Infection, Genetics and Evolution 73 (2019) 98–108

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

# Bioinformatics analysis of EgA31 and EgG1Y162 proteins for designing a multi-epitope vaccine against *Echinococcus granulosus*

Xiao Zhao[a,b,1], Fengbo Zhang[a,c,1], Zhiwei Li[b,1], Hongying Wang[b], Mengting An[b], Yujiao Li[c], Nannan Pang[c], Jianbing Ding[a,b,*]

[a] State Key Laboratory of Pathogenesis, Prevention, Treatment of Central Asian High Incidence Diseases, the First Affiliated Hospital of Xinjiang Medical University, Urumqi 830011, Xinjiang, PR China
[b] College of Basic Medicine of Xinjiang Medical University, Urumqi 830011, Xinjiang, PR China
[c] Department of Clinical Laboratory, the First Affiliated Hospital of Xinjiang Medical University, Urumqi 830011, Xinjiang, PR China

ARTICLE INFO

*Keywords:*
*Echinococcus granulosus*
EgA31
EgG1Y162
Bioinformatics analysis
Multi-epitope
Vaccine

ABSTRACT

Cystic echinococcosis (CE) is a global zoonosis caused by the larvae of the parasite *Echinococcus granulosus (E. granulosus)*. According to its life cycle and previous studies of antigen candidates for vaccines against *E. granulosus*, we chose two proteins expressed at different stages of the *E. granulosus* life cycle to design a multi-epitope vaccine. The EgA31 antigen gene is derived from the adult stage of *E. granulosus,* and the EgG1Y162 antigen gene is derived from the larval stage of *E. granulosus*. In this study, we used several bioinformatics methods to analyze various aspects of the EgA31 and EgG1Y162 proteins, including the physicochemical properties, secondary and tertiary structures, and the dominant T-cell and B-cell epitopes. The results showed that EgA31 protein was an unstable and hydrophilic protein, while EgG1Y162 was stable and hydrophobic. The secondary structure of the EgA31 protein consisted of 82.36% alpha helixes, 4.16% extended strands, 3.16% beta turns and 10.32% random coils. The secondary structure of EgG1Y162 consisted of 33.33% alpha helixes, 25.49% extended strands, 5.88% beta turns and 35.29% random coils. Moreover, our results identified 6 dominant T-cell epitopes and 5 dominant B-cell epitopes in the EgA31 protein structure and 6 dominant T-cell epitopes and 3 dominant B-cell epitopes in EgG1Y162. In conclusion, this study provides comprehensive biological information about the EgA31 and EgG1Y162 proteins, which will lay a theoretical foundation for multi-epitope vaccines against *Echinococcus granulosus*.

## 1. Introduction

Cystic echinococcosis (CE) is a neglected helminthic disease caused by the larvae of *Echinococcus granulosus (E. granulosus)* and remains a serious public health problem in developing countries (Otero-Abad and Torgerson, 2013). Based on the World Health Organization (WHO) report, the cost of controlling echinococcosis is over 3 billion USD annually, while the infection rate continues to increase every year. In China, CE is mainly prevalent in agricultural and pastoral areas, such as Xinjiang, Qinghai, Ningxia, and Gansu, and it is an enormous economic and health burden (Cumino et al., 2012).

*Echinococcus granulosus* could infect humans and many other herbivores. The definitive hosts are carnivores, especially dogs, which produce the gravid proglottid that contains the eggs of *E. granulosus* and releases them into the environment (Ozturk et al., 2007). The intermediate hosts of *E. granulosus* are sheep, cattle, goat and humans. People often are infected with CE by accidentally ingesting the eggs from polluted water and food (Otero-Abad and Torgerson, 2013). The oncosphere is released from the embryophore and then develops in the human lungs, liver and other organs. As the larva enlarges, the cyst

*Abbreviations:* 3D, three-dimensional; AAP, amino acid pair; ANN, artificial neural network; APC, antigen presenting cell; CD, cluster of differentiation; CE, cystic echinococcosis; CTL, cytotoxic T lymphocyte; *E. granulosus*, *Echinococcus granulosus*; GRAVY, grand average of hydropathicity; HLA, human leukocyte antigen; IEDB, immune epitope database; MHC, Major Histocompatibility Complex; MW, molecular weight; NCBI, National Center for Biotechnology Information; PCR-SBT, polymerase chain reaction-sequence-based-typing; Phyre2, Protein homology/analogy Recognition Engine V 2.0; pI, isoelectric point; SDS-page, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; WHO, World Health Organization

* Corresponding author at: Department of Immunology, College of Basic Medicine, Xinjiang Medical University, No. 393, Xinyi Road, Urumqi 830011, Xinjiang, PR China.
*E-mail address:* 1601379937@qq.com (J. Ding).
[1] Equal contribution.

https://doi.org/10.1016/j.meegid.2019.04.017
Received 13 December 2018; Received in revised form 17 April 2019; Accepted 19 April 2019
Available online 22 April 2019

squeezes the surrounding organ tissues, causing their atrophy and dysfunction, thus becoming a choric disease (Qian and Zhou, 2018). For CE patients, common treatment strategies are pharmacotherapy and surgical treatment. However, pharmacotherapy is often associated with inevitable side effects, and surgery requires long-term follow-up along with a high recurrence rate. Thus, a prophylactic treatment (e.g., vaccine) for CE becomes one of the most suitable and effective strategies.

Several studies have been conducted on the vaccine antigen candidates of *E. granulosus,* including Eg95, EgA31, EgDf1, EgAg5, and EgG1Y162 (Pourseif et al., 2018a, 2018b). These antigens are expressed at the different developmental stages of *E. granulosus*. EgA31 was a single gene derived from the cDNA library of the adult stage of *E. granulosus* obtained from the sera of infected dogs. The recombinant EgA31 (rEgA31) was used as a protective vaccine and the intradermal injection of rEgA31 in dogs was associated with an elevated cellular immune response (Fu et al., 1999; Fu et al., 2000). The EgG1Y162 gene was cloned from the cDNA library and amplified from both the larvae and adult stages of *E. granulosus* (Cao et al., 2009; Zhang et al., 2014). Recombinant EgG1Y162 showed a reasonably strong immune response, with high specificity and sensitivity (Ma et al., 2016; Zhang et al., 2018).

The immune response can be activated by pathogens and antigen epitopes with strong immunogenicity and immunoreactivity. In view of the importance of the EgA31 and EgG1Y162 proteins (Fu et al., 1999; Zhang et al., 2018), a multi-epitope vaccine containing proper antigenic peptides of both genes may serve as an efficient vaccine against *E. granulosus* infection. In this study, we used bioinformatic methods to analyze the physicochemical parameters and the secondary and tertiary structures and to predict the B- and T-cell epitopes of the two proteins. Here, we discovered that 11 dominant epitopes (5 for B-cell and 6 for T-cell) and 9 dominant epitopes (3 for B-cell and 6 for T-cell) were predicted for protein EgA31 and EgG1Y162, respectively. These results laid the foundation for designing a multi-epitope against *E. granulosus*.

## 2. Materials and methods

### 2.1. Amino acid sequences

The complete amino acid sequences of EgG1Y162 and EgA31 were obtained from the National Center for Biotechnology Information (NCBI) database.

### 2.2. Prediction of physicochemical parameters

Online tools from Expasy ProtParam (Wilkins et al., 1999) (http://web.expasy.org/protparam/) were used to analyze EgG1Y162 and EgA31 proteins, including the molecular weight (MW), theoretical isoelectric point (pI), extinction coefficient, unstable coefficient, total average hydrophobicity, and other physicochemical parameters.

### 2.3. Prediction of transmembrane domains

The transmembrane domains of the EgA31 and EgG1Y162 proteins were analyzed by the TMHMM Server (Moller et al., 2001) (http://www.cbs.dtu.dk/services/TMHMM-2.0/).

### 2.4. Secondary structure prediction

The secondary structures of EgA31 and EgG1Y162 were predicted by DNAstar software using the Gramier - Robson and Chou - Fasman methods (Li et al., 2015). To predict the protein secondary structure more accurately, we performed prediction analyses using SOPMA online analysis software (Deleage, 2017) (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) to improve the reliability of the predicted protein secondary structure.

### 2.5. Tertiary structure prediction

The Protein homology/analogy Recognition Engine V 2.0 (Phyre2) server (Kelley et al., 2015) (http://www.sbg.bio.ic.ac.uk/phyre2/html) was employed to build the three-dimensional models of the EgG1Y162 and EgA31 proteins. The 3D models were built with this server according to the sequence similarities of amino acids. Next, the protein tertiary structures were selected, and then Raswin software was used to demonstrate the 3D models of the proteins.

