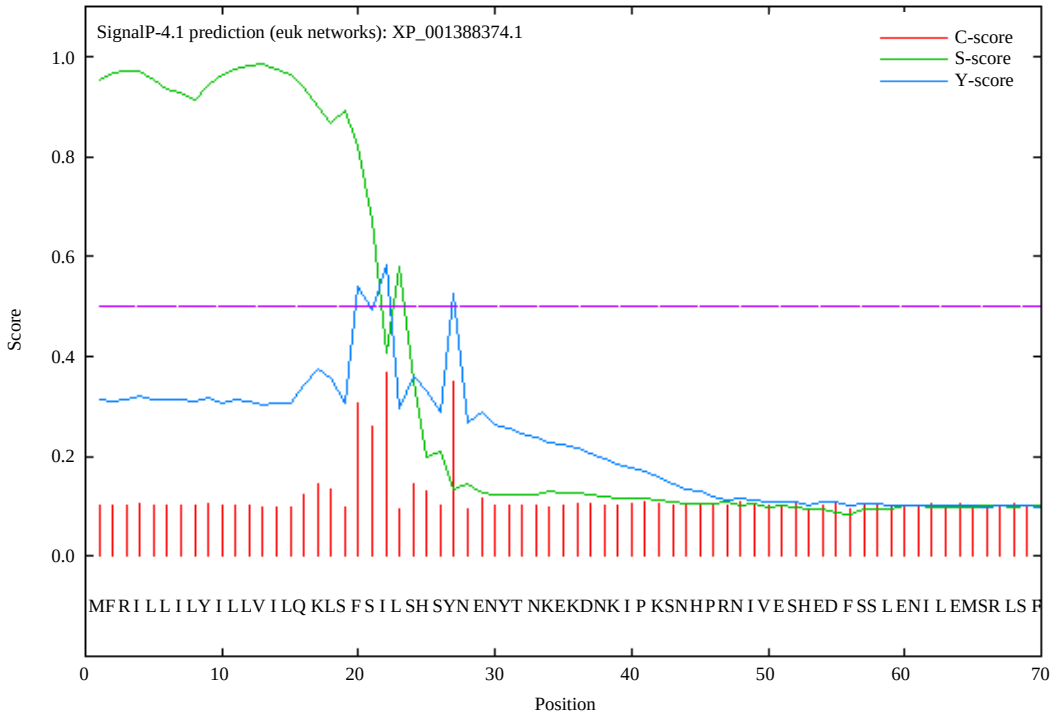

Fig. 4: Signal sequence analysis of *Cryptosporidium* casein kinase II, alpha subunit (accession number: XP_001388374.1). The C-score is a raw cleavage site score and S-score is a signal peptide score, Y-score is a combination of C-score and S-score. The cutoff value of more than 0.45 shows the presence of signal peptide in protein sequence

