Archives of Razi Institute, Vol. 78, No. 1 (2023) 323-335

DOI: 10.22092/ARI.2022.359082.2367

Original Article

In silico Analysis of a 29 kDa *Echinococcus granulosus* Protoscolex Protein (P29) as a Vaccine Candidate against Cystic Echinococcosis

Khazaei, S[1], Dalimi, A[1 *], Pirestani, M[1], Ghafarifar, F[1]

1. Department of Parasitology, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran

Received 5 June 2022; Accepted 29 June 2022
Corresponding Author: dalimi_a@modares.ac.ir

Abstract

Vaccination can be a key step in controlling hydatid cyst infection in humans and livestock in endemic areas of the disease. The aim of the Present study was to determine some of the basal biochemical properties followed by prediction and screening of B-cell and MHC-binding epitopes of EgP29 protein *in silico*. Some of the basic physico-chemical properties along with antigenicity, allergenicity, solubility, post-translational modification (PTM) sites, subcellular localization, signal peptide, transmembrane domain, secondary and tertiary structures followed by refinement and validations were computationally determined for this protein. Also, B-cell epitopes were predicted and screened using various web servers, while MHC-binding and CTL epitopes were predicted using IEDB and NetCTL servers, respectively. The protein is a 238-residue, 27 kDa molecule, with high thermotolerance (aliphatic: 71.81) and hydrophilicity (negative GRAVY). There were several glycosylation and phosphorylation sites in the sequence, without a transmembrane domain and signal peptide. Moreover, several B-cell and MHC-binding epitopes were found in the EgP29 protein, which could be further used in multi-epitope vaccines. In conclusion, results of the present study can be a promising sign for achieving effective approaches to the preparation of a multi-epitope vaccines against echinococcosis. So, it is necessary that the effectiveness of the protein and its epitopes be evaluated *in vitro* and *in vivo*.

Keywords: Hydatid cyst, *Echinococcus granulosus*, P29 protein, Epitopes, *In silico*

1. Introduction

One of the major zoonotic human infections is cystic echinococcosis (CE), caused by the larval stages (metacestodes) of the tapeworm, *Echinococcus granulosus sensu lato* (*E. granulosus s.l.*), inflicting a substantial burden to the public health (1). The life cycle is primarily based on the presence of canid species, in particular dogs, in a given area as the principal definitive hosts, while ungulates are the major intermediate hosts. Notably, humans are considered as dead-end hosts in the life cycle. Upon ingestion of egg-contaminated food, a gradually-growing hydatid cyst would develop mostly within liver and lungs of affected animals and/or humans, which is the major sequela of CE. Reportedly, the incidence of the infection is pronounced in under-developed nations and home slaughtering, dwelling in rural areas and feeding dogs with infected viscera are the major contributing factors (2). Chronically affected human population may suffer from disability and impotency due to CE (2).

Critical control measures against CE stem from the adequate health education, promoted hygiene in slaughterhouses and dosing dogs (3). Indisputably, durable therapies or under-dosing may entail drug resistance and/or drug residues, engendering health problems. Vaccines seem to be a suitable alternative for

aforementioned measures to tackle the infection (3). Thus far, there have been several immunization studies based on different platforms and vaccine candidates against *E. granulosus s.l.* infection, which have shown variable clinical outcomes (4). There have been pilot and field-based trials of vaccination in China and South America using the oncospheral antigen, EG95, demonstrating high rate of prophylactic effects on the CE transmission in sheep (5). Protoscolices are a major source of parasitic antigens to be employed as potential vaccine candidates (6). One of these molecules is a 29-kDa protein, initially characterized by Gonzalez, Spinelli (7) over two decades ago. This protein has, also, been identified in the cyst germinal layer, whereas it is absent in extracts of adult worms (8). The protein may play a crucial role in the host-parasite interactions and being used for diagnostic or vaccination purposes (8).

Prediction of immunogenic epitopes is an essential step in next-generation vaccine design. This is facilitated through computational modelling programs or web servers (9). Previously, several studies have utilized *in silico* methods to demonstrate immunodominant fragments in proteins with parasitic origin from *E. granulosus s.l.* (10, 11), *Leishmania major* (12), *Plasmodium falciparum* (13), *Toxoplasma gondii* (14, 15). The present *in silico* study targeted the P29 antigen of *E. granulosus s.l.* (EgP29) to further characterize its immunogenic regions regarding B-cells and major histocompatibility complex (MHC) molecules for future vaccination studies.

2. Materials and Methods

2.1. Amino Acid Sequence Retrieval

The amino acid sequence of EgP29 protein was obtained as FASTA format via the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/), under accession number AHA85390.1.

2.2. Evaluation of Physico-Chemical Properties of P29 Protein

Preliminary physico-chemical features of the protein, including the grand average of hydropathicity (GRAVY), instability and aliphatic indices, extinction coefficient, estimated half-life, isoelectric point (pI), total number of negatively- and positively-charged residues and molecular weight (MW) were estimated using computational algorithms through the ProtParam online server (https://web.expasy.org/protparam/).

2.3. Prediction of Antigenicity, Allergenicity and Solubility

For antigenicity evaluation, VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) and ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) web tools were utilized, while AllergenFP v1.0 (https://ddg-pharmfac.net/AllergenFP/) and AllerTOPv2.0 (https://www.ddg-pharmfac.net/AllerTOP/) were employed for prediction of allergenicity. VaxiJen performs a target-based prediction with an accuracy of 70-89% (threshold: 0.5), whereas ANTIGENpro predicts relied on auto cross covariance (ACC) transformation of protein sequences into identical vectors of remarkable amino acid features. The solubility of EgP29 protein was, also, predicted using Protein-Sol web server (https://protein-sol.manchester.ac.uk/) and values greater than the population average for the experimental dataset (0.45) are considered as soluble molecules.