### 2.6. Linear B-cell epitopes of the EgA31 and EgG1Y162 proteins

To analyze the B-cell linear epitopes, several prediction software packages were used, including BCPREDS Server 1.0 (Chen et al., 2007) (http://ailab.ist.psu.edu/bcpred/predict.html), ABCpred Prediction (Saha and Raghava, 2006) (http://crdd.osdd.net/raghava/abcpred/ABC_submission.html), BepiPred 1.0 Server (Larsen et al., 2006) (http://www.cbs.dtu.dk/services/BepiPred-1.0/) and SVMTriP (Yao et al., 2012) (http://sysbio.unl.edu/SVMTriP/prediction.php). Linear epitopes of 20-mer length were predicted at the default specificity of 75% on the BCPRED server; the 16-mer linear epitopes were predicted at the default 0.51 threshold on ABCpred; the score threshold for epitope assignment was 0.35 on the BepiPred 1.0 server; and linear epitopes of 20-mer length were predicted on SMVTriP.

### 2.7. T-cell epitopes of the EgA31 and EgG1Y162 proteins

T lymphocytes are divided into two groups: CD8+ and CD4+ T cells. CD4+ T cells are the essential component of protective immunity, so we used SYFPEITHI (Rammensee et al., 1995; Rammensee et al., 1999) (http://www.syfpeithi.de/bin/mhcserver.dll/epitopeprediction.htm) and IEDB (Ostrov et al., 2018) (http://tools.iedb.org/mhcii/) software to predict CD4+ T cell epitopes.

According to a previous study, the high frequency HLA alleles for the Uyghur ethnic minority group were HLA-A*1101 (13.46%), A*0201 (12.50%), and A*0301 (10.10%), HLA-DRB1*0701 (16.35%), DRB1*1501 (8.65%) and DRB1*0301 (7.69%) (Shen et al., 2010). We therefore analyzed the epitopes of CD4+ T cells in the context of HLA-DRB1*0701, HLA-DRB1*1501 and HLA-DRB1*0301. Next, epitopes of CD8+ T cells were predicted using the IEDB online tool (Lanoix et al., 2018) (http://tools.iedb.org/mhci/) in the context of the HLA-A*1101, HLA_A*0201 and HLA_A*0301.

### 2.8. The prediction of dominant T- and B- cell epitopes of the EgA31 and EgG1Y162 proteins

In the first step, every predicted epitope was evaluated for their length and position, excluding some epitopes that were too short to form epitopes. In the second step, only the top 5 high-scoring epitopes from each prediction software were listed. In the third step, based on the listed epitopes, we selected the overlapping peptides as the dominant epitopes.

## 3. Results

### 3.1. Gene information for EgA31 and EgG1Y162 proteins

The amino acid sequence of the EgA31 protein was obtained from GenBank as follows:

TRPQRKKNEYEDLELQLENAQNNIRTQESNCRRLSLEHMKALEEIKM
KQITMEGLETKITELIQRNEDLTKEALNTKNVESSNREEVLLSKIKSLEKTA-
KHLHIVLKEEKHYNNQLKEEIDEIRKENLATLQTRLNEIFQKESEMNSERK-
ALNMRIMALEAENERLRISAAEKQAMDDSTASSDGIYFAIEEERKKSAEL-
RRALITQESRNLELQNDLERLQRETRDELDEQKAEIEKLHKELVGADNSK-
TTVQSIRNEMRGIQVQIQLLRGGYLDLFHDKIGHYKEKLQESETKLLELQ-

GTHDQTKRIHDVEKDKLAQELKYMEKQINLCSNENNKLKDALASMENQITGLLNGNELLKKKIEMLETRAETKEINTKLENLLKTETQELLTTIKDLDSKLTESGEEVKELKKQLEKAEKQIRDAEVLTDEKNKIIEEMKKTINKLSETKKEVEDKNSELRKASIQKQSLIEEKEILGRQLELKDTIISAMKKEKDALRHELHEMRDTINTLTEKIATIKIPEPLPMKPIEIPKPLQKEGVNKAMENEIQTYKDTIKTLKDEIFDKSRVINENQVIIKQLERDLNDMKGLVGFYKAFAKKK.

The amino acid sequence of EgG1Y162 protein, also obtained from GenBank, was:

MVLRFCLILLATSVIAEEIRVDPELMAKLTKELKTTLPEHFRWIHVGSRSLELGWNATGLANLHADHIKLTANLYTTYVTFKYRNVPIERQKLTLEGLKPSTFYEVVVQAFKGGSQVFKYTGFIRTLAPGEDGADRASGFALIFAMAGLLLLT.

### 3.2. The physicochemical parameters of the two proteins

By analyzing the physiochemical parameters of the two proteins with the Expasy ProtParam server, we found that the EgA31 protein was composed of 601 amino acids and the molecular weight of the protein was 70,253.31 Da. The theoretical pI value was 5.93. The chemical formula was $C_{3037}H_{5103}N_{871}O_{989}S_{19}$. The instability coefficient was 44.29. The grand average of hydrophilic (GRAVY) was −0.970.

The EgG1Y162 protein was composed of 153 amino acids, and the molecular weight of the protein was 17,117.04 Da. The theoretical pI value was 9.07 and the chemical formula was $C_{788}H_{1247}N_{203}O_{214}S_4$. The extinction coefficient was 18,450. The instability coefficient was 33.64. The grand average of hydrophilic (GRAVY) was 0.205.

### 3.3. Transmembrane domains

The results of TMHMM server analysis showed that the EgA31 and EgG1Y162 proteins did not have a transmembrane domain (Figs. 1 and 2). The transmembrane region of protein was greater than 1, suggesting that it is an extracellular protein.

### 3.4. Secondary structure predictions

The secondary structures of the EgA31 and EgG1Y162 proteins were determined by the DNAstar software Protein module. The predicted secondary structure of the EgA31 protein is shown in Fig. 3 and Table 1, and the secondary structure of EgG1Y162 protein is shown in Fig. 4 and Table 2.

In addition to DNAstar software, we also used SOPMA software to analyze the secondary structures of the two proteins. The software graphically demonstrated the prediction of protein secondary structure and the proportion of various basic structures, including the alpha-helix, extended strand, beta-turn and random coil. The sequence length of the EgA31 protein was 601 amino acids and its secondary structure composition was as follows: 82.36% alpha helix, 4.16% extended strand, 3.16% beta turn and 10.32% random coil. The sequence length of the EgG1Y162 protein was 153 amino acids and its secondary structure composition was as follows: 33.33% alpha helix, 25.49% extended strand, 5.88% beta turn and 35.29% random coil. These parameters of the proteins EgA31 and EgG1Y162 are summarized in Figs. 5 and 6, respectively.

### 3.5. The tertiary structure prediction

The Phyre2 server was employed to build the three-dimensional models of the EgA31 and EgG1Y162 proteins. Phyre2 software predicted the tertiary structure of a protein by folding recognition modeling. After performing the online analysis, template c4uxvA matched to amino acid residues 95–599 (83% of the EgA31 sequence), which was modeled with 97.5% confidence as the single highest scoring template. Template c4pbxA was chosen for residues 10–128 of the EgG1Y162 protein (75% of EgG1Y162 sequence), which had been modeled with 99.8% confidence as the single highest scoring template. The modeling results are shown in Fig. 7.

### 3.6. B-cell epitope predictions

The BCPRED, ABCpred, BepiPred and SVMTriP online server predictions were employed to predict the linear epitopes of the EgA31 and EgG1Y162 proteins. The predicted epitopes of BCPRED were sorted and are shown in Tables 3 and 4. The results of the ABCpred server are shown in Tables 5 and 6. The results of BepiPred are listed in Tables 7 and 8, and the results of SVMTriP are listed in Tables 9 and 10. All the tables listed the top 5 high-score epitopes. The higher score represents a higher possibility as a potential epitope.

### 3.7. T-cell epitope predictions

We chose SYFPEITHI, IEDB (MHC-II) online prediction software to predict the T-cell epitopes of the EgA31 and EgG1Y162 proteins, with the parameters of HLA-DRB1 *0701, HLA-DRB1 *1501 and HLA-DRB1 *0301 to predict CD4+ T cell epitopes. The results are shown in Tables 11 and 12. Then, IEDB(MHC-I) online prediction software was used to predict CD8+ T cell epitopes of the two proteins, with the parameters of HLA-A*1101, HLA_A*0201 and HLA_A*0301. The results are shown in Tables 13 and 14. All the tables listed the top 5 high-scoring epitopes.