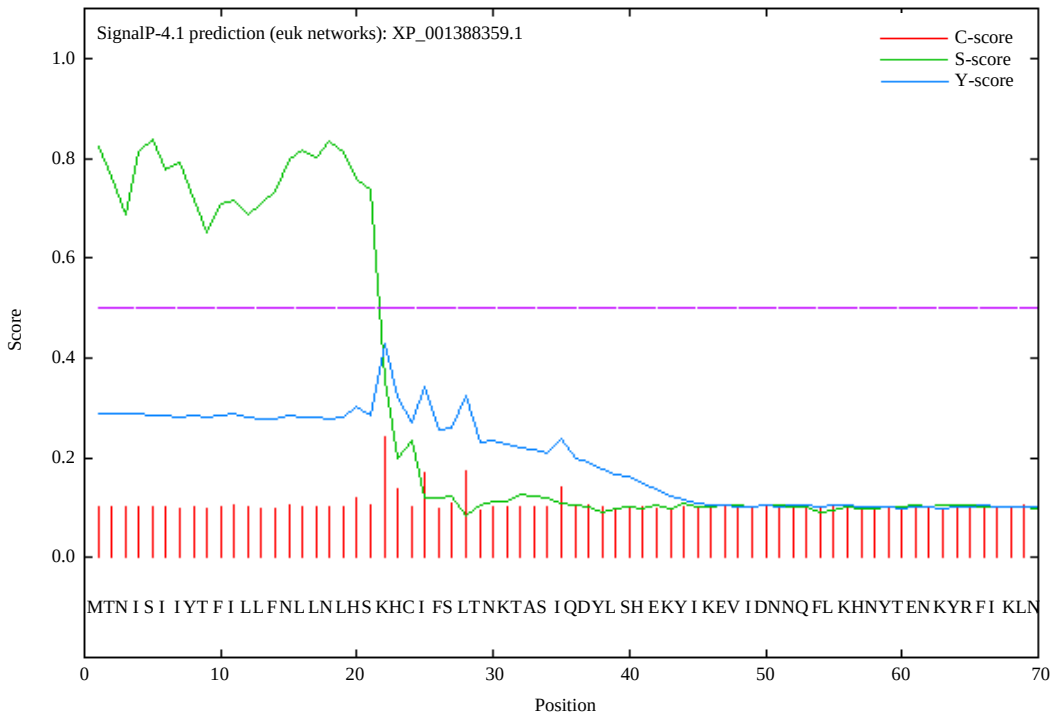

Fig. 5: Signal sequence analysis of *Cryptosporidium* peptidase' insulinase like peptidase' (accession number: XP_001388359.1). The C-score is a raw cleavage site score and S-score is a signal peptide score, Y-score is a combination of C-score and S-score. The cutoff value of more than 0.45 shows the presence of signal peptide in protein sequence

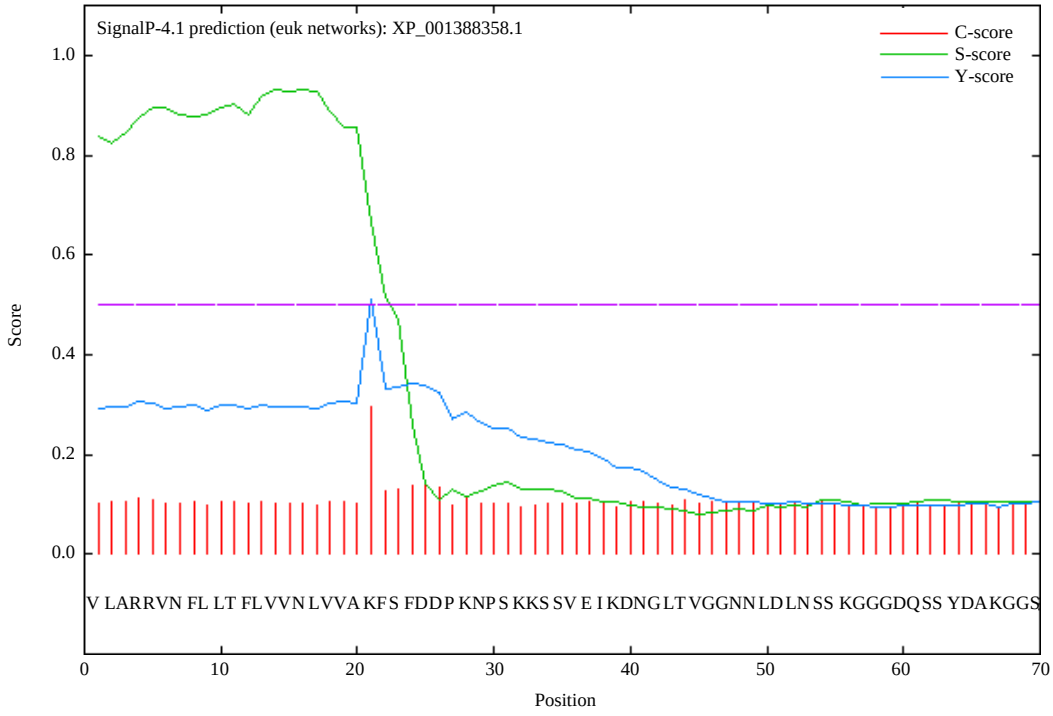

Fig. 6: Signal sequence analysis of *Cryptosporidium* ABC transporter protein (accession number: XP_001388358.1). The C-score is a raw cleavage site score and S-score is a signal peptide score, Y-score is a combination of C-score and S-score. The cutoff value of more than 0.45 shows the presence of signal peptide in protein sequence