2.4. Deciphering Post-Translational Modification (PTM) Sites

The presence of several putative PTM sites were computationally analyzed in the examined protein sequence using online servers such as CSS-Palm (http://csspalm.biocuckoo.org/), GPS-PAIL 2.0 (http://pail.biocuckoo.org/), NetNGlyc (https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0), NetOGlyc (https://services.healthtech.dtu.dk/service.php?NetOGlyc-4.0) and NetPhos (https://services.healthtech.dtu.dk/service.php?NetPhos-3.1) in order to predict palmitoylation, acetylation, N- and O-glycosylation as well as phosphorylation sites.

2.5. Prediction of Transmembrane Domain, Signal Peptide and Subcellular Localization

The presence of a putative signal peptide in the protein sequence was evaluated using SignalP and TargetP web servers, available at https://services.healthtech.dtu.dk/. Moreover, the presence of transmembrane domain and subcellular localization were forecasted using Deep TMHMM and DeepLoc tools using the same URL.

2.6. Secondary structure prediction

The secondary structure of the designed vaccine was predicted using NetSurfP-2.0 web tool, available at https://services.healthtech.dtu.dk/service.php?NetSurfP-2.0. This server predicts the surface accessibility, secondary structure, disorder, and phi/psi dihedral angles of amino acids in a given protein sequence. A single model, using a combination of Convolutional and Bi-Directional Long-Short Term Memory Neural Networks, predicts all structural features together.

2.7. Homology modelling, refinement and validation

Prediction of the three-dimensional (3D) model of EgP29 protein was done using I-TASSER server and enhanced through implementation of an array of refinement approaches by GalaxyRefine server, available at https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE; the server provides five refined models, sorted based on GDT-HA, RMSD, MolProbity, Clash score, Poor rotamers and Rama favored. Subsequently, two web tools, ERRAT and PROCHECK available at https://saves.mbi.ucla.edu/ were utilized to validate the quality of the finally refined model.

2.8. Linear and conformational B-cell epitope prediction

Epitopes corresponding to the B-cells, including continuous and conformation ones, were forecasted using a variety of web servers. Linear epitopes were predicted using B-cell tool of the immune epitope database (IEDB) along with hydrophilicity, antigenicity, surface accessibility, beta-turn and flexibility (http://tools.iedb.org/bcell/), BCPREDS that employs artificial neural network (ANN) (http://crdd.osdd.net/raghava/abcpred/) as well as SVMTriP server based on support vector machine (SVM) combined with Tri-peptide similarity and Propensity scores (http://crdd.osdd.net/raghava/bcepred/). Moreover, conformational B-cell epitopes of the protein were predicted using ElliPro tool of the IEDB online server, available at http://tools.iedb.org/ellipro/.

2.9. Prediction of mouse and human MHC-binding epitopes

For this aim, those residues with higher affinity (lower percentile ranks) to bind with mouse and human MHC-I (http://tools.iedb.org/mhci/) and MHC-II (http://tools.iedb.org/mhcii/) alleles were predicted using IEDB server. Accordingly, eight mouse MHC-I alleles (H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Kk, H2-Ld, H2-Qa1 and H2-Qa2) and three mouse MHC-II alleles (H2-IAb, H2-IAd and H2-IEd) were used. In addition, HLAA*02:01, HLA-A*24:02 (MHC-I), DRB1*01:02 and DQA1*05:01/DQB1*03:01 (MHC-II) were targeted as human alleles. For all predictions, 12-mer and 15-mer epitopes were forecasted using IEDB recommended 2020.09 (NetMHCpan EL 4.1) and IEDB recommended 2.22 method, respectively. All high-rated epitopes were then screened regarding antigenicity and allergenicity using VaxiJen v2.0 and AllergenFP v1.0 servers, respectively.

3. Results

3.1. Physico-Chemical, Solubility, Antigenicity and Allergenicity Properties of EgP29 Protein

This 238 amino acid protein of *E. granulosus s.l.* had a MW of 27096.59 kDa and most of its residues were negatively-charged (Asp + Glu) (41), while 37 residues were positively-charged (Arg + Lys). The speculated pI of the protein was calculated to be 5.63 and total number of atoms were 3805. The estimated half-life of

EgP29 was 30 h, >20 h and >10 h in mammalian reticulocytes, yeast and *Escherichia coli*, respectively. The protein was demonstrated to be stable (instability index: 35.16), highly thermotolerant (71.81) and hydrophilic (GRAVY: -0.697) in nature. Attributable to the Protein-Sol output, this protein was extremely soluble with predicted scaled solubility of 0.782 (threshold: 0.45) (Figure 1). Although the protein was not antigenic, as indicated by VaxiJen v2.0 server (0.4194), it was shown to be antigenic by ANTIGENpro server (0.951959). Of note, no allergenic traits were predicted for the protein, as substantiated by AllergenFP and AllerTOP web servers.

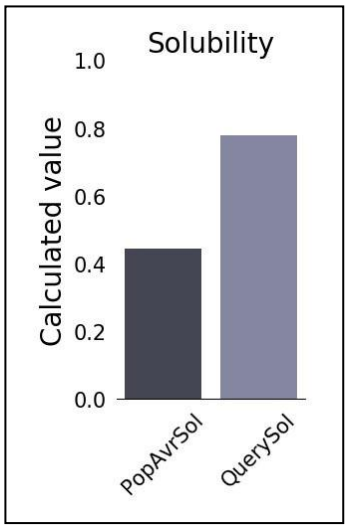

Figure 1. Calculated solubility of the EgP29 protein using Protein-Sol web server

3.2. Prediction of PTM Sites, Signal Peptide, Transmembrane Domain and Subcellular Localization

Although no transmembrane domain was predicted within protein sequence, it constituted of several PTM sites, including 10 acetylation, 1 palmitoylation, 2 N-glycosylation, 5 O-glycosylation and 26 phosphorylation (12 tyrosine, 11 serine and 3 tyrosine) regions (Figure 2). Also, based on the output of two servers (TargetP and SignalP), no signal peptide was predicted for EgP29. Based on DeepLoc server prediction, this protein was allocated to the cytoplasm with a probability of 0.5221 (threshold: 0.4761).