### 3.8. The dominant T- and B-cell epitope predictions

Based on the process in the methods section, 11 dominant epitopes (5 for B-cells and 6 for T-cells) and 9 dominant epitopes (3 for B-cells and 6 for T-cell) s were predicted for protein EgA31 and EgG1Y162,

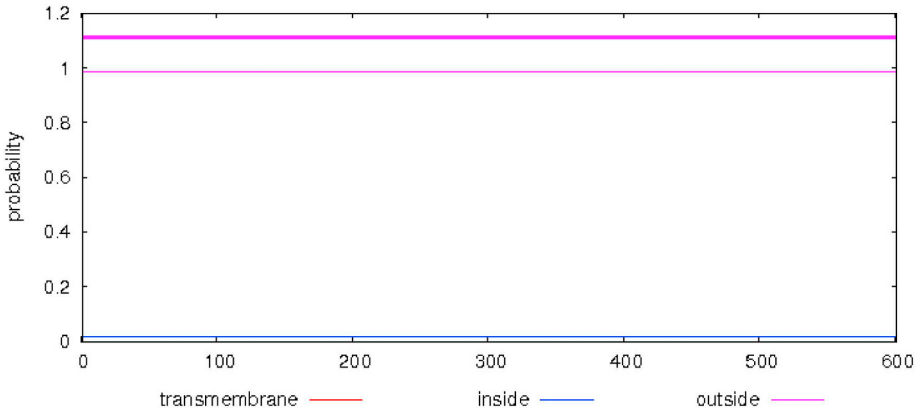


**Fig. 1.** Transmembrane domain of the EgA31 protein.

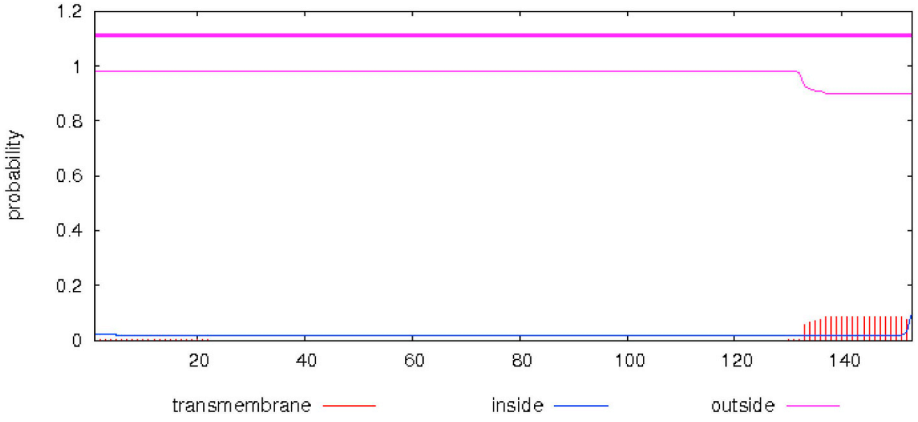


**Fig. 2.** Transmembrane domain of the EgG1Y162 protein.

respectively. Detailed dominant epitopes are listed in Tables 15 and 16.

## 4. Discussion

Humans are often infected with *E. granulosus* through drinking polluted water and eating polluted food (Ma et al., 2016; Tuerxun et al., 2013). After *E. granulosus* infection, the human patient often needs surgical resection and may undergo relapses that bring great economic pressure and physical pain to the family. Therefore, with the recent developments in the fields of technology and bioinformatics, designing novel vaccines consisting of multi-epitopes of a single antigen has become a new approach to elicit both humoral and cellular immunity to enhance the host protective immune response (Khatoon et al., 2017; Pandey et al., 2018). In this study, we used bioinformatics methods to predict the epitopes for designing a multi-epitope vaccine with the ability to elicit specific B- and T-cell immune responses against *E. granulosus* infection.

Previously, we had discovered that there were two antigen candidate proteins of *E. granulosus*, protein EgA31 and EgG1Y162. The EgA31 protein is mainly expressed in the adult stages of *E. granulosus* (Li et al., 2012), while EgG1Y162 protein is mainly expressed in the larval stages of *E. granulosus* (Cao et al., 2009; Ma et al., 2016), and both proteins can stimulate immune responses in vivo. However, there was no information regarding their structures and the antigenic epitopes.

Structurally, proteins are divided into four levels, from low to high: the primary structure is the amino acid sequence, the secondary structure is the local spatial structure formed by the main chain atoms such as the α-helix and β-fold, the tertiary structure is the three-dimensional structure of the protein in space, and the quaternary

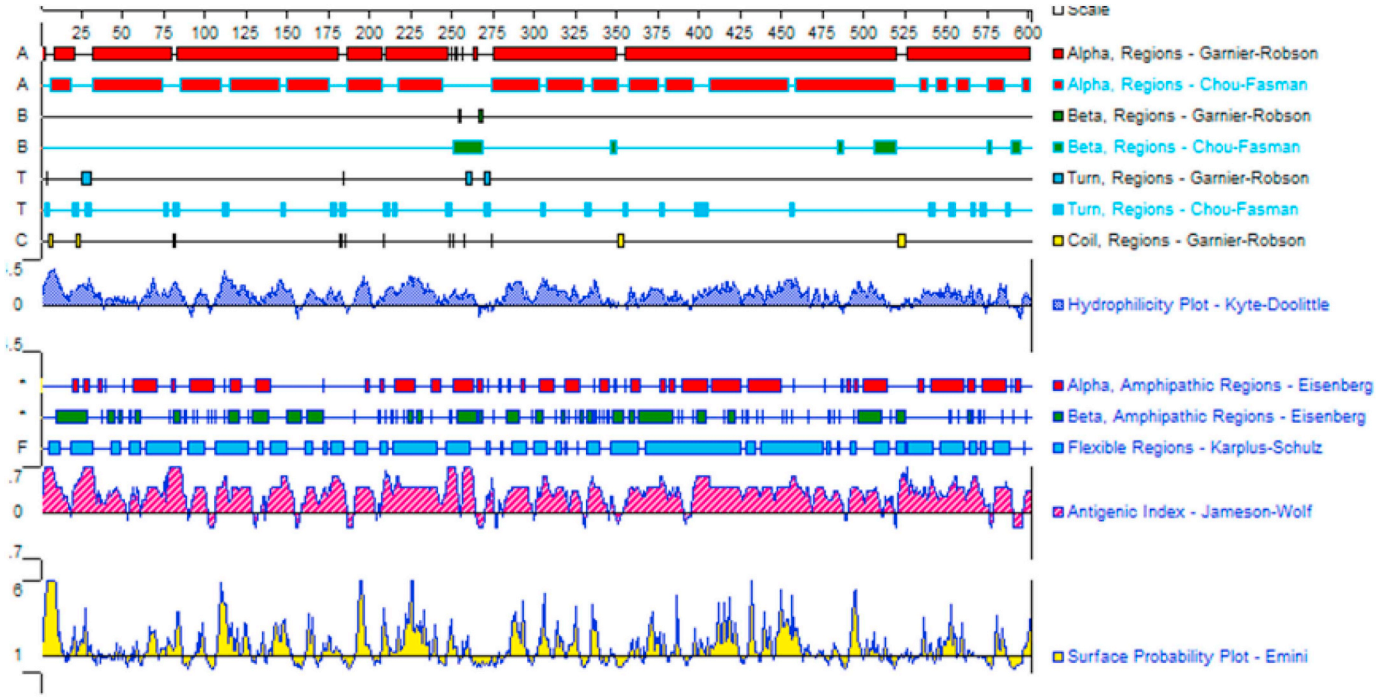


**Fig. 3.** The predicted secondary structure of the EgA31 protein.
Gramier - Robson and Chou - Fasman methods were used to analyze the secondary structure of the EgA31 protein. Lines 1, 3, 5, and 7 are the Gramier - Robson methods; the red represents the alpha helix, the green represents the beta fold, the blue represents the turn, the yellow represents the random coil. Lines 2, 4, 6 are the Chou - Fasman method; the red represents the alpha helix, the green represents the beta fold, the blue for the turn, without random coil prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**
DNAStar Protean predicted the secondary structure of the EgA31 protein.

| Methods | Alpha-helix | Beta-sheet area | Beta turn | Random coil |
|---|---|---|---|---|
| Garnier-Robson | 1–3,9-21,32–80<br>83–181,186–208,<br>210–248, 275–350,<br>355–520, 526–601 | 254–255<br>266–269 | 25–31<br>259–262<br>270–273 | 5–8<br>22–24<br>81–82<br>182–183 |
| Chou-Fasman | 7–9,32–74<br>85–110,115–145<br>150–176,186–208<br>218–244,274–303<br>308–330,335–351<br>358–375,380–397<br>534–539,544–551<br>556–564,575–585<br>596–601 | 251–269<br>347–350<br>484–488<br>506–520<br>575–578<br>590–595 | 3–6,20–23<br>28–31,75–78<br>81–84,111–114<br>146–149, 177–180<br>182–185, 209–212<br>214–217, 247–250<br>331–334, 354–357<br>376–379, 398–405<br>455–458, 540–543,552–555,565–568,571–574,586–589 | |

