Tehran University of Medical Sciences Publication
http://tums.ac.ir

Iran J Parasitol

Open access Journal at
http://ijpa.tums.ac.ir

Iranian Society of Parasitology
http://isp.tums.ac.ir

Original Article

Biochemical Properties and Immunogenic Epitopes of *Echinococcus granulosus* Glutathione S-Transferase as a Vaccine Target: *In-Silico* Study

Sasan Khazaei, *Abdolhossein Dalimi, Majid Pirestani, Fatemeh Ghafarifar

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

Received 22 Sep 2023
Accepted 19 Nov 2023

Keywords:
Echinococcus granulosus;
Glutathione S-transferase;
Bioinformatics;
Vaccine

***Correspondence Email:**
dalimi_a@modares.ac.ir

Abstract

Background: The current *in silico* study was done to determine the primary biochemical features and immunogenic epitopes of *Echinococcus granulosus* glutathione S-transferase protein as a potential vaccine candidate.

Methods: Several web tools were employed to predict physico-chemical properties, antigenicity, allergenicity, solubility, post-translational modification (PTM) sites, subcellular localization, signal peptide, transmembrane domain, secondary and tertiary structure followed by refinement and validations. In addition, B-cell epitopes were predicted and were screened using various web servers, while MHC-binding and CTL epitopes were predicted using IEDB and NetCTL servers, respectively.

Results: The protein had 219 residues with a molecular weight of 25.55 kDa and alkaline isoelectric pH (7.5). This protein was stable, thermotolerant (aliphatic index: 78.04) and hydrophilic (GRAVY: -0.440). The predicted antigenicity scores were low and the protein was non-allergenic in nature. There were no transmembrane domain and signal peptide in the sequence. Moreover, several B-cell, MHC-binding and CTL epitopes were found in the EgGST protein, which could be further used in multi-epitope vaccines.

Conclusion: Further studies are needed on the development of vaccines in vivo using EgGST alone or in combination with other antigens in the future.

Introduction

The larval stages of *Echinococcus granulosus* are the causative agents of a significant zoonosis in humans and livestock, i.e. cystic echinococcosis (CE) or hydatidosis (1). The disease represents remarkable public health concern, mostly in developing countries (2). Canine hosts are the major definitive hosts for *E. granulosus*; within their intestine, gravid proglottids containing thousands of eggs are produced by fertile worms. Hence, fecal matter of infected carnivores is the main source for the environmental contamination and subsequent infection of the herbivorous preys such as ungulates (intermediate hosts) (3). Hydatid cysts are dominantly found in liver and lungs, while other organs may, also, be affected. Human infection may occur only accidentally *via* ingesting parasite ova by contaminated food or water (4). Reportedly, feeding dogs with infected viscera, free-roaming dogs, dog ownership, home slaughtering, low country income and living in rural regions are the main risk factors for human CE, which provokes impotency and disability in affected populations (5).

Dosing dogs, hygienic slaughter in abattoirs and health education are the main control measures taken to obstruct parasite transmission from dogs to livestock and humans (6). Unavoidably, therapeutic options may be associated with several pharmacologic side effects, drug residues in meat and milk of affected livestock and may facilitate the drug resistance concern because of long-term treatment and/or under-dosing (7). Promising advances in vaccination strategies entail higher efficiency to prevent CE, both in definitive and intermediate hosts (8). In this sense, an ideal vaccine would completely prevent the oncospheral maturation and hydatid cyst development in humans and sheep and hampers worm maturation in canid intestine (9). Thus far, a wide number of vaccine antigens have been introduced for *E. granulosus*, with degrees of inconsistency in clinical outcomes (reviewed in (2)). A well-known representative of vaccine candidates is EG95 (oncospheral antigen), causing high-degrees of prophylactic effects on CE transmission in pilot and field trials in China and South America (10). The glutathione S-transferase (GST) activity of *E. granulosus* was initially found within the cytosol of protoscoleces (PSCs) (11). Today, they are known as one of the several immunomodulatory molecules, assisting the worm to survive in immunocompetent hosts (12). Therefore, it deserves to be further explored in-depth using immunoinformatics methods.

The present *in silico* study was performed to determine the biochemical features and the most potent immunogenic epitopes of EgGST as a foundation for future vaccination approaches.

Methods

Amino acid sequence retrieval

The amino acid sequence of EgGST protein was retrieved as FASTA format through the national center for biotechnology information (NCBI) database (https://www.ncbi.nlm.nih.gov/) under accession number of AAD16438.1.

Ethical approval

This study was approved by the Ethical Committee of Tarbiat Modares University, Iran. (IR.MODARES.REC.1397.008).

Assessment of physico-chemical parameters

The ProtParam online server was employed to evaluate some of the most basic but important physico-chemical properties of EgGST, 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).

Prediction of antigenicity, allergenicity and solubility profiles

The antigenicity of protein was analyzed by: ANTIGENpro and VaxiJen v2.0. ANTIGENpro is an online tool of the SCRATCH protein predictor server that is only sequence-based and independent of alignment and pathogen type. VaxiJen v2.0 server performs based on auto cross-covariance (ACC) transformation of protein sequences into uniform vectors of significant amino acid properties. The prediction occurs according to the target organism, varying from 70% to 89%. The allergenicity of the vaccine candidate was predicted using AllergenFP v1.0, AllerTOP v2.0 and AlgPred (Combined approach). A descriptor-based fingerprint platform is used by AllergenFP v1.0 to differ antigens from allergens with estimated accuracy of 88.9%, while AllerTOP v2.0 employs various machine-learning methods to sort allergens, including *k*-nearest neighbors, E-descriptors as well as auto and cross variance transformation. In order to predict the protein solubility, ProteinSol server was used, which provides a population average for the experimental dataset of 0.45 and any values above this score is considered soluble.

Identification of post-translational modification (PTM) sites

PTM sites were computationally analyzed in the examined protein sequence using online servers such as CSS-Palm and GPS-PAIL 2.0 (The cuckoo workgroup) and NetNGlyc, NetOGlyc and NetPhos (DTU health tech services) in order to predict palmitoylation, acetylation, N- and O-glycosylation as well as phosphorylation sites.