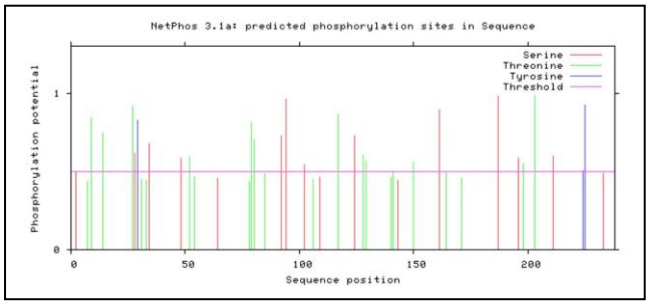

Figure 2. Prediction of phosphorylation sites of the EgP29 protein using NetPhos web server suggested 26 putative regions

3.3. Secondary Structure Prediction

Pertinent to the secondary structure prediction using NetSurfP-2.0 server, most of the residues were exposed and the predicted structures were helices followed by random coils across the sequence. It is, also, noticeable that the first 27 amino acid residues of the EgP9 protein were disordered. Figure 3 illustrates in details the secondary structures predicted by this server.

3.4. 3D Homology Modelling, Rehashing and Validation

For modelling the 3D structure of the protein, we used a powerful web server, I-TASSER. Based on the results, top-ten threading templates with highest significance were chosen from LOMETS database, among which template 6 (sequence identity: 0.18, coverage: 0.91) was used to generate the 3D model. Five models were accordingly predicted and model number 1 had the lowest C-score (higher prediction confidence) (-0.07), with estimated TM-score of 0.7±0.12 and estimated RMSD of 5.8±3.7 Å (Figure 4).

This model was further subjected to refining, using GalaxyRefine web server. Based on its output, model number 1, having GDT-HA of 0.9779, RMSD of 0.321, MolProbity of 1.835, Clash score of 8.5, Poor rotamers of 0.9 and Rama favored of 94.5, was selected as the best refined model. Comparison of structural improvements between crude and refined model was done using ERRAT and PROCHECK tools. Based on ERRAT, the overall quality factor of the crude and refined models were 96.957 and 100.000, respectively.

Moreover, Ramachandran plot analysis of the refined model demonstrated improvements in this model, with 211 (92.1%), 13 (5.7%), 2 (0.9%) and 3 (1.3%) of the residues allocated to the most favored, additional allowed, generously allowed and disallowed regions, respectively (Figure 5).