Table 2: Excretory/secretory virulence protein sequences with/out N-terminal signal peptides in *Cryptosporidium parvum*

Protein name	Accession number	References
Proteins without N-terminal signal peptides		
Aminopeptidase	EAK87928.1	Abrahamsen *et al.*[17]
Heat shock protein 70	XP_625373.1	Abrahamsen *et al.*[17]
Heat shock protein 90	XP_626924.1	Abrahamsen *et al.*[17]
Glycoprotein 900	AF527875.1	Sturbaum *et al.*[39]
RNA polymerase III C5 subunit	XP_625493.1	Abrahamsen *et al.*[17]
CP47	AAM46174.1	Nil
Cysteine protease	XP_001388329.1	Abrahamsen *et al.*[17]
Serine protease	XP_627811.1	Abrahamsen *et al.*[17]
Phospholipase C	EAK87844.1	Abrahamsen *et al.*[17]
ATPase	XP_626796.1	Abrahamsen *et al.*[17]
Acyl-CoA synthetases	XP_626845.1	Abrahamsen *et al.*[17]
Type I fatty acid synthase	XP_626786.1	Abrahamsen *et al.*[17]
Polyketide synthase	EAK87820.1	Abrahamsen *et al.*[17]
Haemolytic protein	EAK88625.1	Abrahamsen *et al.*[17]
CpCCp3	XP_626313.1	Abrahamsen *et al.*[17]
ATP-binding cassette protein	XP_001388145.1	Abrahamsen *et al.*[17]
ATP-dependent helicase	XP_001388086.1	Abrahamsen *et al.*[17]
RNase L inhibitor-like protein	XP_001388150.1	Abrahamsen *et al.*[17]
Acyltransferase	XP_628464.1	Abrahamsen *et al.*[17]
Glycerol-3-phosphate acyltransferase	EAK89355.1	Abrahamsen *et al.*[17]
Diacylglycerol acyltransferase	XP_626337.1	Abrahamsen *et al.*[17]
Histone acetyltransferase	XP_627612.1	Abrahamsen *et al.*[17]
Proteins with N-terminal signal peptides		
P23	XP_627530.1	Abrahamsen *et al.*[17]
TRAP-C1	XP_628162.1	Abrahamsen *et al.*[17]
Mucin	EAK89464.1	Abrahamsen *et al.*[17]
CP2	AAR36877.1	O'Hara *et al.*[40]
CpCCP2	XP_628222.1	Abrahamsen *et al.*[17]
CpCCp1/Cpa135	XP_628351.1	Abrahamsen *et al.*[17]

This table shows several protein sequences of *Cryptosporidium parvum* that have the N-terminal signal peptides or not. Some of them are regarded as having characteristics of virulence for this parasite that could be potentially vital for having signal peptide to act as virulence factor

twenty first position of amino acid residues position. Last but not least, signal sequence analysis of *Cryptosporidium* ABC transporter protein (Fig. 6, accession number: XP_001388358.1) has D-score value of 0.714 occupying up to 20 amino acid residues from the beginning residue of signal peptide sequence. The C-score is a raw cleavage site score and S-score is a signal peptide score. The Y-score is a combination of C-score and S-score. The cutoff value of more than 0.45 shows the presence of signal peptide in protein sequence and sensitivity of the output value for the signal peptide regions[22]. In principle, N-terminal or amino-terminal signal peptides are known as short sequences that function to target and guide only secretory proteins through translocation heading to subcellular compartments like endoplasmic reticulum inside the cell[34]. Once it reaches the site that it is destined for, signal peptide will be cleaved off from the passenger protein sequences. On the other hand, the output of SignalP 4.1 that has very low scores corresponds only to non-secretory proteins. In fact, non-secretory proteins without N-terminal signal peptide are commonly known as leaderless secretion protein or extracellular proteins[35]. It somehow can be exported out without a guide of N-terminal signal peptide. However, the rest of other analysed protein sequences were not found to have N-terminal signal peptides and hence, further studies need to be done for revealing the real matter behind it. Among most commonly used signal peptide prediction tools that are used, SignalP tool is more superior to give accuracy and consistency of the prediction results based on its composite scoring schemes besides its capability to differentiate between secretory and non-secretory proteins[34]. However, identification of correct signal peptide cleavage site remains a challenge to be clearly pinpointed.