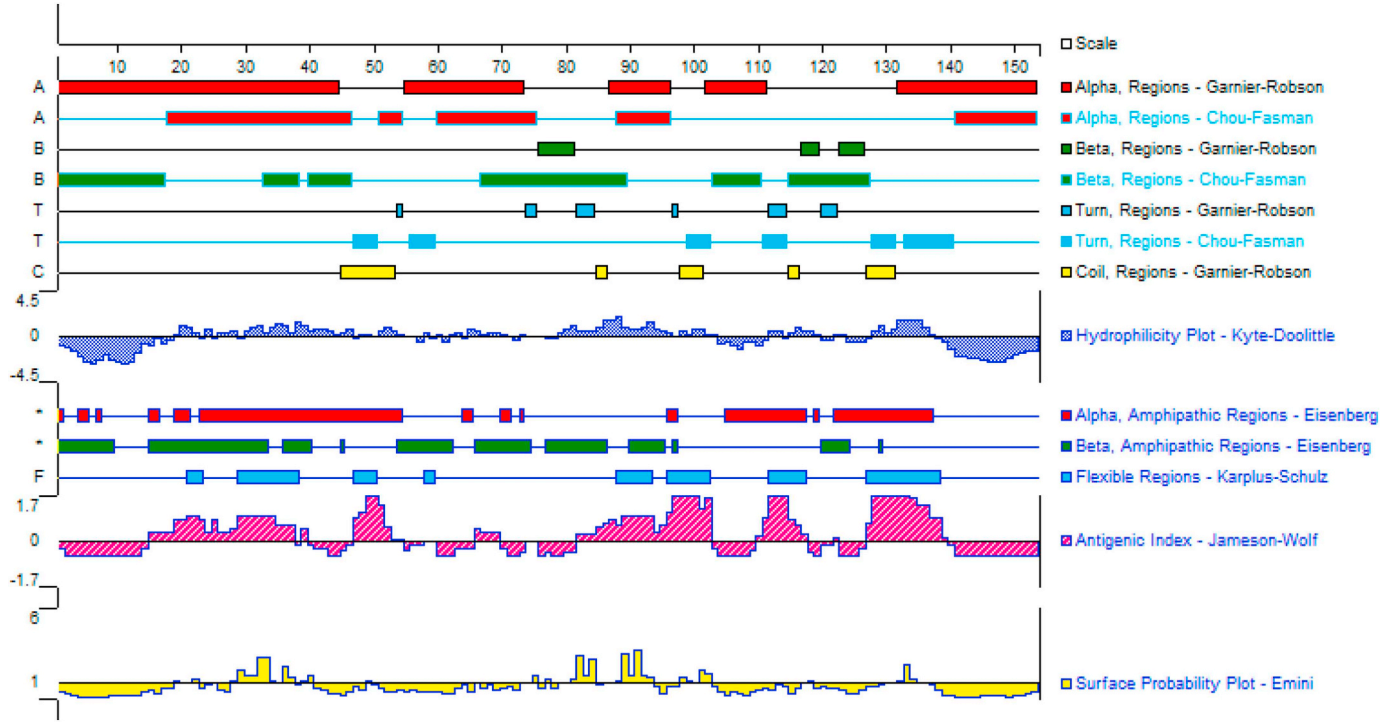


**Fig. 4.** Predicting the secondary structure of the EgG1Y162 protein.
Gramier - Robson and Chou - Fasman methods were used for the prediction of the secondary structure of the EgA31 protein. Lines 1, 3, 5, and 7 are the Gramier - Robson methods; the red represents the alpha helix, the green represents the beta fold, the blue for the turn, the yellow for random coil. Lines 2, 4, 6 are the Chou - Fasman method; the red represents the alpha helix, the green represents the beta fold, the blue for the turn, without random coil prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**
DNAStar Protean predicted the secondary structure of the EgG1Y162 protein.

| Methods | Alpha-helix region | Beta-sheet area | Beta turn | Random coil |
|---|---|---|---|---|
| Garnier-Robson | 1–44<br>55–73<br>87–96<br>102–111<br>132–153 | 76–81<br>117–119<br>123–126 | 74–75<br>82–84<br>112–114<br>120–122 | 45–53<br>98–101<br>127–131 |
| Chou-Fasman | 18–46<br>51–54<br>60–75<br>88–96<br>141–153 | 1–17<br>33–38<br>40–46<br>67–89<br>103–110<br>115–127 | 47–50<br>56–59<br>99–102<br>111–114<br>128–131<br>133–140 | |

structure is the number and arrangement of multiple folded protein subunits in a multisubunit complex (Rehman and Botelho, 2018).

From the ProtParam server results, we discovered various physicochemical properties of the EgA31 protein. Its molecular weight (MW) was 70 kDa, which was useful in SDS-page electrophoresis and western blotting. The theoretical protrusion index was 5.93, which was useful for isoelectric focusing, separation and purification of proteins by ion exchange chromatography. The instability index was estimated to check the stability of the protein. The results showed that the instability index of the EgA31 protein was 44.29, which was classified as an unstable protein (more than 40 indicates protein instability). The grand average of hydrophilic (GRAVY) range was between −2 and 2. The negative value indicates a hydrophilic protein; the GRAVY value of the EgA31 protein was −0.970, which is classified as a hydrophilic protein.

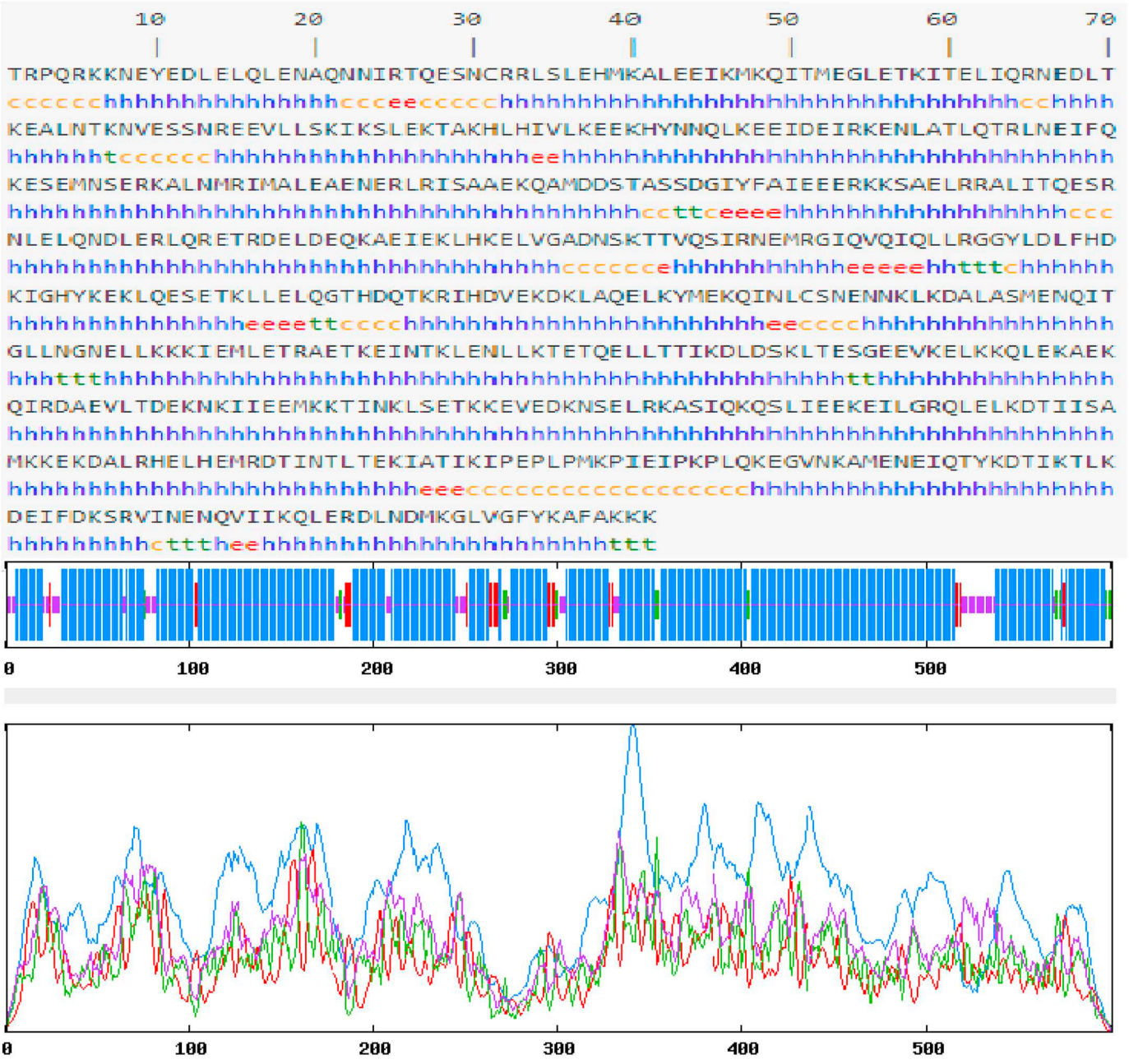


**Fig. 5.** The predicted secondary structure of the EgA31 Protein.
SOPMA server results indicate that the blue H = alpha helix accounted for approximately 82.36%. The red E = beta sheet accounted for approximately 4.16%. The green T = beta turn accounted for approximately 3.16%. The yellow C = random coil accounted for approximately 10.32%. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Compared to the EgA31 protein, the molecular weight of the EgG1Y162 protein was 17 kDa. The instability index was 33.64, classified as stable, and the average hydrophilic coefficient GRAVY was 0.205, classified as a hydrophobic protein. These biochemical parameters would help us to design strategies for the extraction of the proteins and subsequent separation and purification of the proteins in future studies.