Prediction of transmembrane domain, signal peptide and subcellular localization

The presence of transmembrane helices in protein was estimated using TMHMM 2.0 server from DTU health tech services, which is accessible *via*. Moreover, the presence of a signal peptide and subcellular localization of EgGST protein were predicted using SignalP and DeepLoc web tools of the DTU health tech services.

Secondary structure and disulfide bond prediction

The secondary structure of EgGST was predicted using PSI-blast based secondary structure PREDiction (PSIPRED) server. Moreover, the DiNNNA online service (unified software for cysteine state and disulfide bond partner prediction) was utilized to predict the putative disulfide bonds in the protein sequence.

Three-dimensional (3D) model construction, refinement and validations

Homology modelling of the protein is a substantial step in vaccine design; hence, we employed SWISS-MODEL to develop models out of the provided templates. To enhance the quality of the best-fit 3D model provided by the SWISS-MODEL server, refinement procedure was applied using DeepRefiner server. This server provides five refined models, based on different scores, comprising predicted global quality score, Rosetta energy score, MolProbity scores, GOAP score, OPSUS-PSP score, DFIRE score and RWPlus score. Next, three web servers including Prosa-Web and PROCHECK validated the quality of the refined model.

Continuous and conformational B-cell epitope prediction

Regarding linear B-cell epitopes, several web servers were employed such as Bcepred that utilizes physico-chemical properties including accessibility, hydrophilicity, polarity, flexibility/mobility, exposed surface and turns for prediction (accuracy: 58.70%), ABCpred server that applies artificial neural network (ANN) for the prediction, and B-cell tool of the Immune Epitope Database (IEDB) server which predicts B-cell epitopes along with hydrophilicity, antigenicity, surface accessibility, beta-turn and flexibility. In addition, conforma-

tional B-cell epitopes were predicted using ElliPro tool of the IEDB server.

Prediction of mouse and human MHC-binding epitopes

MHC-binding epitopes with specific affinity to different mouse and human MHC alleles were predicted by. Accordingly, eight mouse MHC-I alleles and three mouse MHC-II alleles were used. Furthermore, HLA-A*02:01, HLA-A*24:02 (MHC-I), DRB1*01:02 and DQA1*05:01/DQB1*03:01 (MHC-II) were used as human alleles. For all MHC-I and MHC-II prediction, 12-mer and 15-mer epitopes were predicted using IEDB recommended 2020.09 (NetMHCPan EL 4.1) and IEDB recommended 2.22 method, respectively. The outputs are provided in percentile ranks, so that lower ranks demonstrate higher epitope affinity to a given MHC allele. All high-ranked epitopes were then screened in terms of antigenicity and allergenicity using VaxiJen v2.0 and AllergenFP v1.0 servers, respectively.

Prediction of the cytotoxic T-lymphocyte (CTL) epitopes

NetCTL 1.2 web server was utilized which predicts CTL epitopes with respect to the 12 major MHC-I supertypes. Here, we predicted the CTL epitopes with respect to the most common supertypes (A1, A2, A3 and B7) to more cover the global human population. Next, these epitopes were screened regarding immunogenicity and allergenicity using MHC-I immunogenicity tool of the IEDB and AllerTOP v2.0 server, respectively.

Results

Physico-chemical properties

This protein had 219 amino acids (MW: 25553.51) and a theoretical pI is 7.5. The number of positively- (Arg + Lys) and negatively-charged residues (Asp + Glu) was almost equal (32 vs 31) and the total number of atoms was estimated as 3582. The extinction coefficients (M^{-1} cm^{-1}) at 280 nm in water were 41745 (assuming all pairs of Cys residues form cystines) and 41370 (all Cys residues are reduced). In addition, the half-life of the protein was calculated to be 30 hours (mammalian reticulocytes, *in vitro*), >20 hours (yeast, *in vivo*) and >10 hours (*Escherichia coli*, *in vivo*). The EgGST protein was rendered as a stable protein, since it demonstrated an instability score of 31.48. Moreover, the aliphatic index and GRAVY score of this protein were 78.04 (high thermotolerance) and -0.440 (hydrophilic), respectively.

Antigenic, allergenic and solubility profiles of EgGST

The antigenicity was estimated to be 0.3209 (VaxiJen) and 0.360934 (ANTIGENpro). In addition, AllergenFP and AllerTOP demonstrated that the protein is not allergenic in nature and does not contain IgE epitopes. The protein was demonstrated to be highly soluble, with a solubility score of 0.494 predicted by Protein-Sol (Fig. 1).

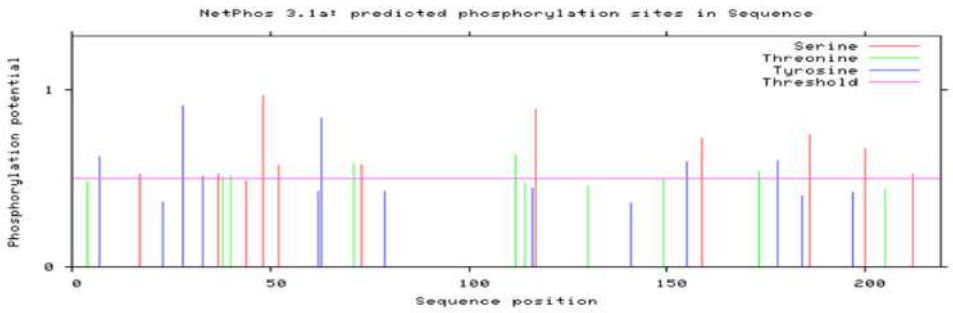

Fig. 1: Calculated solubility of the EgGST protein using Protein-Sol web server

Prediction of subcellular localization, PTM sites, transmembrane domain and signal peptide

Based on DeepLoc tool, EgGST is a cytoplasmic (likelihood: 0.9606) and soluble (likelihood: 0.9926) protein, while no transmembrane domain and signal peptide was predicted for the protein using TMHMM and SignalP, respectively. Regarding PTM sites, no N-glycosylation sites were found within the sequence, whereas there exist several phosphorylation sites (21), single acetylation, and O-glycosylation site in EgGST protein. With respect to phosphorylation regions, most of them (10) were serine phosphorylation sites, followed by threonine phosphorylation (6) and tyrosine phosphorylation (5) (Fig. 2).