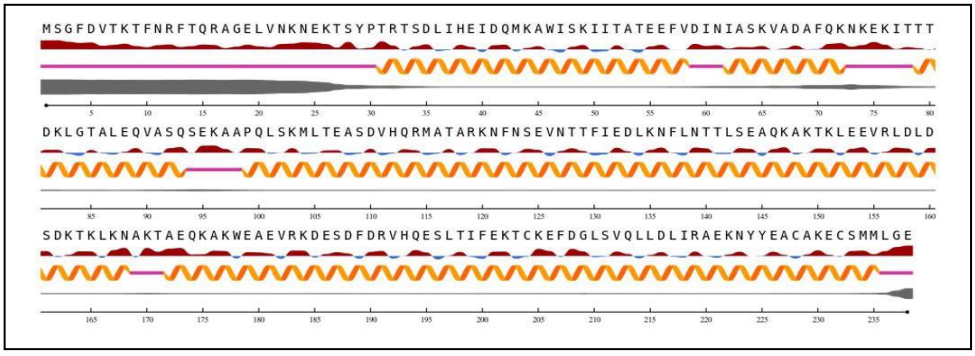

Figure 3. Secondary structure prediction using NetsurfP-2.0 web server

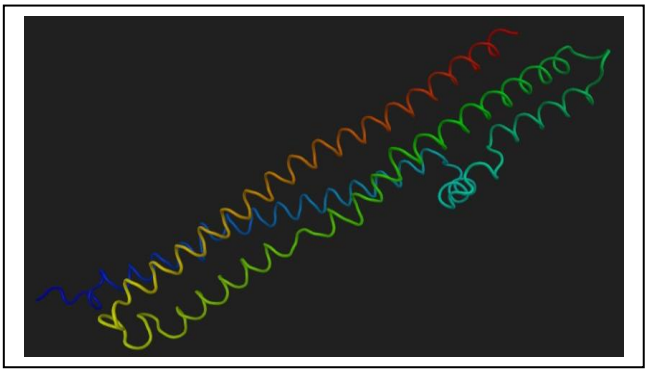

Figure 4. 3D model of the EgP29 protein predicted using ITASSER web server

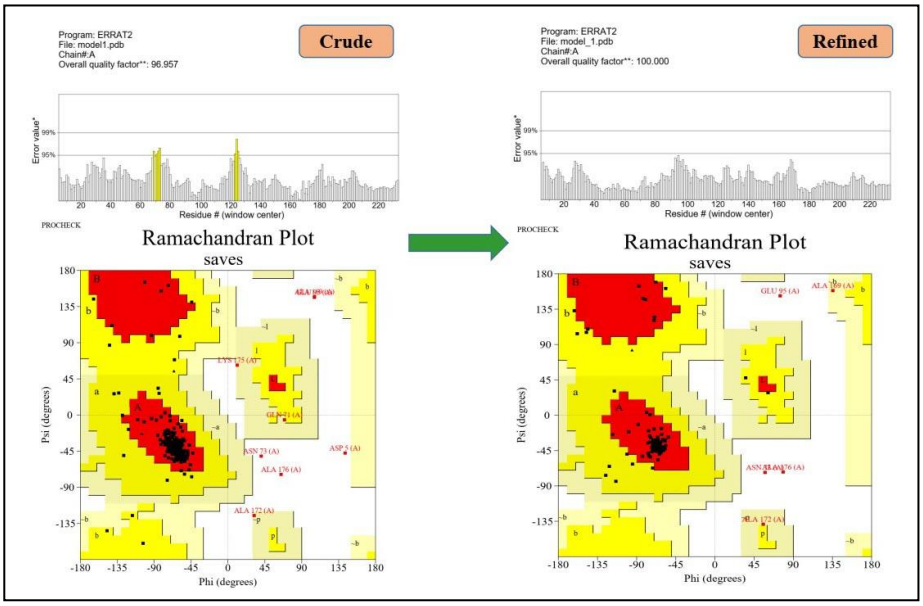

Figure 5. Evaluation of the refined model in comparison with the crude model showed improvements as indicated by ERRAT and PROCHECK web tools

3.5. Continuous and Conformational B-Cell Epitope Prediction

The BCPREDS server demonstrated 8 potential linear B-cell epitopes within the EgP29 sequence using a prediction threshold of 75% (Table 1). In addition, top 10 continuous B-cell epitopes were forecasted by SVMTriP servers, among which "QLSKMLTEASDVHQ" (100-113) possessed the highest score (1.000) and was recommended by this server (Table 2).

Moreover, B-cell tool of the IEDB server predicted B-cell epitopes regarding Chou and Fasman beta-turn, Emini surface accessibility, Karplus & Schulz flexibility, Kolaskar & Tongaonkar antigenicity, Parker hydrophilicity and BepiPred linear epitopes with averages scores of 0.933, 1.000, 1.011, 1.001, 2.387 and 1.005, respectively (Figure 6). Five conformational B-cell epitopes, predicted by ElliPro tool of IEDB, were including: 1) 42 residues, score: 0.781; 2) 16 residues, score: 0.773; 3) 24 residues, score: 0.734; 4) 4 residues, score: 0.685; and 5) 19 residues, score: 0.677 (Figure 7).

3.6. Prediction of Mouse and Human MHC-Binding Epitopes

Several mouse MHC alleles were employed to predict potential MHC binders regarding EgP29 protein in this study. Accordingly, top 3 high-ranked (lower percentile rank) epitopes were selected and further screened in terms of antigenicity and allergenicity. The obtained results are provided in details in table 3. Similar procedures were, also, accomplished in case of prediction of human MHC binders, showing 6 potential epitopes with antigenicity scores in parentheses within the examined protein sequence, including "VRLDLDSDKTKL" (0.8514), "YPTRTSDLIHEI" (0.5052), "VNTTFIEDLKNF" (0.7472), "NYYEACAKECSM" (0.5245), "YYEACAKECSMM" (0.5555) and "EEFVDINIASKVADA" (0.6286) (Table 4).

Table 1. Prediction of linear B-cell epitopes (14-mer) in EgP29 protein using BCPREDS server (Threshold: 75%)

Start Position	Epitope Sequence	Score
72	KNKEKITTTDKLGT	0.999
22	NKNEKTSYPTRTSD	0.982
181	EVRKDESDFDRVHQ	0.933
88	EQVASQSEKAAPQL	0.9
6	VTKTFNRFTQRAGE	0.854
116	ATARKNFNSEVNTT	0.847
219	RAEKNYYEACAKEC	0.722
133	DLKNFLNTTLSEAQ	0.718

Table 2. Specific linear B-cell epitopes (14-mer) of EgP29 predicted by the SVMTriP web server

Rank	Position	Epitope	Score	Recommended
1	100-113	QLSKMLTEASDVHQ	1.000	
2	31-44	TRTSDLIHEIDQMK	0.780	
3	210-223	LSVQLLDLIRAEKN	0.525	
4	180-193	AEVRKDESDFDRVH	0.409	
5	150-163	TKLEEVRLDLDSDK	0.366	
6	79-92	TTDKLGTALEQVAS	0.366	
7	47-60	ISKIITATEEFVDI	0.351	
8	123-136	NSEVNTTFIEDLKN	0.279	
9	62-75	IASKVADAFQKNKE	0.263	
10	1-14	MSGFDVTKTFNRFT	0.241	