We also identified that these two proteins did not have any transmembrane domains and might be fully contacted by antigen-presenting cells to initiate T- and B- cell priming and strong immune responses. The secondary structures of the two proteins were obtained using the DNAstar software and the SOPMA online server. The advantages of the DNAstar Protein Module are to present the secondary structures in a graph with several methods at the same time (Fig. 3). Among the predictions of the secondary structures, we mainly focused on the Garnier-Robson method (calculating the possibility of specific amino acid residues inside a specific structure) and the Chou-Fasman method (predicting protein secondary structures by the crystal structure of the amino acid sequence). According to the characteristics of spatial conformation, the α-helix and β-sheet structures of the secondary structure of the protein are maintained by hydrogen bonds, which are not easily deformed and are mostly located inside the protein. It is difficult to be recognized and bound by antibodies. The β- turns and random coil are mostly located on the surface of the protein as protruding structures (Zhang et al., 2011). The precise prediction by DNAstar software contains the starting point to the end of the amino acid. The proportions of various protein secondary structures were analyzed by the SOMPA

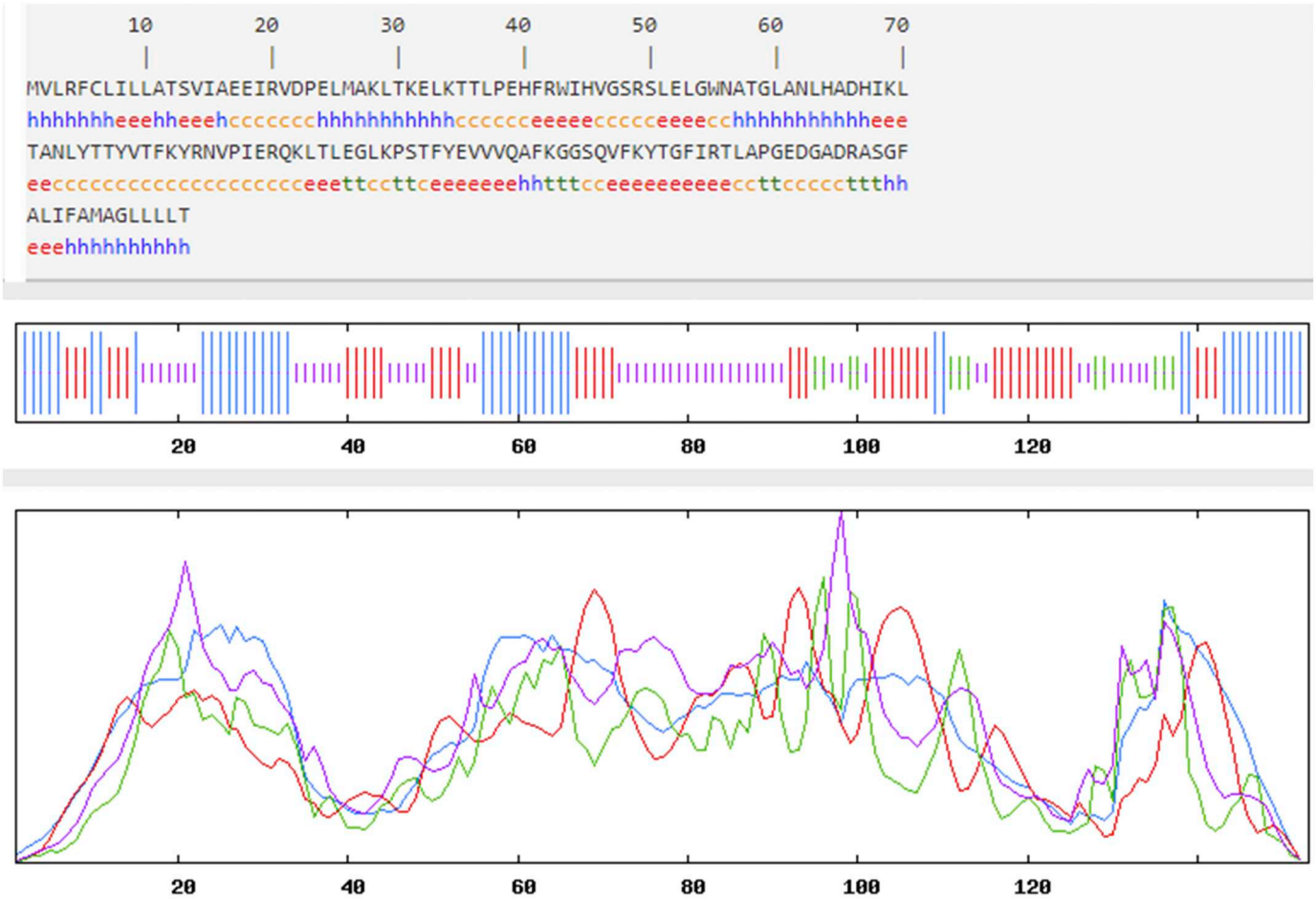


**Fig. 6.** The predicted secondary structure of the EgG1Y162 Protein.
SOPMA server results indicate that the blue H = alpha helix accounted for approximately 33.33%. The red E = extended strand accounted for approximately 25.49%. The green T = beta turn accounted for approximately 5.88%. The yellow C = random coil accounted for approximately 35.29%. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

online server.

The tertiary structure of the two proteins was modeled by the Phyre2 server, which used fold recognition modeling (protein threading) to construct a tertiary structure of proteins (Kelley et al., 2015). The fold recognition modeling, as a method of protein modeling, constructs proteins by using the same fold as known protein structures (Brylinski, 2013). Its principle is to find and determine the possible folding model of the target amino acid sequences by using the known protein folding as a template, and then score a series of folding models, where the highest score is the most likely tertiary structure (Zhu et al., 2018). In our study, the template amino acid of EgA31 was at 97.5 confidence and the template amino acid of EgG1Y162 protein was at 99.8% confidence. Confidence represents the probability (from 0 to 100) of a match between target sequence and template. According to the prediction of secondary structure and tertiary structure modeling, there are multiple random coil structures in the EgA31 and EgG1Y162 proteins, showing great antigenic potential.

*E. granulosus* infection occurs in the host and induces a strong immune response associated with B and T lymphocyte stimulation (Pourseif et al., 2018a, 2018b). B-cell epitopes are divided into linear and conformational epitopes. Although there are several methods to predict the conformational epitopes, we focused on designing a multi-epitope-based vaccine against *E. granulosus*, which elicits humoral and cellular immune responses at the same time. To increase the accuracy of the epitope predictions, we used several different software packages. The BCPREDS Server predicts the linear B-cell epitopes based on the amino acid pair (AAP) antigenicity scale, which reflects some special sequence-coupled features in B-cell epitopes (Chen et al., 2007). Linear epitopes of 20-mer length of the EgA31 and EgG1Y162 proteins were predicted at the default specificity of 75% on the BCPRED server (Tables 3 and 4). The ABCpred Prediction Server predicted linear B-cell epitopes based on a recurrent neural network (Jordan network). The network yielded an overall prediction accuracy of 65.93% when tested by fivefold cross-validation (Saha and Raghava, 2006). B-cell epitopes of both proteins were predicted with a 16-mer length at a 0.51 threshold default on ABCpred (Tables 5 and 6). The BepiPred server predicted B-cell epitopes based on the hidden Markov model with propensity scale methods (Larsen et al., 2006). The SVMTriP Server predicted the B-cell epitopes based on the Tri-peptide similarity and Propensity scores (Yao et al., 2012). Linear epitopes of 20-mer length were predicted at the default specificity of 75% on the BepiPred server (Table 7) and those of 16-mer length were predicted at the 0.51 default threshold on SVMTriP (Tables 9 and 10). By analyzing the results of the four software packages, we found overlapping peptides of predicted linear epitopes, although there were no completely consistent predictions. These overlapping peptides had high scores in different software algorithms. In this study, we selected overlapping peptides as the final dominant epitopes. There were 5 dominant epitopes of 4–11 amino acids, 177–184 amino acids, 223–237 amino acids, 520–533 amino acids, and 582–588 amino acids for the EgA31 protein, and 3 dominant