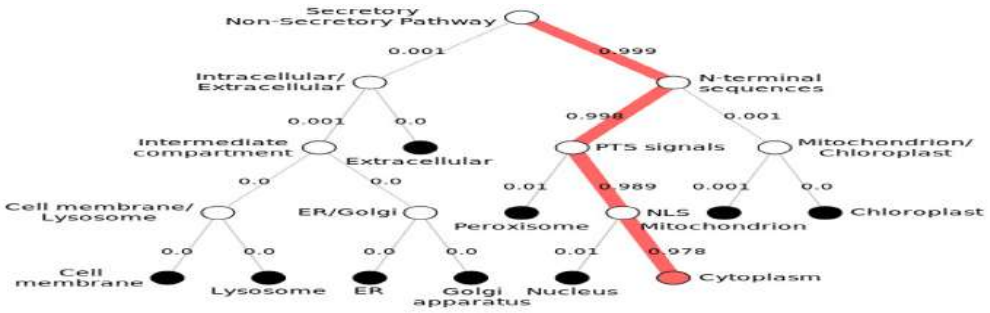

Fig. 2: Prediction of phosphorylation sites of the EgGST protein using NetPhos web server suggested 21 putative regions

Secondary structure and disulfide bond prediction

Secondary structure analysis showed the predominance of helices in most parts of the sequence, followed by coils and strand without disordered regions (Fig. 3). The output of DiANNA showed that the cysteines are expected to form disulfide bonds at the following positions: 90-102 (GMIPDCKKRRA – LHMLQCEVVDL), 115-219 (AFTRTCYSPDF – KWRGDCXXXXX) and 161-208 (PDFSLCELLNQ – FKTRPCNGASA) (Table 1).

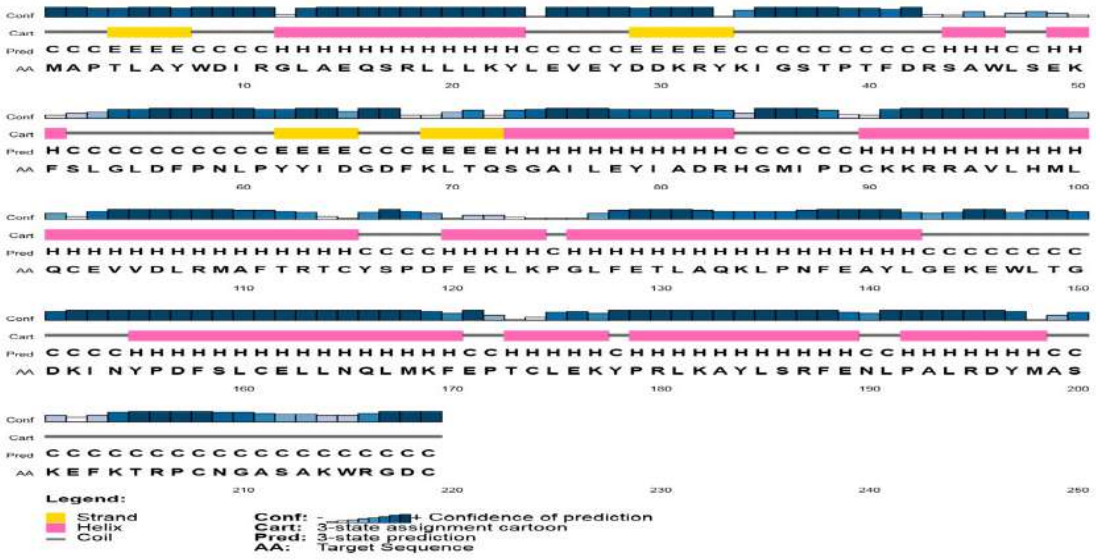

Fig. 3: Secondary structure prediction using PSIPRED web server

Table 1: Predicted disulfide bonds of the EgGST protein using DiANNA web server

Disulfide bond scores			
Cysteine sequence position	***Distance***	***Bond***	***Score***
90-102	12	GMIPDCKKRRA-LHMLQCEVVDL	0.01157
90-115	25	GMIPDCKKRRA-AFTRTCYSPDF	0.01061
90-161	71	GMIPDCKKRRA-PDFSLCELLNQ	0.01059
90-147	84	GMIPDCKKRRA-KFEPTCLEKYP	0.01039
90-208	118	GMIPDCKKRRA-FKTRPCNGASA	0.01064
20-219	129	GMIPDCKKRRA-KWRGDCXXXXX	0.01042
102-115	13	LHMLQCEVVDL-AFTRTCYSPDF	0.01051
102-161	59	LHMLQCEVVDL-PDFSLCELLNQ	0.01039
102-174	72	LHMLQCEVVDL-KFEPTCLEKYP	0.01045
102-208	106	LHMLQCEVVDL-FKTRPCNGASA	0.01096
102-219	117	LHMLQCEVVDL-KWRGDCXXXXX	0.01262
115-161	46	AFTRTCYSPDF-PDFSLCELLNQ	0.01051
115-174	59	AFTRTCYSPDF-KFEPTCLEKYP	0.01046
115-208	93	AFTRTCYSPDF-FKTRPCNGASA	0.01188
115-219	104	AFTRTCYSPDF-KWRGDCXXXXX	0.01442
161-174	13	PDFSLCELLNQ-KFEPTCLEKYP	0.01109
161-208	47	PDFSLCELLNQ-FKTRPCNGASA	0.10508
161-219	58	PDFSLCELLNQ-KWRGDCXXXXX	0.01059
174-208	34	KFEPTCLEKYP-FKTRPCNGASA	0.01095
174-219	45	KFEPTCLEKYP-KWRGDCXXXXX	0.0113
208-219	11	FKTRPCNGASA-KWRGDCXXXXX	0.01042
Predicted bonds			
90-102		GMIPDCKKRRA - LHMLQCEVVDL	
115-219		AFTRTCYSPDF - KWRGDCXXXXX	
161-208		PDFSLCELLNQ - FKTRPCNGASA	
Predicted connectivity			
1-2, 3-7, 4-6			

3D structure, refinement and validation

The SWISS-MODEL server predicted 50 templates for our submitted EgGST sequence, among which only one model was provided for further refinement and validation. This 3D model (homo-dimer) was derived from template 1gta.1.A, having 43.40% sequence identity and relatively full coverage (0.97) with the initially-submitted sequence (Fig. 4) illustrates some details of the chosen 3D model.