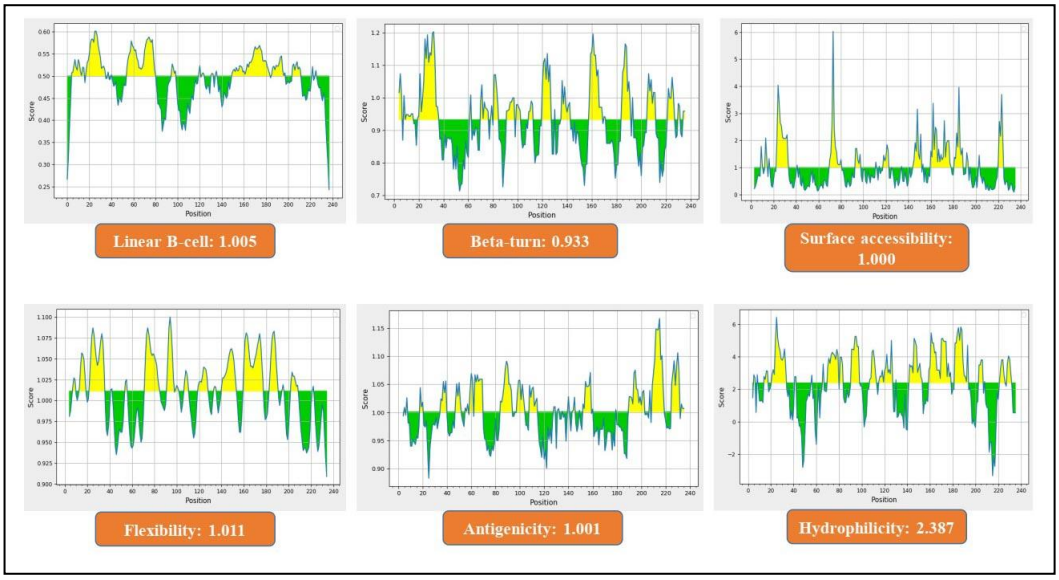

Figure 6. Predicted linear B-cell epitopes regarding antigenicity, BepiPred, beta-turn, flexibility, hydrophilicity and surface accessibility using B-cell online tool of the IEDB server

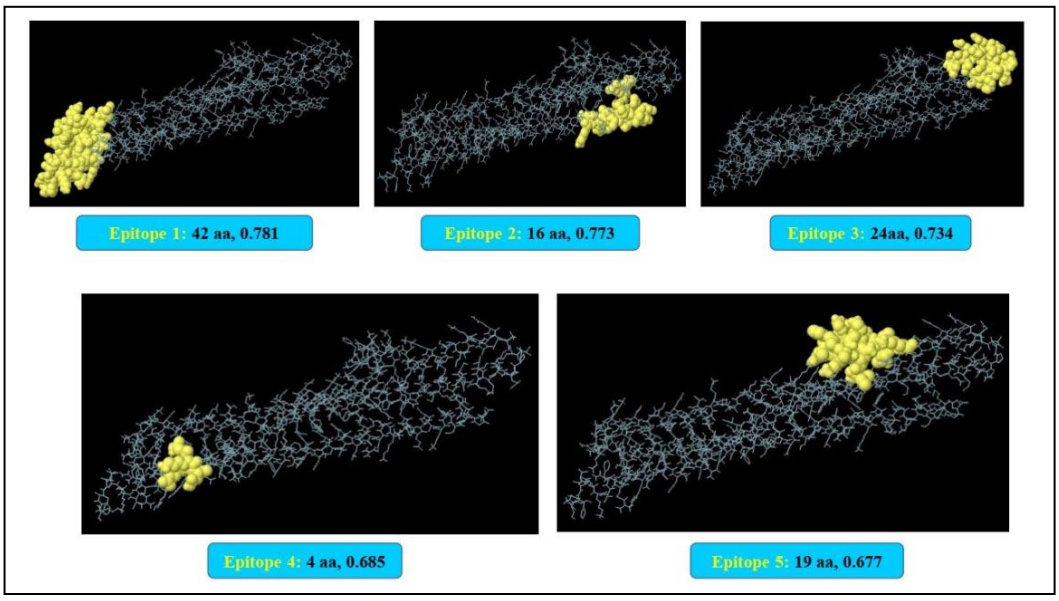

Figure 7. Predicted conformational B-cell epitopes of the EgP29 protein using ElliPro tool of IEDB server

Table 3. Prediction of the high-ranked epitopes of EgP29 with specific binding capacity for mouse MHC molecules using IEDB server with subsequent screening regarding antigenicity and allergenicity

Mouse MHC alleles	Peptide sequence	Start-End	Percentile rank	Antigenicity	Allergenicity
H2-Db (MHC-I)	KNFNSEVNTTFI	50-61	0.69	1.3266	No
	NSEVNTTFIEDL	53-64	2.9	1.2469	No
	AAPQLSKMLTEA	27-38	3.3	0.2095	Yes
H2-Dd (MHC-I)	SEKAAPQLSKML	24-35	2.0	0.0843	No
	SGFDVTKTFNRF	2-13	2.3	-1.4030	No
	RKNFNSEVNTTF	49-60	4.5	1.0987	No
H2-Kb (MHC-I)	SGFDVTKTFNRF	2-13	1.9	-1.4030	No
	KNKEKITTTDKL	50-61	8.4	0.5802	No
	SEKAAPQLSKML	24-35	11.0	0.0843	No
H2-Kd (MHC-I)	NYYEACAKECSM	13-24	3.9	0.5245	No
	FDRVHQESLTIF	49-60	5.4	-0.6716	No
	YYEACAKECSMM	14-25	6.7	0.5555	No
H2-Kk (MHC-I)	EEFVDINIASKV	55-66	1.2	0.6304	No
	SEKAAPQLSKML	24-35	1.2	0.0843	No
	YPTRTSDLIHEI	29-40	2.7	0.5052	No
H2-Ld (MHC-I)	YPTRTSDLIHEI	29-40	0.48	0.5052	No
	KITTTDKLGTAL	6-17	4.2	0.1480	No
	SGFDVTKTFNRF	2-13	4.9	-1.4030	No
H2-Qa1 (MHC-I)	RTSDLIHEIDQM	32-43	2.2	-0.3976	No
	SGFDVTKTFNRF	2-13	5.6	-1.4030	No
	QSEKAAPQLSKM	27-38	6.9	0.0135	No
H2-Qa2 (MHC-I)	SEKAAPQLSKML	24-35	0.7	0.0843	No
	DLIHEIDQMKAW	35-46	3.1	-0.8833	No
	EEFVDINIASKV	55-66	3.1	0.6304	No
H2-IAb (MHC-II)	ASQSEKAAPQLSKML	21-35	9.7	0.3236	No
	SQSEKAAPQLSKMLT	22-36	10.15	0.1879	No
	QSEKAAPQLSKMLTE	23-37	12.6	0.2334	No
H2-IAd (MHC-II)	EIDQMKAWISKIITA	39-53	2.31	0.6187	No
	IDQMKAWISKIITAT	40-54	2.38	0.8088	No
	HEIDQMKAWISKIIT	38-52	2.52	0.4094	No
H2-IEd (MHC-II)	EQKAKWEAEVRKDES	33-47	3.3	1.1510	No
	AEQKAKWEAEVRKDE	32-46	3.85	1.1873	No
	QKAKWEAEVRKDESD	34-48	4.8	1.2224	No