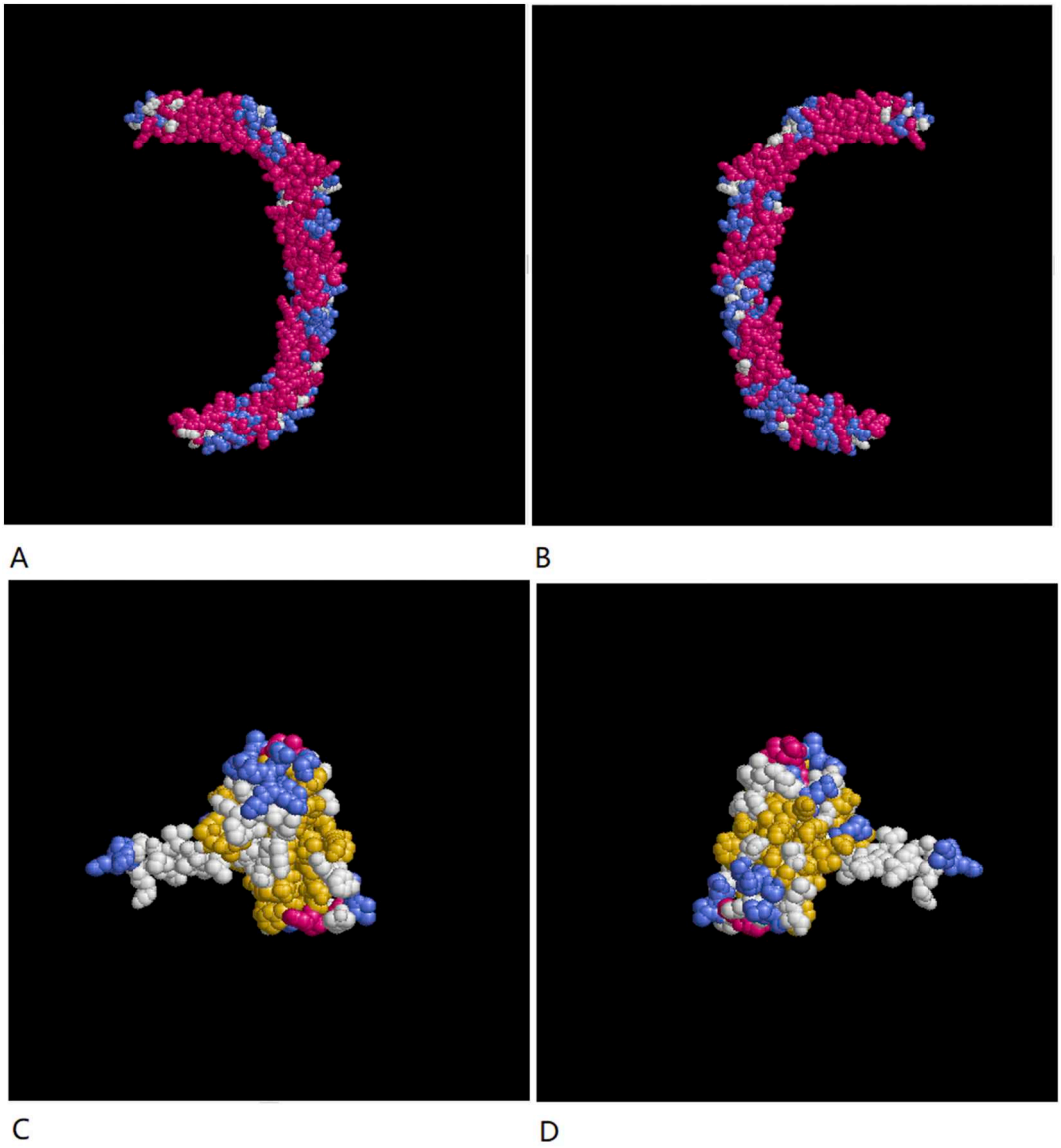


**Fig. 7.** The predicted tertiary structures of the EgA31 and EgG1Y162 proteins.
The prediction of the tertiary structures of the EgA31 and EgG1Y162 proteins was performed by the Phyre2 online server and illustrated by the Raswin software. After selecting the structure mode parameter for the Raswin software, the alpha-helix was marked in red, the beta-sheet in yellow, the random coil in blue, and other residues tags in white.
(A) the front of EgA31 protein, (B) the back of the EgA31 protein (C) the front of EgG1Y162 protein (D) the back of the EgG1Y162 protein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

epitopes of 39–47 amino acids, 96–109 amino acids, and 128–136 amino acids for the EgG1Y162 protein.

Cytotoxic T-lymphocytes (CD8+ CTL) and helper T-lymphocytes (CD4+ Th) play an important role in immune responses against *Echinococcosis* (Apan and Kaygusuz, 2006). Antigenic epitopes of CD4+ T cell were presented by MHC II expressed on the surface of antigen presenting cells (APCs). Once the CD4+ T cell is activated to assist the immune responses of CTL and B-cells, CTL can recognize antigenic epitopes via MHC I expressed on APCs. HLA is a gene complex composed of more than 200 gene loci. It is the most complex gene polymorphism of alleles known to date (Faridi et al., 2018). The distribution frequency of HLA alleles and haploid types varies with different ethnic groups or members of the same ethnic group in different geographic regions. Previous studies showed that the high frequency alleles were HLA-A*1101 (13.46%), A*0201 (12.50%), and A*0301 (10.10%); and HLA-DRB1*0701 (16.35%), DRB1*1501 (8.65%) and DRB1*0301 (7.69%) in an ethnic Uyghur population in the Xinjiang region of China when the distribution of HLA-DRB1 alleles were examined by PCR-SBT (polymerase chain reaction-sequence-based-typing) (Shen et al., 2010). The high incidence of *Echinococcosis* is mainly concentrated in pastoral areas and semi-agricultural and semi-pastoral areas, such as Xinjiang, China. Therefore, in this experiment, combined with SYFPEITHI, IEDB selected HLA-DRB1*0701, DRB1*1501 and DRB1*0301 as the parameter to predict the EgA31 protein and the EgG1Y162 protein (Tables 11 and 12). According to another study, HLA-A*0201 is one of the most prevalent MHC-I alleles, with a frequency over 30% in most populations (Lazarevic and Flynn, 2002). Based on these studies, we analyzed the epitopes of CD8+ T cells using the IEDB online tool (http://tools.iedb.

**Table 3**
Linear B-cell epitopes of EgA31 by BCPREDS.

| Position | Epitope | Score |
|---|---|---|
| 177 | DDSTASSDGIYFAIEEERKK | 0.997 |
| 520 | KIPEPLPMKPIEIPKPLQKE | 0.993 |
| 4 | QRKKNEYEDLELQLENAQNN | 0.991 |
| 461 | RKASIQKQSLIEEKEILGRQ | 0.985 |
| 218 | LERLQRETRDELDEQKAEIE | 0.945 |
| 56 | ETKITELIQRNEDLTKEALN | 0.913 |
| 426 | EVLTDEKNKIIEEMKKTINK | 0.902 |
| 582 | RDLNDMKGLVGFYKAFAKKK | 0.854 |
| 276 | DLFHDKIGHYKEKLQESETK | 0.828 |
| 488 | ISAMKKEKDALRHELHEMRD | 0.82 |
| 344 | SMENQITGLLNGNELLKKKI | 0.818 |
| 244 | VGADNSKTTVQSIRNEMRGI | 0.763 |
| 318 | AQELKYMEKQINLCSNENNK | 0.76 |

**Table 4**
Linear B-cell epitopes of EgG1Y162 by BCPREDS.

| Position | Epitope | Score |
|---|---|---|
| 120 | YTGFIRTLAPGEDGADRASG | 0.997 |
| 96 | EGLKPSTFYEVVVQAFKGGS | 0.95 |
| 28 | KLTKELKTTLPEHFRWIHVG | 0.899 |

**Table 5**
Linear B-cell epitopes of EgA31 by the ABCpred Prediction server.

| Rank | Sequence | Start position | Score |
|---|---|---|---|
| 1 | TIKIPEPLPMKPIEIP | 518 | 0.94 |
| 2 | TTVQSIRNEMRGIQVQ | 251 | 0.90 |
| 3 | NQVIIKQLERDLNDMK | 573 | 0.88 |
| 3 | ENEIQTYKDTIKTLKD | 546 | 0.88 |
| 4 | ETKITELIQRNEDLTK | 56 | 0.87 |
| 4 | DGIYFAIEEERKKSAE | 184 | 0.87 |
| 4 | HIVLKEEKHYNNQLKE | 103 | 0.87 |
| 5 | HEMRDTINTLTEKIAT | 503 | 0.86 |
| 6 | QELLTTIKDLDSKLTE | 388 | 0.84 |
| 6 | TKEINTKLENLLKTET | 372 | 0.84 |
| 6 | SLEHMKALEEIKMKQI | 35 | 0.84 |
| 6 | KYMEKQINLCSNENNK | 322 | 0.84 |

**Table 6**
Linear B-cell epitopes of EgG1Y162 identified by the ABCpred Prediction server.

| Rank | Sequence | Start position | Score |
|---|---|---|---|
| 1 | ANLYTTYVTFKYRNVP | 72 | 0.92 |
| 1 | AEEIRVDPELMAKLTK | 16 | 0.92 |
| 2 | TGFIRTLAPGEDGADR | 121 | 0.85 |
| 3 | TLEGLKPSTFYEVVVQ | 94 | 0.82 |
| 4 | YVTFKYRNVPIERQKL | 78 | 0.81 |
| 5 | PGEDGADRASGFALIF | 129 | 0.80 |
| 6 | HFRWIHVGSRSLELGW | 40 | 0.77 |
| 6 | GSQVFKYTGFIRTLAP | 114 | 0.77 |
| 7 | DHIKLTANLYTTYVTF | 66 | 0.76 |
| 8 | NATGLANLHADHIKLT | 56 | 0.73 |
| 9 | NVPIERQKLTLEGLKP | 85 | 0.70 |
| 10 | DPELMAKLTKELKTTL | 22 | 0.69 |