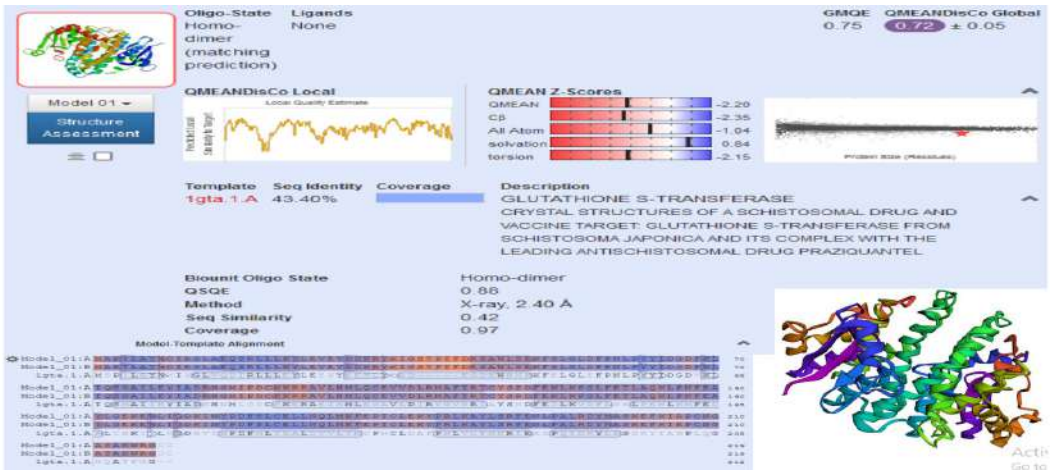

Fig. 4: Evaluation of the refined model in comparison with the crude model showed improvements as indicated by Prosa-Web and PROCHECK web tools

Based on DeepRefiner server output, model number 4 was selected with the following scores: predicted global quality score of 0.265, Rosetta energy score of -1163.060, MolProbity score of 1.119, GOAP score of -62241.600, OPSUS-PSP score of -7902.856, DFIRE score of -834.650 and RWPlus score of -99003.536. Based on Prosa-Web, a Z-score of -7.87 was calculated for the crude mode, while the Z-score of refined model was improved to -8.14. Moreover, Ramachandran analysis showed that 346 (90.1%), 32 (8.3%), 4 (1.0%) and 2 (0.5%) of residues were allocated to the most favored, additional allowed, generously allowed and disallowed regions, respectively. These allocations were changed in the refined model as follows: 347 (90.4%), 34 (8.9%), 3 (0.8%) and 0 (0.0%) the most favored, additional allowed, generously allowed and disallowed regions, respectively (Fig. 5).

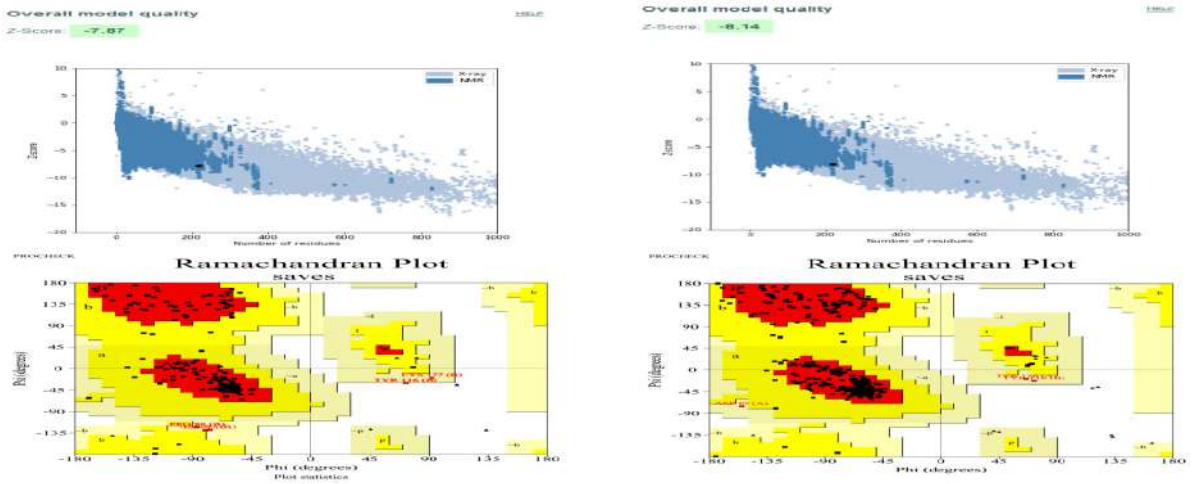

Fig. 5: Homo-dimer 3D model of the EgGST protein and other conformational parameters predicted using SWISS-MODEL web server

Prediction of linear and conformational B-cell epitopes

Based on ABCpred output, 7 linear B-cell epitopes were predicted in EgGST protein, all having a threshold scores of over 0.75 (Table 2).

Table 2: Prediction of linear B-cell epitopes in EgGST protein using ABCpred server (Threshold: 0.75)

Start Position	*Epitope Sequence*	*Score*
178	YPRLKAYLSRFENL	0.94
29	DDKRYKIGSTPTFD	0.86
124	KPGLFETLAQKLPN	0.86
3	PTLAYWDIRGLAEQ	0.85
66	GDFKLTQSGAILEY	0.83
96	VLHMLQCEVVDLRM	0.78
72	QSGAILEYIADRHG	0.77

In addition, Table 3 demonstrates the predicted continuous B-cell epitopes regarding hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity and antigenic propensity. The B-cell tool of the IEDB, also, predicted epitopes regarding Chou and Fasman turn, Emini surface accessibility, Karplus & Schulz flexibility, Kolaskar & Tongaonkar antigenicity, Parker hydrophilicity and Bepipred linear epitopes with averages scores of 0.977, 1.000, 0.992, 1.027, 0.985 and -0.112, respectively (Fig. 6).