Besides, the present results suggest that N-terminal signal peptide might be important in virulence and pathogenesis of cryptosporidiosis. Based on the details of virulence proteins from *Cryptosporidium parvum*, several proteins also have been shown to have no signal peptide sequences (Table 2). In fact, secretory proteins have a signal peptide that is cleaved off during pathogenicity and virulence process of potential pathogens like *Mycobacterium tuberculosis* that are destined to be exported via the signal peptidase I-dependent pathway[36]. On the other hand, hypothetical protein,

Table 3: Result of SignalP 4.1 output for protein sequences of *Cryptosporidium parvum* with signal peptides

Accession number of protein	Measure	Position	Value	Cutoff value	Signal peptide
XP_001388414.1	Max. C	27	0.710	0.45	Yes
	Max. Y	27	0.761		
	Max. S	15	0.913		
	Mean S	1-26	0.813		
	D-score	1-26	0.789		
XP_001388393.1	Max. C	23	0.357	0.45	Yes
	Max. Y	23	0.565		
	Max. S	14	0.945		
	Mean S	1-22	0.895		
	D-score	1-22	0.743		
XP_001388384.1	Max. C	18	0.458	0.45	Yes
	Max. Y	18	0.565		
	Max. S	14	0.817		
	Mean S	1-17	0.716		
	D-score	1-17	0.647		
XP_001388374.1	Max. C	22	0.368	0.45	Yes
	Max. Y	22	0.584		
	Max. S	13	0.984		
	Mean S	1-21	0.927		
	D-score	1-21	0.769		
XP_001388359.1	Max. C	22	0.241	0.45	Yes
	Max. Y	22	0.429		
	Max. S	5	0.838		
	Mean S	1-21	0.760		
	D-score	1-21	0.608		
XP_001388358.1	Max. C	21	0.296	0.45	Yes
	Max. Y	21	0.511		
	Max. S	14	0.932		
	Mean S	1-20	0.886		
	D-score	1-20	0.714		

This table shows several protein sequences of *Cryptosporidium parvum* that have higher value of SignalP 4.1 output ranging from 0.5 up to close to 1.0. Signal peptide is considered available if the maximum S value shows the significant value at specific position of amino acid residue on protein sequences. Generally, cutoff value shows 0.45 for signal peptide availability and D-score is normally higher than cutoff value for discriminating signal peptides from non-signal peptides

S1/P1nuclease, trans membrane protein, alpha subunit casein kinase II, peptidase'insulinase like peptidase' and ABC transporter protein that were reported in this study have N-terminal signal peptides besides they are known as virulence factors. In comparison with other virulence proteins of *Cryptosporidium parvum*, many of those proteins do not contain signal peptide and most of them are housekeeping proteinsin this parasite (Table 2). This housekeeping proteins can also be known as moonlighting proteins that have lacking signal peptides even though they somehow can enhance putative role of virulence factors especially on extracellular localization of parasites[37]. However, several virulence protein sequences such as P23, TRAP-C1, mucin, CP2, CpCCP2 and CpCCp1/Cpa135 have been predicted to have N-terminal signal peptides. These virulence proteins are regarded to function in virulence and pathogenicity of this parasite along with those proteins which are non-signal peptide N-terminal protein sequences[38].