Table 4. Prediction of the high-ranked epitopes of EgP29 with specific binding capacity for human MHC molecules using IEDB server with subsequent screening regarding antigenicity and allergenicity

Human leukocyte antigen alleles	Peptide sequences	Start-End	Percentile rank	VaxiJen Antigenicity	AllergenFP Allergenicity
HLA-A*02:01	TLSEAQKAKTKL	1-12	3.4	0.6951	Yes
	QMKAWISKIITA	42-53	4.0	1.5841	Yes
	TTDKLGTALEQV	9-20	4.7	0.1268	Yes
	VRLDLDSDKTKL	15-26	4.9	0.8514	No
	YPTRTSDLIHEI	29-40	5.3	0.5052	No
HLA-A*24:02	VNTTFIEDLKNF	56-67	1.2	0.7472	No
	FDRVHQESLTIF	49-60	3.2	-0.6716	No
	NYYEACAKECSM	13-24	4.1	0.5245	No
	YYEACAKECSMM	14-25	4.4	0.5555	No
	SGFDVTKTFNRF	2-13	5.4	-1.4030	No
DRB1*01:02	SVQLLDLIRAEKNYY	1-15	7.4	-0.0551	No
	VQLLDLIRAEKNYYE	2-16	7.4	0.0590	No
	ATEEFVDINIASKVA	53-67	26.0	0.2734	No
	EEFVDINIASKVADA	55-69	26.0	0.6286	No
	EFVDINIASKVADAF	56-70	26.0	0.2046	No
DQA1*05:01/DQB1*03:01	EFVDINIASKVADAF	56-70	14.0	0.2046	No
	QVASQSEKAAPQLSK	19-33	15.0	0.2377	No
	VASQSEKAAPQLSKM	20-34	16.0	0.1757	No
	ASQSEKAAPQLSKML	21-35	17.0	0.3236	No
	MKAWISKIITATEEF	43-57	18.0	1.0687	Yes

4. Discussion

The incidence and public health significance of CE has long been emphasized, particularly in those areas that sheep raising is a traditionally-established practice (3). Accumulating evidence demonstrates that G1 is the primary and most abundant genotype across the globe, found in different animal hosts and human populations, hence the sheep-dog cycle must be primarily targeted for prevention purposes (16). This could be afforded via safe animal slaughtering, treatment of owned dogs (*i.e.*, shelter, pet, etc.) on a routine basis and vaccination practices to cut the life cycle and decrease the chance of disease transmission in both intermediate and definitive hosts. In previous investigations, a considerable number of vaccine candidates has been introduced and examined regarding safety and protective efficacy against *E. granulosus s.l.* infection. Conventional method of vaccine generation is a laborious and costly procedure, and deeply requires hard-working and passing through different experimental and clinical phases. This is unexceptionally more complicated in case of parasitic agents having complex life cycles, including *E. granulosus s.l.* By adjoining computational methods and biological data, we can utilize an easier framework for designing potent vaccine candidates to promote vaccine effectiveness, hence establishing a profitable pipeline of vaccine Research & Development (R&D) in shade of immunoinformatics. In this sense, progressively-developing machine-learning methods and algorithms accelerate to attain such an encouraging goal (17).

Although CE is known as an ancient disease, the underlying molecular mechanisms of immune responses elicited against the parasite have been characterized during last decades using unprecedented Omics-based technologies. A substantial increase in IgE, IgM and IgG (particularly IgG1 and IgG4 subtypes) accompanies an established CE infection (18), whereas a dichotomy is observed in cell-mediated immune responses. This dichotomy is represented by Th1 dominance in early infection favoring host through

inhibiting the parasite growth, while Th2 inversely act and any imbalance in both responses may entail immunopathogenesis. This is, also, reflected and should be considered in the vaccine R&D against CE, the procedure which could be accurately enhanced through comprehensive *in silico* methods (19). In 2000, EgP29 protein was first discovered as a novel 29 kDa antigen in search of alternative antigens in the hydatid cyst fluid (7). In the following, the protein was identified in the protoscolex-derived soluble somatic antigen of G1 genotype, being rendered as a biomarker to monitor CE patients (20). Recently, it was demonstrated that recombinant EgP29 protein possesses desirable protection against CE infection in sheep (8) and mouse experimental models (21). There is no information regarding potential immunodominant epitopes in EgP29 protein thus far. The purpose of the present study was to predict some of the preliminary physico-chemical characteristics of the EgP29 protein and to highlight some of the potential immunogenic epitopes with higher affinity for B-cells as well as mouse and human MHC molecules.