Table 3: Specific B-cell linear epitopes of EgGST based on different physico-chemical parameters predicted by the Bcepred web server

Physico-chemical parameter	*B-cell epitopes*
Hydrophilicity	EVEYDDKRYK, PDCKKRRA, KTRPCNG
Flexibility	GMIPDCKKR, YMASKEFK
Accessibility	KYLEVEYDDKRYKIGSTPTFDRS, EYIADRH, IPDCKKRRAVLH, RTCYSPDFEKLKPGL, QKLPNFEAYLGEKEW, TGDKINYPDF, CLEKYPRLKAYLSRFENLP, RDYMASKEFKTRPCNG, SAKWRGD
Turns	-
Exposed surface	EVEYDDKRYKIG, IPDCKKRRAVL, PDFEKLKP, EKYPRLKA, SKEFKTRP
Polarity	KYLEVEYDDKRYKIGS, EYIADRH, IPDCKKRRAVLHM, PDFEKLKP, YLGEKEWLTG, EPTCLEKYPRLKA, MASKEFKTRPC
Antigenic propensity	SRLLLKYLEVE, PNLPYYI, VLHMLQCEVVDLR, YPDFSLCELLNQL

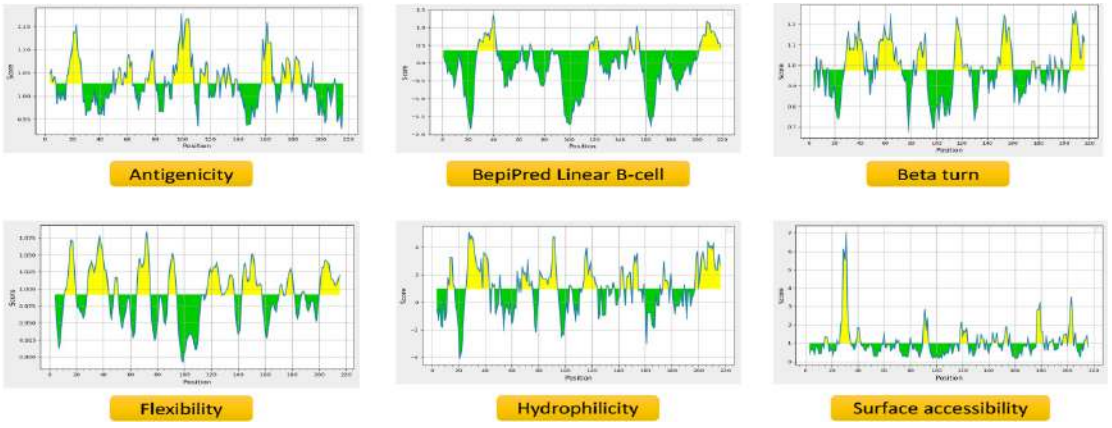

Fig. 6: Predicted linear B-cell epitopes regarding antigenicity, BepiPred, beta-turn, flexibility, hydrophilicity and surface accessibility using B-cell online tool of the IEDB server

The ElliPro output for conformation B-cell epitopes were as follows: 1) 46 residues, score: 0.743; 2) 46 residues, score: 0.727; 3) 67 residues, score: 0.695; 4) 50 residues, score: 0.68; 5) 5 residues, score: 0.621; 6) 14 residues, score: 0.618; and 7) 8 residues, score: 0.578 (Fig. 7).

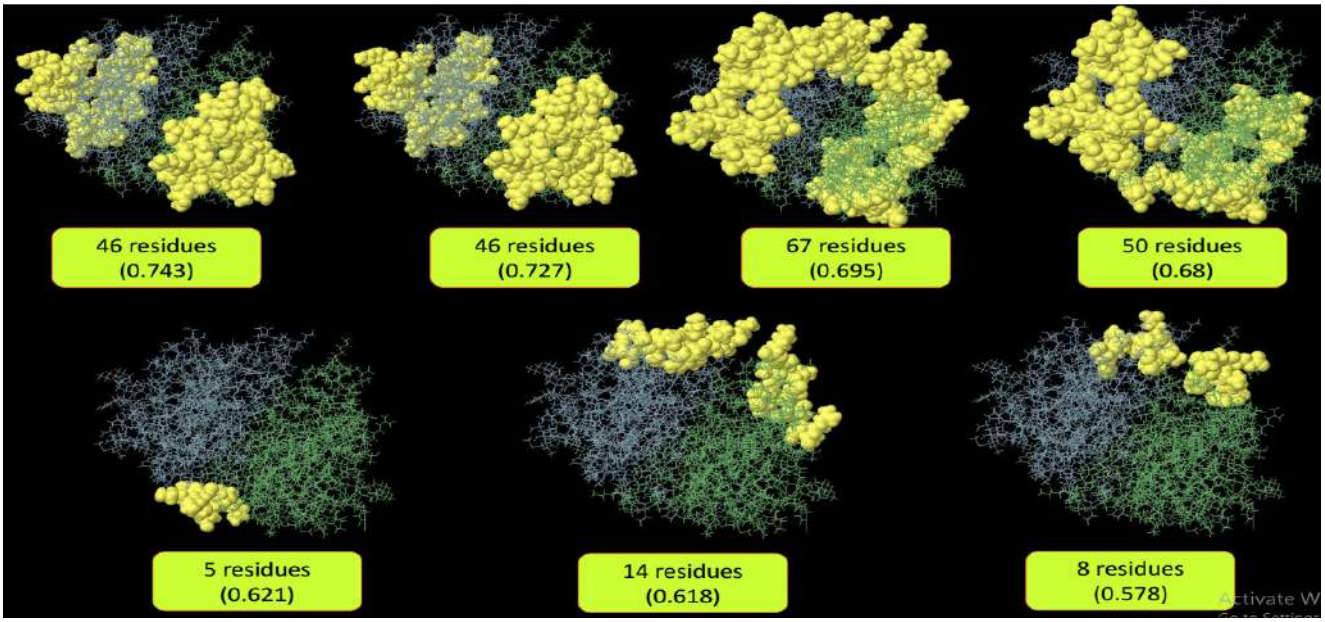

Fig. 7: Predicted conformational B-cell epitopes of the EgGST protein using ElliPro tool of IEDB server

Prediction of mouse and human MHC-binding epitopes

Mouse MHC-binding epitopes were predicted with respect to several MHC-I and MHC-II alleles. Those high-ranked and qualified epitopes were then screened in terms of antigenicity and allergenicity, resulting in 7 antigenic, non-allergenic epitopes, including 2 MHC-I-binding epitopes of "FPNLPYYIDGDF" and "SPDFEKLKPGLF" as well as 5 MHC-II-binding epitopes of "GDFKLTQSGAILEYI", "DGDFKLTQSGAILEY", "CLEKYPRLKAYLSRF", "TCLEKYPRLKAYLSR", "EPTCLEKYPRLKAYL" (Table 4).