SignalP 4.1 server is a standalone software that runs the prediction of signal peptide cleavage sites on amino acid sequences in terms of application in combined artificial neural networks[14,16]. In fact, it is recently the most commonly used program for signal peptides prediction. However, proteins containing signal peptides are not commonly secreted even though it is destined to the secretory pathway of post-translational protein synthesis[16]. In this study, *Cryptosporidium* is a eukaryotic organism and it is predicted to have signal peptide for certain protein sequences using this server (Table 3). From Table 3, C-score is the output of cleavage site networks and functions in differentiating signal peptide cleavage site from the rest of non-related particles[41]. S-score is also one of the important outputs in SignalP 4.1 server that works from the signal peptide score network. It is essential for distinguishing positions of signal peptides from both mature part of proteins and non-signal peptides of proteins[15]. Besides, Y-score is a mix of C-score and S-score that is also used for distinguishing signal peptides from non-signal peptides. In addition, SignalP 4.1 server also implemented the score such as mean S and discrimination score (D-score). Both are used in average S-score of the predicted signal peptide and weighted average of the mean S with the maximal Y-score, respectively[16].

CONCLUSION

As a conclusion, signal peptides of *Cryptosporidium parvum* basically can be analysed with the most widely used signal peptide prediction tool like SignalP 4.1. The N-terminal signal peptides commonly are destined for translocation in subcellular compartments of cells via secretory pathway. In contrast, some of them are regarded to be leaderless region or moonlighting proteins without having N-terminal signal peptide. However, both types are regarded to be potentially involved in virulence and pathogenicity of this parasite. It is about a matter of time to be revealed on signal peptide studies among other parasites as well including this waterborne parasite, *Cryptosporidium*.

SIGNIFICANCE STATEMENT

This study discover the importance of signal peptide to act as a secretory signal for certain proteins that are destined for certain functions through secretory pathways like virulence and pathogenicity process of this parasite. Essentially, the results of this study can be beneficial for other researchers to explore in depth for signal peptides analysis for other parasites as well, not only to this parasite. Indirectly, this study will help the researcher to uncover the other protein regions like moonlighting proteins that are without N-terminal signal peptide. It is hypothesized that this type of protein could be potentially enhance the virulence factors as similar to those of N-terminal signal peptide protein regions. Last but not least, exposure on signal peptide studies could be possibly driven for the discovery of new approach on how to pinpoint accurately the signal peptide coverage site on the protein sequences besides using the signal peptide prediction tools like SignalP 4.1. Thus a new theory on developing a new prediction tool for improving its accuracy may be arrived at in future.

ACKNOWLEDGMENT

The study was funded by an IIUM Research Initiative Grant Scheme (RIGS) grant under research grant No. RIGS 16-301-0465.