**Table 7**
Linear B-Cell epitopes of EgA31 protein predicted by the BepiPred 1.0 server.

| Position | Epitope | Score |
|---|---|---|
| 1 | TRPQRKKNEYE | 1.300 |
| 171 | AEKQAMDDSTASSD | 1.089 |
| 446 | LSETKKEVEDKNSEL | 0.939 |
| 223 | RETRDELDEQKAE | 0.887 |
| 399 | SKLTESGEEVKE | 0.859 |
| 74 | LNTKNVESSN | 0.812 |
| 244 | VGADNSKTTVQ | 0.807 |
| 521 | IPEPLPMKPIEIPKPLQKEGVNKAMEN | 0.748 |
| 20 | AQNNIRTQES | 0.655 |
| 302 | THDQTKRIHDV | 0.654 |
| 416 | EKAEKQIR | 0.588 |

**Table 8**
Linear B-cell epitopes of the EgG1Y162 protein predicted by the BepiPred 1.0 server.

| Position | Epitope | Score |
|---|---|---|
| 128–137 | APGEDGADRA | 1.544 |

**Table 9**
Linear B-Cell epitope of the EgA31 protein by SVMTriP.

| Position | Epitope | Score |
|---|---|---|
| 582–601 | RDLNDMKGLVGFYKAFAKKK | 1.000 |
| 296–315 | LLELQGTHDQTKRIHDVEKD | 0.927 |
| 338–357 | LKDALASMENQITGLLNGNE | 0.801 |
| 38–57 | HMKALEEIKMKQITMEGLET | 0.770 |
| 160–179 | EAENERLRISAAEKQAMDDS | 0.630 |
| 243–242 | LVGADNSKTTVQSIRNEMRG | 0.617 |

**Table 10**
Linear B-cell epitope of EgG1Y162 by SVMTriP.

| Position | Epitope | Score |
|---|---|---|
| 39–58 | EHFRWIHVGSRSLELGWNAT | 1.000 |
| 80–99 | TFKYRNVPIERQKLTLEGLK | 0.820 |

org/mhci/) in the context of the HLA-A*1101, HLA_A*0201 and HLA_A*0301 (Tables 13 and 14). After comparing the different results, 6 dominant T-cell epitopes of the EgA31 protein were predicted: 86–97 amino acids, 274–288 amino acids, 286–297 amino acids, 352–359 amino acids, 379–391 amino acids, and 424–435 amino acids, and 6 dominant T-cell epitopes of the EgG1Y162 protein were predicted: 5–16 amino acids, 25–28 amino acids, 34–42 amino acids, 40–51 amino acids, 60–74 amino acids, and 114–128 amino acids.

Epitopes would be potentially useful as an effective vaccine. However, vaccines with only one protein epitope often induce less effective protective immunity because of the complexity of the parasite. For these reasons, our group plans to design a multi-epitope vaccine using the dominant epitopes. We will construct a multi-epitope vaccine that contains dominant epitopes from specific proteins at different growth stages to produce strong immune responses against *E. granulosus* in the future.

**Table 11**
Parameter HLA-DRB1 *0701, HLA-DRB1*1501 and HLA-DRB1*0301 EgA31 protein CD4 + T cell epitope prediction.

| | SYFPEITHI | | | IEDB | | |
|---|---|---|---|---|---|---|
| | Position | Sequence | Score | Position | Sequence | Percentile rank |
| HLA-DRB1*0701 | 379 | LENLLKTETQELLTT | 32 | 261 | RGIQVQIQLLRGGYL | 3.08 |
| | 286 | KEKLQESETKLLELQ | 30 | 283 | GHYKEKLQESETKLL | 3.35 |
| | 538 | KEGVNKAMENEIQTY | 28 | 377 | TKLENLLKTETQELL | 4.19 |
| HLA-DRB1*1501 | 86 | **E**V**L**L**S**KI**K**SL**E**KTA**K** | 24 | 587 | MKGLVGFYKAFAKKK | 0.70 |
| | 272 | **G**G**Y**L**D**L**F**HDKIGHYK | 24 | 263 | IQVQIQLLRGGYLDL | 0.79 |
| | 321 | **L**KY**M**E**K**Q**I**NLCSNEN | 24 | 83 | NREEVLLSKIKSLEK | 8 |
| HLA-DRB1*0301 | 424 | **D**AEV**L**T**D**E**K**NKIIEE | 31 | 421 | QIRDAEVLTDEKNKI | 0.18 |
| | 274 | **Y**LD**L**F**H**DK**I**GHYKEK | 29 | 274 | YLDLFHDKIGHYKEK | 0.89 |
| | 211 | **N**LE**L**Q**N**DLERLQRET | 28 | 577 | IKQLERDLNDMKGLV | 0.91 |

**Table 12**
Parameters HLA-DRB1 *0701, HLA-DRB1*1501 and HLA-DRB1*0301 EgG1Y162 protein T cell epitope prediction.

| | SYFPEITHI | | | IEDB | | |
|---|---|---|---|---|---|---|
| | Position | Sequence | Score | Position | Sequence | Percentile rank |
| HLA-DRB1*0701 | 5 | FCLILLATSVIAEEI | 26 | 37 | LPEHFRWIHVGSRSL | 0.11 |
| | 40 | HFRWIHVGSRSLELG | 26 | 137 | ASGFALIFAMAGLLL | 0.29 |
| | 52 | ELGWNATGLANLHAD | 26 | 2 | VLRFCLILLATSVIA | 1.48 |
| HLA-DRB1*1501 | 114 | **G**S**Q**V**F**K**Y**TGFIRTLA | 34 | 1 | MVLRFCLILLATSVI | 0.78 |
| | 105 | **E**VV**V**QA**F**KGGSQVFK | 28 | 114 | GSQVFKYTGFIRTLA | 1.15 |
| | 5 | **F**CL**I**L**L**ATSVIAEEI | 24 | 65 | ADHIKLTANLYTTYV | 1.30 |
| HLA-DRB1*0301 | 16 | **A**EEI**R**V**D**P**E**LMAKLT | 29 | 16 | AEEIRVDPELMAKLT | 0.03 |
| | 26 | **M**AK**L**T**K**E**L**K**T**TLPEH | 26 | 2 | VLRFCLILLATSVIA | 1.01 |
| | 60 | **E**IR**V**D**P**EL**M**AKLTKE | 22 | 60 | LANLHADHIKLTANL | 1.04 |

**Table 13**
Parameter HLA-A*1101, HLA_A*0201 and HLA_A*0301 EgA31 protein CD8 + T cell epitope prediction.

| | IEDB | | |
|---|---|---|---|
| | Position | Sequence | Percentile rank |
| HLA-A*1101 | 50 | ITMEGLETK | 0.32 |
| | 392 | TTIKDLDSK | 0.47 |
| | 442 | TINKLSETK | 0.77 |
| HLA-A*0201 | 351 | GLLNGNELL | 1.7 |
| | 400 | KLTESGEEV | 1.7 |
| | 151 | ALNMRIMAL | 1.8 |
| HLA_A*0301 | 165 | RLRISAAEK | 0.26 |
| | 352 | LLNGNELLK | 0.36 |
| | 520 | KIPEPLPMK | 0.72 |

**Table 14**
Parameter HLA-A*1101, HLA_A*0201 and HLA_A*0301 EgG1Y162 protein CD8 + T cell epitope prediction.

| | IEDB | | |
|---|---|---|---|
| | Position | Sequence | Percentile rank |
| HLA-A*1101 | 20 | RVDPELMAK | 0.31 |
| | 76 | TTYVTFKYR | 0.52 |
| | 34 | KTTLPEHFR | 1.95 |
| HLA-A*0201 | 144 | FAMAGLLLL | 0.8 |
| | 6 | CLILLATSV | 1.2 |
| | 25 | LMAKLTKEL | 3.0 |
| HLA_A*0301 | 20 | RVDPELMAK | 0.65 |
| | 76 | TTYVTFKYR | 1.6 |
| | 34 | KTTLPEHFR | 4.35 |

**Table 15**
Prediction of T- and B-cell dominant epitopes of the EgA31 protein.

| Episode | Methods | Location | Sequence |
|---|---|---|---|
| B-cell epitope | ABCpred, BepiPred, BCPREDS, SVMTriP | 4–11 | QRKKNEYE |
| | | 177–184 | DDSTASSD |
| | | 223–237 | RETRDELDEQKAEIE |
| | | 520–533 | KIPEPLPMKPIEIP |
| | | 582–588 | RDLNDMK |
| T-cell epitope | SYFPEITHI、IEDB | 86–97 | EVLLSKIKSLEK |
| | | 274–286 | YLDLFHDKIGHYK |
| | | 288–297 | KLQESETKLL |
| | | 352–359 | LLNGNELL |
| | | 379–391 | LENLLKTETQELL |
| | | 424–435 | DAEVLTDEKNKI |