At first, preliminary physico-chemical features of the protein were predicted using ProtParam server; this 238 amino acid protein was immunogenic, with a molecular weight of over 5-10 kDa (~ 27 kDa). The isoelectric pH (pI) was rendered to be relatively acidic in nature, as substantiated by a pI score of 5.63. Instability index below 40 is considered as a stable protein, such as this protein, showing instability index of 35.16. Negative GRAVY scores (-0.697) indicate to a hydrophilic molecule with enhanced interaction with the surrounding water-based milieu, whereas high aliphatic index (71.81) suggest higher molecular tolerance against wide range of temperatures. A proper vaccine candidate should not be allergenic in nature (22); here, no allergenicity was predicted for EgP29 protein. Moreover, it was rendered as highly antigenic by ANTIGENpro tool, with antigenicity score of 0.951959. Overall, such basic features on the physical, chemical and biological characteristics of the EgP29 protein seem to be essential in designing wet lab experiments and future vaccinology studies. Based on several web servers, no signal peptide and transmembrane domains were present within the protein sequence (23), while several PTM regions existed in the protein. Actually, such modification sites are critical for a number of cellular processes and their prediction seem to be reasonable and advantageous regarding selection of suitable prokaryotic and/or eukaryotic protein expression platforms (24). In the present study, online tools were used to predict phosphorylation, acetylation, palmitoylation and glycosylation sites in the EgP29 protein. Our results indicated that the most abundant PTM sites was phosphorylation, followed by acetylation (5), O-glycosylation (2), N-glycosylation and palmitoylation (1).

In the second step, secondary and tertiary structure of this protein were modelled using machine-learning based approaches developed by NetSurfP-2.0 and I-TASSER web servers. According to NetSurfP server, helices and random coils were predominant secondary structures in the protein of interest and no specific disordered regions were detected throughout the sequence, except in the first 27 residues. Homology modelling is performed in order to predict the conformation and 3D structure of a given molecule. For this aim, some web servers have been introduced and here we used I-TASSER. Although, "A major concern in structural biology is the identification of faults in experimental and theoretical models of protein structures"; hitherto, enhancements in the reliability and quality of the 3D model is highly significant through structural rehashing, performed by the GalaxyRefine server. In current study, model number 1 was selected, based on GDT-HA of 0.9779, RMSD of 0.321, MolProbity of 1.835, Clash score of 8.5, Poor rotamers of 0.9 and Rama favored of 94.5. Comparatively, adequately acceptable results predicted by ERRAT and PROCHECK online tools showed structural improvements between the crude and refined models of EgP29 protein.

The immune responses against CE infection depends on both humoral and cell-mediated arms. In case of

humoral immunity, specific IgG is the earliest detected immunoglobulin in the circulation, particularly those against hydatid cyst fluid (2 weeks) and oncospheral antigens (11 weeks) in challenged mice and sheep (25). Such antibodies play a crucial role in parasite killing and a protective response. Furthermore, cellular immune responses of Th1-type are highly protective against *E. granulosus s.l.* larval stages. A critical finding of present *in silico* study was the prediction and screening of B-cell and MHC-binding epitopes of EgP29 protein using a set of online immunoinformatics tools to increase the prediction accuracy. In this sense, B-cell epitopes were forecasted by BCPREDS, SVMTriP and B-cell tool of IEDB server. With respect to proper antigen-antibody interaction, conformational B-cell epitopes were predicted, indicating 5 non-continuous epitopes within the protein sequence. Moreover, T-cell and MHC interplay are a cornerstone for an appropriate cell-mediate immune response; accordingly, several mouse and human MHC alleles in IEDB server were employed for the prediction of MHC binders. Regarding mouse MHC binding epitopes, 15 candidate peptides were qualified after screening in terms of antigenicity and allergenicity, including KNFNSEVNTTFI, NSEVNTTFIEDL, RKNFNSEVNTTF, KNKEKITTTDKL, NYYEACAKECSM, YYEACAKECSMM, EEFVDINIASKV, YPTRTSDLIHEI, KITTTDKLGTAL, EEFVDINIASKV (MHC-I), EIDQMKAWISKIITA, IDQMKAWISKIITAT, EQKAKWEAEVRKDES, AEQKAKWEAEVRKDE and QKAKWEAEVRKDESD (MHC-II). Moreover, 5 potential human MHC binder peptides were accurately screened, regarding VRLDLDSDKTKL, YPTRTSDLIHEI, NYYEACAKECSM, YYEACAKECSMM (MHC-I) and EEFVDINIASKVADA (MHC-II). Altogether, such qualified epitopes can be targeted alone and/or combined with potential epitopes derived from other vaccine candidates to design, engineer and implement an efficacious multi-epitope vaccine construct directed against CE in future.

5. Conclusion

As a final word, CE is still a neglected and silent threat to both livestock and human population in endemic areas of the Middle East and South America. Attempt to develop a highly protective vaccine against the infection is ongoing, being accelerated by the advent of computer-aided machine-learning based algorithms for immunoinformatics, in order to sense the immunodominant epitopes and construct finely-tuned multi-epitope immunogenic structures. The present *in-silico* study provided some of the basic physico-chemical properties of EgP29 protein, its structural modelling and predicted some of the antigenic, non-allergenic B- and T-cell epitopes for this protein. Undoubtedly, the findings of the present investigation should be validated in experimental settings and would assist researchers for future immunization against CE.

Authors' Contribution

Study concept and design: S. Kh. and A. D.
Acquisition of data: S. Kh.
Analysis and interpretation of data: A. D. and M. P.
Drafting of the manuscript: S. Kh., A. D. and M. P.
Critical revision of the manuscript for important intellectual content: A. D.
Statistical analysis: A. D.
Administrative, technical, and material support: A. D., M. P. and F. Gh.

Ethics

This study was confirmed by the Ethical Committee of Tarbiat Modares University.

Conflict of Interest

The authors declare that they have no conflict of interest.

Grant Support

This study was financially supported by Tarbiat Modares University.