REFERENCES

1. Laurent, F. and S. Lacroix-Lamande, 2017. Innate immune responses play a key role in controlling infection of the intestinal epithelium by *Cryptosporidium*. Int. J. Parasitol., 47: 711-721.
2. Checkley, W., A.C. White, D. Jaganath, M.J. Arrowood and R.M. Chalmers *et al*., 2015. A review of the global burden, novel diagnostics, therapeutics and vaccine targets for *Cryptosporidium*. Lancet Infect. Dis., 15: 85-94.
3. Ying, X., X. Chen, Y. Wang, W. Fang and D. Luo, 2009. Signal sequence analysis of protein sequences from the filarial nematode parasite *Brugia malayi* and the evolution of secreted proteins in parasites. Parasitol. Res., 104: 1321-1326.
4. Crowe, J., F.E. Lumb, M.M. Harnett and W. Harnett, 2017. Parasite excretory secretory products and their effects on metabolic syndrome. Parasite Immunol., Vol. 39, No. 5. 10.1111/pim.12410.
5. Marcilla, A., M. Trelis, A. Cortes, J. Sotillo and F. Cantalapiedra *et al*., 2012. Extracellular vesicles from parasitic helminths contain specific excretory/secretory proteins and are internalized in intestinal host cells. PloS One, Vol. 7, No. 9. 10.1371/journal.pone.0045974.
6. Hewitson, J.P., J.R. Grainger and R.M. Maizels, 2009. Helminth immunoregulation: The role of parasite secreted proteins in modulating host immunity. Mol. Biochem. Parasitol., 167: 1-11.
7. Coakley, G., A.H. Buck and R.M. Maizels, 2016. Host parasite communications-messages from helminths for the immune system: Parasite communication and cell-cell interactions. Mol. Biochem. Parasitol., 208: 33-40.
8. Jimenez, J.C., J. Fontaine, C. Creusy, L. Fleurisse, J.M. Grzych, M. Capron and E. Dei-Cas, 2014. Antibody and cytokine responses to *Giardia* excretory/secretory proteins in *Giardia intestinalis*-infected BALB/c mice. Parasitol. Res., 113: 2709-2718.
9. Gonzalez-Paez, G.E., F. Alba-Hurtado, C.G. Garcia-Tovar and R. Arguello-Garcia, 2014. Proteinases in excretory-secretory products of *Toxocara canis* second-stage larvae: Zymography and modeling insights. BioMed. Res. Int., Vol. 2014. 10.1155/2014/418708.
10. Ganga, G., J.P. Varshney and R.C. Patra, 2007. Activity of antioxidant enzymes in excretory-secretory fluid and somatic extracts of *Fasciola gigantic*. J. Vet. Parasitol., 21: 51-52.
11. Gomez, S., L. Adalid-Peralta, H. Palafox-Fonseca, V.A. Cantu-Robles and X. Soberon *et al*., 2015. Genome analysis of excretory/secretory proteins in *Taenia solium* reveals their Abundance of Antigenic Regions (AAR). Scient. Rep., Vol. 5. 10.1038/srep09683.
12. Montaner, S., A. Galiano, M. Trelis, L. Martin-Jaular, H.A. del Portillo, D. Bernal and A. Marcilla, 2014. The role of extracellular vesicles in modulating the host immune response during parasitic infections. Front. Immunol., Vol. 5. 10.3389/fimmu.2014.00433.
13. Liao, Q., X. Yuan, H. Xiao, C. Liu, Z. Lv, Y. Zhao and Z. Wu, 2011. Identifying *Schistosoma japonicum* excretory/secretory proteins and their interactions with host immune system. PloS One, Vol. 6, No. 8. 10.1371/journal.pone.0023786.