A similar approach was exerted to determine human MHC-binding epitopes. The output revealed 5 top-ranked, antigenic and non-allergenic epitopes, comprising "KYPRLKAYLSRF", "DFKLTQSGAILEYIA", "FKLTQSGAILEYIAD", "GDFKLTQSGAILEYI" and "KLTQSGAILEYIADR" (Table 5).

Prediction of human CTL epitopes

Based on NetCTL 1.2 server, 3 CTL epitopes were predicted to possess high antigenicity scores and without allergenicity, encompassing "TQSGAILEY" (A1 supertype), "SLGLDFPNL" (A2 supertype) and "GLFETLAQK" (A3 supertype) (Table 6).

Table 4: Prediction of the high-ranked epitopes 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)	AYLSRFENLPAL	43-54	1.7	-0.4482	Yes
	RFENLPALRDYM	47-58	3.2	-0.4182	Yes
	SLGLDFPNLPYY	52-63	3.8	-0.0745	Yes
H2-Dd (MHC-I)	YSPDFEKLKPGL	46-57	0.72	0.3268	No
	LGLDFPNLPYYI	53-64	2.1	0.1802	Yes
	FEPTCLEKYPRL	30-41	3.0	-0.1890	No
H2-Kb (MHC-I)	INYPDFSLCELL	13-24	0.7	-0.0649	No
	YSPDFEKLKPGL	46-57	2.2	0.3268	No
	AYLSRFENLPAL	43-54	3.8	-0.4482	Yes
H2-Kd (MHC-I)	AYLSRFENLPAL	43-54	0.29	-0.4482	Yes
	NLPYYIDGDFKL	59-70	1.9	0.7856	Yes
	RLKAYLSRFENL	40-51	2.0	0.8088	Yes
H2-Kk (MHC-I)	FEKLKPGLFETL	50-61	2.0	0.3035	Yes
	LEVEYDDKRYKI	24-35	2.6	1.3909	Yes
	FEPTCLEKYPRL	30-41	3.5	-0.1890	No
H2-Ld (MHC-I)	KPGLFETLAQKL	54-65	0.47	-0.3559	Yes
	FPNLPYYIDGDF*	57-68	0.71	0.9888	No
	SPDFEKLKPGLF*	46-58	0.92	0.8299	No
H2-Qa1 (MHC-I)	SLGLDFPNLPYY	52-63	1.6	-0.0745	Yes
	GLAEQSRLLLKY	12-23	4.9	0.2913	No
	FSLGLDFPNLPY	51-62	6.3	-0.2369	Yes
H2-Qa2 (MHC-I)	FEKLKPGLFETL	50-61	2.6	0.3035	Yes
	LEVEYDDKRYKI	24-35	2.9	1.3909	Yes
	FEPTCLEKYPRL	30-41	3.1	-0.1890	No
H2-IAb (MHC-II)	GDFKLTQSGAILEYI*	66-80	8.05	0.5829	No
	DGDFKLTQSGAILEY*	65-79	8.85	0.7770	No
	DKRYKIGSTPTFDRS	30-44	8.90	-0.3659	No
H2-IAd (MHC-II)	LEKYPRLKAYLSRFE	175-189	2.57	-0.7481	No
	CLEKYPRLKAYLSRF*	174-188	3.06	0.8769	No
	TCLEKYPRLKAYLSR*	173-187	3.5	0.8828	No
H2-IEd (MHC-II)	PTCLEKYPRLKAYLS	172-186	1.21	0.7676	Yes
	EPTCLEKYPRLKAYL*	171-185	1.5	0.8369	No
	FEPTCLEKYPRLKAY	170-184	1.6	0.4151	No

(*) indicates potentially qualified epitopes

Table 5: Prediction of the high-ranked epitopes 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*	*Antigenicity*	*Allergenicity*
HLA-A*02:01	NLPYYIDGDFKL	59-70	0.91	0.7856	Yes
	YLSRFENLPALR	44-55	1.9	-0.5568	No
	AYLSRFENLPAL	43-54	2.3	-0.4482	Yes
	GLFETLAQKLPN	56-67	5.3	-0.4198	Yes
	RLKAYLSRFENL	40-51	7.4	0.8088	Yes
HLA-A*24:02	KYPRLKAYLSRF*	37-48	0.25	1.3018	No
	TRTCYSPDFEKL	42-53	0.47	-0.0078	No
	NLPYYIDGDFKL	59-70	0.81	0.7856	Yes
	AYLSRFENLPAL	43-54	0.99	-0.4482	Yes
	DKRYKIGSTPTF	30-41	1.1	-0.2191	Yes
	AVLHMLQCEVVDLRM	95-109	1.9	0.4135	Yes
	KKRRAVLHMLQCEVV	91-105	1.9	0.2785	Yes
DRB1*01:02	KRRAVLHMLQCEVVD	92-106	1.9	0.0761	No
	LHMLQCEVVDLRMAF	97-111	1.9	0.3121	No
	RAVLHMLQCEVVDLR	04-108	1.9	0.2724	No
	DFKLTQSGAILEYIA*	67-81	16.0	0.5560	No
DQA1*05:01/DQB1*03:01	FKLTQSGAILEYIAD*	68-82	19.0	0.5178	No
	GDFKLTQSGAILEYI*	66-80	20.0	0.5829	No
	KLTQSGAILEYIADR*	69-83	20.0	0.5984	No
	LTQSGAILEYIADRH	70-84	24.0	0.4237	No

Table 6: Prediction of the high-ranked human CTL epitopes from EgGST protein using NetCTL web server, with subsequent screening using immunogenicity and allergenicity.