References

1. Khademvatan S, Majidiani H, Foroutan M, Hazrati Tappeh K, Aryamand S, Khalkhali HR. Echinococcus granulosus genotypes in Iran: a systematic review. J Helminthol. 2019;93(2):131-8.
2. Tamarozzi F, Deplazes P, Casulli A. Reinventing the Wheel of Echinococcus granulosus sensu lato Transmission to Humans. Trends Parasitol. 2020;36(5):427-34.
3. Larrieu E, Gavidia CM, Lightowlers MW. Control of cystic echinococcosis: Background and prospects. Zoonoses Public Health. 2019;66(8):889-99.
4. Anvari D, Rezaei F, Ashouri A, Rezaei S, Majidiani H, Pagheh AS, et al. Current situation and future prospects of Echinococcus granulosus vaccine candidates: A systematic review. Transbound Emerg Dis. 2021;68(3):1080-96.
5. Craig PS, Hegglin D, Lightowlers MW, Torgerson PR, Wang Q. Echinococcosis: Control and Prevention. Adv Parasitol. 2017;96:55-158.
6. Carmena D, Benito A, Eraso E. Antigens for the immunodiagnosis of Echinococcus granulosus infection: An update. Acta Trop. 2006;98(1):74-86.
7. Gonzalez G, Spinelli P, Lorenzo C, Hellman U, Nieto A, Willis A, et al. Molecular characterization of P-29, a metacestode-specific component of Echinococcus granulosus which is immunologically related to, but distinct from, antigen 5. Mol Biochem Parasitol. 2000;105(2):177-84.
8. Wang H, Li Z, Gao F, Zhao J, Zhu M, He X, et al. Immunoprotection of recombinant Eg.P29 against Echinococcus granulosus in sheep. Vet Res Commun. 2016;40(2):73-9.
9. Parvizpour S, Pourseif MM, Razmara J, Rafi MA, Omidi Y. Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. Drug Discov Today. 2020;25(6):1034-42.
10. Khazaei S, Moghadamizad Z. Echinococcus granulosus cyclophilin: Immunoinformatics analysis to provide insights into the biochemical properties and immunogenic epitopes. Inform Med Unlocked. 2022;30:100925.
11. Shams M, Javanmardi E, Nosrati MC, Ghasemi E, Shamsinia S, Yousefi A, et al. Bioinformatics features and immunogenic epitopes of Echinococcus granulosus Myophilin as a promising target for vaccination against cystic echinococcosis. Infect Genet Evol. 2021;89:104714.
12. Shams M, Nourmohammadi H, Basati G, Adhami G, Majidiani H, Azizi E. Leishmanolysin gp63: Bioinformatics evidences of immunogenic epitopes in Leishmania major for enhanced vaccine design against zoonotic cutaneous leishmaniasis. Inform Med Unlocked. 2021;24:100626.
13. Asghari A, Nourmohammadi H, Majidiani H, Shariatzadeh SA, Shams M, Montazeri F. In silico analysis and prediction of immunogenic epitopes for pre-erythrocytic proteins of the deadly Plasmodium falciparum. Infect Genet Evol. 2021;93:104985.
14. Majidiani H, Dalimi A, Ghaffarifar F, Pirestani M. Multi-epitope vaccine expressed in Leishmania tarentolae confers protective immunity to Toxoplasma gondii in BALB/c mice. Microb Pathog. 2021;155:104925.
15. Majidiani H, Dalimi A, Ghaffarifar F, Pirestani M, Ghaffari AD. Computational probing of Toxoplasma gondii major surface antigen 1 (SAG1) for enhanced vaccine design against toxoplasmosis. Microb Pathog. 2020;147:104386.
16. Galeh TM, Spotin A, Mahami-Oskouei M, Carmena D, Rahimi MT, Barac A, et al. The seroprevalence rate and population genetic structure of human cystic echinococcosis in the Middle East: A systematic review and meta-analysis. Int J Surg. 2018;51:39-48.
17. Oli AN, Obialor WO, Ifeanyichukwu MO, Odimegwu DC, Okoyeh JN, Emechebe GO, et al. Immunoinformatics and Vaccine Development: An Overview. Immunotargets Ther. 2020;9:13-30.
18. Mourglia-Ettlin G, Miles S, Hernandez A, Dematteis S. Ageing impairs parasite-specific antibody responses in cystic echinococcosis. Parasite Immunol. 2018;40(1).
19. Wen H, Vuitton L, Tuxun T, Li J, Vuitton DA, Zhang W, et al. Echinococcosis: Advances in the 21st Century. Clin Microbiol Rev. 2019;32(2).
20. Ben Nouir N, Nunez S, Gianinazzi C, Gorcii M, Muller N, Nouri A, et al. Assessment of Echinococcus granulosus somatic protoscolex antigens for serological follow-up of young patients surgically treated for cystic echinococcosis. J Clin Microbiol. 2008;46(5):1631-40.

21. Gharibi Z, Rahdar M, Pirestani M, Tavalla M, Tabandeh MR. The Immunization of Protoscolices P29 DNA Vaccine on Experimental Cystic Echinococosis in Balb/c Mice. Acta Parasitol. 2021;66(4):1114-21.
22. Abraham Peele K, Srihansa T, Krupanidhi S, Ayyagari VS, Venkateswarulu TC. Design of multi-epitope vaccine candidate against SARS-CoV-2: a in-silico study. J Biomol Struct Dyn. 2021;39(10):3793-801.
23. Safavi A, Kefayat A, Sotoodehnejadnematalahi F, Salehi M, Modarressi MH. In Silico Analysis of Synaptonemal Complex Protein 1 (SYCP1) and Acrosin Binding Protein (ACRBP) Antigens to Design Novel Multiepitope Peptide Cancer Vaccine Against Breast Cancer. Int J Pept Res Ther. 2019;25(4):1343-59.
24. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. Nat Biotechnol. 2003;21(3):255-61.
25. Zhang W, You H, Li J, Zhang Z, Turson G, Aili H, et al. Immunoglobulin profiles in a murine intermediate host model of resistance for Echinococcus granulosus infection. Parasite Immunol. 2003;25(3):161-8.