14. Klee, E.W. and L.B. Ellis, 2005. Evaluating eukaryotic secreted protein prediction. BMC Bioinform., Vol. 6, No. 1. 10.1186/1471-2105-6-256.
15. Klee, E.W. and C.P. Sosa, 2007. Computational classification of classically secreted proteins. Drug Discov. Today, 12: 234-240.
16. Petersen, T.N., S. Brunak, G. von Heijne and H. Nielsen, 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods, 8: 785-786.
17. Abrahamsen, M.S., T.J. Templeton, S. Enomoto, J.E. Abrahante and G. Zhu *et al*., 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. Science, 304: 441-445.
18. Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis and T.L. Madden, 2008. NCBI BLAST: A better web interface. Nucl. Acids Res., 36: W5-W9.
19. Madden, T., 2013. The BLAST Sequence Analysis Tool. In: The NCBI Handbook, 2nd Edn., McEntyre, J. and J. Ostell (Eds.). National Center for Biotechnology Information, Bethesda MD, USA.
20. Pearson, W.R., 2013. An introduction to sequence similarity ("Homology") searching. Curr. Protoc. Bioinform., 42: 3.1.1-3.1.8.
21. Coletta, A., J.W. Pinney, D.Y.W. Solis, J. Marsh, S.R. Pettifer and T.K. Attwood, 2010. Low-complexity regions within protein sequences have position-dependent roles. BMC Syst. Biol., Vol. 4, No. 1. 10.1186/1752-0509-4-43
22. Nielsen, H., 2017. Predicting Secretory Proteins with SignalP. In: Protein Function Prediction, Kihara, D. (Ed.). Methods in Molecular Biology, Vol. 1611. Humana Press, New York, pp: 59-73.
23. Mousavi, P., Z. Mostafavi-Pour, M.H. Morowvat, N. Nezafat, M. Zamani, A. Berenjian and Y. Ghasemi, 2017. *In silico* analysis of several signal peptides for the excretory production of reteplase in *Escherichia coli*. Curr. Proteomics, 14: 326-335.
24. Wang, T., J. Li, Y. Yang and T. Zhao, 2016. Analysis of coding region for proteins containing signal peptides of *Pseudomonas syringae* pv. tabaci yuexi-1 strain. Acta Tabacaria Sin., 22: 92-100.
25. Kumari, S. and A.K. Chaurasia, 2016. *In silico* analysis and experimental validation of lipoprotein and novel Tat signal peptides processing in *Anabaena* sp. PCC7120. J. Microbiol., 54: 837-846.
26. Massahi, A. and P. Calik, 2015. *In-silico* determination of *Pichia pastoris* signal peptides for extracellular recombinant protein production. J. Theor. Biol., 364: 179-188.
27. Massahi, A. and P. Calik, 2016. Endogenous signal peptides in recombinant protein production by *Pichia pastoris*: From *in-silico* analysis to fermentation. J. Theor. Biol., 408: 22-33.
28. Sun, P.N., X.L. Zhou and Z.X. Wang, 2009. Application of bioinformatics analysis of signal peptide in the identification of *Neurospora crassa* phyA gene. J. Southern Med. Univ., 29: 1098-1101.
29. Huang, J.L., J.Z. Wu, C.G. Xiao, C.J. Li and G.X. Wang, 2007. Analysis of signal peptides of the secreted proteins in *Ralstonia solanacearum* GMI1000. Hereditas, 29: 1409-1416.
30. Wu, H.Z., C.Y. Li, Y.Y. Zhu and Y.F. Bi, 2006. Computational analysis of signal peptide-dependent secreted protein in *Caenorthaditis elegans* ws123. Hereditas, 28: 470-478.
31. Yang, J., C.Y. Li, Y.Y. Wang, Y.Y. Zhu and J.B. Li *et al*., 2006. Computational analysis of signal peptide-dependent secreted proteins in *Saccharomyces cerevisiae*. Agric. Sci. China, 5: 221-227.
32. Pearson, M.S., D.P. McManus, D.J. Smyth, F.A. Lewis and A. Loukas, 2005. *In vitro* and *in silico* analysis of signal peptides from the human blood fluke, *Schistosoma mansoni*. FEMS Immunol. Med. Microbiol., 45: 201-211.
33. Fan, C.M., C.Y. Li, M.F. Zhao and Y.Q. He, 2005. Analysis of signal peptides of the secreted proteins in *Agrobacterium tumefaciens* C58. Acta Microbiol. Sin., 45: 561-566.
34. Choo, K.H., T.W. Tan and S. Ranganathan, 2009. A comprehensive assessment of N-terminal signal peptides prediction methods. BMC Bioinform., 10(Suppl. 15): S2-S2.
35. Bendtsen, J.D., L.J. Jensen, N. Blom, G. von Heijne and S. Brunak, 2004. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng. Design Select., 17: 349-356.
36. Leversen, N.A., G.A. de Souza, H. Malen, S. Prasad, I. Jonassen and H.G. Wiker, 2009. Evaluation of signal peptide prediction algorithms for identification of mycobacterial signal peptides using sequence data from proteomic methods. Microbiology, 155: 2375-2383.
37. Karkowska Kuleta, J. and A. Kozik, 2014. Moonlighting proteins as virulence factors of pathogenic fungi, parasitic protozoa and multicellular parasites. Mol. Oral Microbiol., 29: 270-283.
38. Bouzid, M., P.R. Hunter, R.M. Chalmers and K.M. Tyler, 2013. *Cryptosporidium* pathogenicity and virulence. Clin. Microbiol. Rev., 26: 115-134.
39. Sturbaum, G.D., B.H. Jost and C.R. Sterling, 2003. Nucleotide changes within three *Cryptosporidium parvum* surface protein encoding genes differentiate genotype I from genotype II isolates. Mol. Biochem. Parasitol., 128: 87-90.
40. O'Hara, S.P., J.R. Yu and J.J.C. Lin, 2004. A novel *Cryptosporidium parvum* antigen, CP2, preferentially associates with membranous structures. Parasitol. Res., 92: 317-327.
41. Zhang, Z. and W.J. Henzel, 2004. Signal peptide prediction based on analysis of experimentally verified cleavage sites. Protein Sci., 13: 2819-2824.