Human MHC class I supertype	*CTL epitopes*	*Immunogenicity score*	*Allergenicity*
A1 Supertype	GLDFPNLPY	0.09227	No
	LSEKFSLGL	-0.21579	Yes
	TQSGAILEY*	0.19292	No
	RMAFTRTCY	0.21828	Yes
	EVEYDDKRY	-0.0805	Yes
A2 Supertype	RLLLKYLEV	-0.1791	Yes
	MLQCEVVDL	0.09231	No
	KLPNFEAYL	0.229	Yes
	SLCELLNQL	-0.01113	No
	SLGLDFPNL*	0.1185	No
A3 Supertype	GLDFPNLPY	0.09227	No
	RMAFTRTCY	0.21828	Yes
	GLFETLAQK*	0.13145	No
	CLEKYPRLK	-0.16134	Yes
	ALRDYMASK	-0.19342	Yes
	YPDFSLCEL	-0.03354	No
	LPYYIDGDF	0.18712	Yes
B7 Supertype	TPTFDRSAW	0.08396	Yes
	LAEQSRLLL	-0.21228	Yes
	RPCNGASAK	-0.07094	No

(*) indicates potentially qualified epitopes

Discussion

CE transmission could be averted by implementation of preventive measures such as safe animal slaughtering, routine dosing of dogs and immunization strategies targeting both definitive and/or intermediate hosts (9). Previously, several antigenic compounds of *E. granulosus* were produced in recombinant form and shown to play a protective role against CE as a potent vaccine candidate (13,14). Traditional vaccine Research and Development is a costly and laborious procedure by itself, in particular to combat parasites having complex lifestyles such as *E. granulosus*, while this could be facilitated using finely-tuned, accurate computational methods. Immune cells may efficiently recognize several immunodominant epitopes. Hence, accurate discovery of these epitopes directs us for a rational vaccine design using multi-epitope vaccine compartments. Machine-learning methods and immunoinformatics algorithms using web servers and standalone software pave the way to achieve such a promising goal (2,15).

Our knowledge on the immunological interactions at the host-CE interface have been mostly gained during 21st century using omics-based technologies (16). Increase in IgE, IgM and IgG is the usual humoral immunity reaction against CE, in particular IgG1 and IgG4 subtypes, whereas cell-mediated response is represented as a dichotomy. Th1 responses are highlighted in early infection, which benefit the host by inhibiting parasite growth, while Th2 responses act inversely. Therefore, any imbalance in both responses is substantial-

ly associated with immunopathogenesis- control in CE. For this purpose, identification and proper targeting of novel vaccine antigens and/or epitopes is of utmost significance (17). Herein, we evaluated some basic biochemical properties and potential immunogenic epitopes of EgGST protein as a cornerstone for future vaccination studies.

According to ProtParam server, the protein had 219 amino acids with a MW of 25.55 kDa, which indicates adequate antigenicity in nature, since those antigens having MW of < 5-10 kDa are poor immunogens (7, 18). The aliphatic index and GRAVY score of EgGST protein were predicted as 78.04 and -0.440, respectively. The higher is the aliphatic index, the more stable is the protein in wide ranges of temperatures; hence, EgGST is relatively thermostable. In addition, a negative GRAVY value implies to a hydrophilic molecule that better interacts in the surrounding water-based milieu (19). Despite low antigenic index predicted by the VaxiJen and ANTIGENpro web servers, EgGST protein has been known to induce specific immune responses in vaccination trials (20). Thus, it deserves to be excavated for the identification of novel antigenic epitopes. A good vaccine candidate should not be allergenic in nature (21). Accordingly, this protein neither was allergenic nor possessed IgE epitopes. Altogether, these basic features are necessary for future wet laboratory experiments. It is well-known that PTMs play a critical role in different cellular processes (22). For this aim, we utilized various servers to determine phosphorylation, glycosylation and acetylation sites in the protein sequence. Based on our results, only one acetylation and O-glycosylation sites were found in the protein, while there were numerous (21) phosphorylation sites in EgGST, among which serine phosphorylation sites prevailed. The protein had no transmembrane domain and signal peptide as well.

Prediction of protein secondary and tertiary structure is a key step in bioinformatics analysis of a vaccine candidate. Here, we predicted secondary structure using PSIPRED server, which demonstrated the frequency of helices in most parts of the sequence, followed by coils and strand without disordered regions. In addition, DiANNA server revealed the presence of putative disulfide bonds at the positions 90-102 (GMIPDCKKRRA – LHMLQCEVVDL), 115-219 (AFTRTCYSPDF – KWRGDCXXXXX) and 161-208 (PDFSLCELLNQ – FKTRPCNGASA). Generally, disulfide bonds would strengthen the protein conformation. Tertiary structure of the protein was predicted using a homology-modelling server, SWISS-MODEL. "A major concern in structural biology is the identification of faults in experimental and theoretical models of protein structures" (23). Refinement and validation steps are auxiliary to enhance the quality and reliability of an *in silico* predicted model (24); therefore, we employed DeepRefiner web server to rehash the 3D model generated by SWISS-MODEL server. Subsequently, the refined model was submitted for validation using Prosa-Web and Ramachandran plot analysis (PROCHECK). Comparable with the initial model, refined model showed some degrees of quality enhancements.

Improved prevention and/or control of CE completely depends on the induction of acquired immunity, i.e. humoral and cellular immune responses. By the advent of computational vaccine design strategies, we can efficiently induce these immunological pathways to confine the parasitic infection (17). In this study, we utilized several techniques and web tools to predict B- and T-cell epitopes of EgGST protein to achieve the highest accuracy. In this sense, B-cell epitopes were predicted using Bcepred, IEDB, and ABCpred online servers. The prediction accuracy of Bcepred and ABCpred servers are 52.92-57.93% and 65.93%, respectively. Prediction of conformational B-cell epitopes is another important step in bioinformatics analysis of vaccine candidates. For this aim, we used ElliPro tool of the IEDB server, which demonstrated seven

conformational B-cell epitopes for EgGST protein that are significant with respect to antigen-antibody interactions. We also predicted those mouse- and human-specific epitopes that are important during T-cell/MHC interplay using MHC-binding epitope prediction tool of the IEDB server. Accordingly, 2 MHC-I-binding and 5 MHC-II-binding epitopes were found to possess high affinity to mouse MHC molecules and they were highly antigenic and non-allergenic in nature. Moreover, 5 high-ranked epitopes specific to human MHC molecules were predicted as follows: "KYPRLKAYLSRF", "DFKLTQSGAILEYIA", "FKLTQSGAILEYIAD", "GDFKLTQSGAILEYI" and "KLTQSGAILEYIADR". In addition, NetCTL server was used in the present study to predict CTL epitopes in terms of different MHC supertypes worldwide. Based on the outputs of this server, 3 epitopes were found to possess antigenicity without allergenicity, including "TQSGAILEY" (A1 supertype), "SLGLDFPNL" (A2 supertype) and "GLFETLAQK" (A3 supertype). Altogether, these qualified epitopes can be utilized alone and/or combined with potential epitopes derived from other vaccine antigens to design and engineer an efficacious multi-epitope vaccine construct to be evaluated against CE in future.

Conclusion

This paper provided some biophysical characteristics of EgGST protein, along with its subcellular localization, transmembrane domain, secondary and tertiary structure as well as B- and T-cell epitopes. It is suggested to validate the vaccine candidate regarding effective dose, cross-reaction, total antibody titers, lymphocyte proliferation and cytokine production assays in the context of animal model challenge studies. Inevitably, these findings would assist us for future immunization investigations against CE.

Acknowledgements

The authors would like to thank all staff of Department of Parasitology of Tarbiat Modares University. This article is part of PhD thesis of Medical Parasitology.

References

1. 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):e00075-18.
2. Pourseif MM, Moghaddam G, Saeedi N, Barzegari A, Dehghani J, Omidi Y. Current status and future prospective of vaccine development against *Echinococcus granulosus*. Biologicals. 2018;51:1-11.
3. Shams M, Khazaei S, Naserifar R, Shariatzadeh SA, Anvari D, Montazeri F, Pirestani M, Majidiani H. Global distribution of *Echinococcus granulosus* genotypes in domestic and wild canids: a systematic review and meta-analysis. Parasitology. 2022;149: 1147-1159.
4. McManus DP, Gray DJ, Zhang W, Yang Y. Diagnosis, treatment, and management of echinococcosis. BMJ. 2012;344: e3866.
5. Budke CM, Deplazes P, Torgerson PR. Global socioeconomic impact of cystic echinococcosis. Emerg Infect Dis. 2006;12(2):296-303.
6. Otero-Abad B, Torgerson PR. A systematic review of the epidemiology of echinococcosis in domestic and wild animals. PLoS Negl Trop Dis. 2013;7(6):e2249.
7. 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.
8. Zhang W, McManus DP. Vaccination of dogs against *Echinococcus granulosus*: a means to control hydatid disease? Trends Parasitol. 2008;24(9):419-24.
9. Larrieu E, Gavidia CM, Lightowlers MW. Control of cystic echinococcosis:

background and prospects. Zoonoses Public Health. 2019;66(8):889-99.
10. Craig P, Hegglin D, Lightowlers M, Torgerson PR, Wang Q. Echinococcosis: control and prevention. Adv Parasitol. 2017;96:55-158.
11. Morello A, Repetto Y, Atias A. Characterization of glutathione S-transferase activity in *Echinococcus granulosus*. Comp Biochem Physiol B. 1982;72(3):449-52.
12. Dowling DJ, Hamilton CM, Donnelly S, et al. Major secretory antigens of the helminth *Fasciola hepatica* activate a suppressive dendritic cell phenotype that attenuates Th17 cells but fails to activate Th2 immune responses. Infect Immun. 2010;78(2):793-801.
13. Gauci C, Vural G, Öncel T, et al. Vaccination with recombinant oncosphere antigens reduces the susceptibility of sheep to infection with *Taenia multiceps*. Int J Parasitol. 2008;38(8-9):1041-50.
14. Pourseif MM, Moghaddam G, Nematollahi A, et al. Vaccination with rEGVac elicits immunoprotection against different stages of *Echinococcus granulosus* life cycle: a pilot study. Acta Trop. 2021;218:105883.
15. Del Tordello E, Rappuoli R, Delany I. Reverse vaccinology: exploiting genomes for vaccine design. Hum Vaccin. 2017. P. 65-86.
16. Gottstein B, Soboslay P, Ortona E, Wang J, Siracusano A, Vuitton D. Immunology of alveolar and cystic echinococcosis (AE and CE). Adv Parasitol. 2017;96:1-54.
17. Zhang W, Wen H, Li J, Lin R, McManus DP. Immunology and immunodiagnosis of cystic echinococcosis: an update. Clin Dev Immunol. 2012;2012: 101895.
18. Khazaei S, Dalimi A, Pirestani M, Ghafarifar F. *In silico* analysis of a 29 kDa *Echinococcus granulosus* protoscolex protein (P29) as a vaccine candidate against cystic echinococcosis. Arch Razi Inst. 2023; 78: 323-335.
19. Shams M, Khazaei S, Nazari N, Majidiani H, Kordi B. Shedding light on biochemical features and potential immunogenic epitopes of *Neospora caninum* SAG1: In silico study. Inform Med Unlocked. 2021;27:100785.
20. Zhu M, Wang X, Wang H, et al. Mechanism of protective immunity by vaccination with recombinant *Echinococcus granulosus* glutathione S-transferase (Chinese strain) in mice. Exp Ther Med. 2015;10(3):1127-32.
21. Asghari A, Majidiani H, Fatollahzadeh M, et al. Insights into the biochemical features and immunogenic epitopes of common bradyzoite markers of the ubiquitous *Toxoplasma gondii*. Infect Genet Evol. 2021;95:105037.
22. Lee TY, Hsu JBK, Chang WC, et al. A comprehensive resource for integrating and displaying protein post-translational modifications. BMC Res Notes. 2009;2:111.
23. Ghaffari AD, Dalimi A, Ghaffarifar F, Pirestani M. Structural predication and antigenic analysis of ROP16 protein utilizing immunoinformatics methods in order to identification of a vaccine against *Toxoplasma gondii*: an in silico approach. Microb Pathog. 2020;142:104079.
24. Lamzin VS, Wilson KS. Automated refinement of protein models. Acta Crystallogr D Biol Crystallogr. 1993;49(Pt 1):129-47.