**Table 16**
The T- and B-cell dominant epitope prediction of the EgG1Y162 protein.

| Episode | Methods | Location | Sequence |
|---|---|---|---|
| B-cell epitope | ABCpred, BepiPred, BCPREDS, SVMTriP | 39–47 | EHFRWIHVG |
| | | 96–109 | EGLKPSTFYEVVVQ |
| | | 128–136 | APGEDGADR |
| T-cell epitope | SYFPEITHI、IEDB | 5–16 | FCLILLATSVIA |
| | | 25–28 | LMAK |
| | | 34–42 | KTTLPEHFR |
| | | 40–51 | HFRWIHVGSRSL |
| | | 60–74 | EIRVDPELMAKLTKE |
| | | 114–128 | GSQVFKYTGFIRTLA |

## 5. Conclusion

In conclusion, *E. granulosus* has been reported all over the world and has become a public health emergency of global concern. In this study, we described the basic biophysical features and the secondary and tertiary structures of the EgA31 and EgG1Y162 proteins expressed at different stages of the *E. granulosus* life cycle using several bioinformatics methods. We determined the location of several dominant B-cell and T- cell epitopes in both proteins. These findings will pave the way for designing ideal multi-epitope vaccines against *E. granulosus*.

## Financial support

This study was supported by grants (No. 81460307, No. 31560262) from the National Natural Science Foundation of China and funds for the Xinjiang Key construction Project of the 13th Five-Year Plan (basic medicine).

## Authors' contributions

This study was conceived and designed by Ding Jian-bing. Bioinformatic analysis was performed by Zhao Xiao, Wang Hong-ying, An Meng-ting, and Wulamu Mamuti. The manuscript was drafted by Zhao Xiao and Zhang Feng-bo and edited by Ding Jian-bing.

## Ethics approval and consent to participate

This article does not contain any studies using human participants or animals.

## Conflict of interest

The authors declared no potential conflicts of interest.

## Acknowledgments

Zhao Xiao is thankful to Professor Tianju Liu, the University of Michigan and Professor Buka Samten, the University of Texas Health Science Center, for the revision and all the members of Department of Immunology of the Basic Medical College of Xinjiang Medical University, PR China. Ding Jian-bing is thankful to the State Key Laboratory of Pathogenesis, Prevention, Treatment of Central Asian High Incidence Diseases, The First Affiliated Hospital of Xinjiang Medical University, PR China.

## References

Apan, T.Z., Kaygusuz, S., 2006. Echinococcosis: a possible etiology in psoriatic disease. Dermatology 213, 334–336.
Brylinski, M., 2013. The utility of artificially evolved sequences in protein threading and fold recognition. J. Theor. Biol. 328, 77–88.
Cao, C.B., Ma, X.M., Ding, J.B., Jia, H.Y., Mamuty, W., Ma, H.M., Wen, H., 2009. Cloning and sequence analysis of the egG1Y162 gene of Echinococcus granulosus. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi 27, 177–179.
Chen, J., Liu, H., Yang, J., Chou, K.C., 2007. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 33, 423–428.
Cumino, A.C., Nicolao, M.C., Loos, J.A., Denegri, G., Elissondo, M.C., 2012. Echinococcus granulosus tegumental enzymes as in vitro markers of pharmacological damage: a biochemical and molecular approach. Parasitol. Int. 61, 579–585.
Deleage, G., 2017. ALIGNSEC: viewing protein secondary structure predictions within large multiple sequence alignments. Bioinformatics 33, 3991–3992.
Faridi, P., Li, C., Ramarathinam, S.H., Vivian, J.P., Illing, P.T., Mifsud, N.A., Ayala, R., Song, J., Gearing, L.J., Hertzog, P.J., Ternette, N., Rossjohn, J., Croft, N.P., Purcell, A.W., 2018. A subset of HLA-I peptides are not genomically templated: evidence for cis- and trans-spliced peptide ligands. Sci. Immunol. 3.
Fu, Y., Martinez, C., Chalar, C., Craig, P.S., Ehrlich, R., Petavy, A.F., Bosquet, G., 1999. A new potent antigen from Echinococcus granulosus associated with muscles and tegument. Mol. Biochem. Parasitol. 102, 43–52.
Fu, Y., Saint-Andre, M.I., Marchal, T., Bosquet, G., Petavy, A.F., 2000. Cellular immune response of lymph nodes from dogs following the intradermal injection of a recombinant antigen corresponding to a 66 kDa protein of Echinococcus granulosus. Vet. Immunol. Immunopathol. 74, 195–208.
Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858.
Khatoon, N., Pandey, R.K., Prajapati, V.K., 2017. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. Sci. Rep. 7, 8285.
Lanoix, J., Durette, C., Courcelles, M., Cossette, E., Comtois-Marotte, S., Hardy, M.P., Cote, C., Perreault, C., Thibault, P., 2018. Comparison of the MHC I immunopeptidome repertoire of B-cell lymphoblasts using two isolation methods. Proteomics 18, e1700251.
Larsen, J.E., Lund, O., Nielsen, M., 2006. Improved method for predicting linear B-cell epitopes. Immun. Res. 2, 2.
Lazarevic, V., Flynn, J., 2002. CD8+ T cells in tuberculosis. Am. J. Respir. Crit. Care Med. 166, 1116–1121.
Li, Y.J., Yang, J., Zhao, H., Jia, H.Y., Zhang, L.N., Liu, X.X., Ma, X.M., Wen, H., Ding, J.B., 2012. Bioinformatics prediction of egA31 recombinant antigen epitopes of Echinococcus granulosus. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi 30, 78–80.
Li, J., Bai, X., Liang, Y., Zhang, J., Yang, Y., Zhao, W., Wu, X., 2015. Prediction of epitopes of Rv1410c mycobacterium tuberculosis protein using DNAStar software. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi 31, 474–477.
Ma, X., Zhao, H., Zhang, F., Zhu, Y., Peng, S., Ma, H., Cao, C., Xin, Y., Yimiti, D., Wen, H., Ding, J., 2016. Activity in mice of recombinant BCG-EgG1Y162 vaccine for Echinococcus granulosus infection. Hum. Vaccin. Immunother. 12, 170–175.
Moller, S., Croning, M.D., Apweiler, R., 2001. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17, 646–653.
Ostrov, D.A., Alkanani, A., McDaniel, K.A., Case, S., Baschal, E.E., Pyle, L., Ellis, S., Pollinger, B., Seidl, K.J., Shah, V.N., Garg, S.K., Atkinson, M.A., Gottlieb, P.A., Michels, A.W., 2018. Methyldopa blocks MHC class II binding to disease-specific antigens in autoimmune diabetes. J. Clin. Invest. 128, 1888–1902.
Otero-Abad, B., Torgerson, P.R., 2013. A systematic review of the epidemiology of echinococcosis in domestic and wild animals. PLoS Negl. Trop. Dis. 7, e2249.
Ozturk, G., Aydinli, B., Yildirgan, M.I., Basoglu, M., Atamanalp, S.S., Polat, K.Y., Alper, F., Guvendi, B., Akcay, M.N., Oren, D., 2007. Posttraumatic free intraperitoneal rupture of liver cystic echinococcosis: a case series and review of literature. Am. J. Surg. 194, 313–316.
Pandey, R.K., Bhatt, T.K., Prajapati, V.K., 2018. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. Sci. Rep. 8, 1125.
Pourseif, M.M., Moghaddam, G., Naghili, B., Saeedi, N., Parvizpour, S., Nematollahi, A., Omidi, Y., 2018a. A novel in silico minigene vaccine based on CD4(+) T-helper and B-cell epitopes of EG95 isolates for vaccination against cystic echinococcosis. Comput. Biol. Chem. 72, 150–163.
Pourseif, M.M., Moghaddam, G., Saeedi, N., Barzegari, A., Dehghani, J., Omidi, Y., 2018b. Current status and future prospective of vaccine development against Echinococcus granulosus. Biologicals 51, 1–11.
Qian, M.B., Zhou, X.N., 2018. Walk together to combat echinococcosis. Lancet Infect. Dis. 18, 946.
Rammensee, H.G., Friede, T., Stevanoviic, S., 1995. MHC ligands and peptide motifs: first listing. Immunogenetics 41, 178–228.
Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., Stevanovic, S., 1999. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50, 213–219.
Rehman, I., Botelho, S., 2018. Biochemistry, Tertiary Structure, Protein.
Saha, S., Raghava, G.P., 2006. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65, 40–48.
Shen, C.M., Zhu, B.F., Deng, Y.J., Ye, S.H., Yan, J.W., Yang, G., Wang, H.D., Qin, H.X., Huang, Q.Z., Zhang, J.J., 2010. Allele polymorphism and haplotype diversity of HLA-A, -B and -DRB1 loci in sequence-based typing for Chinese Uyghur ethnic group. PLoS One 5, e13458.
Tuerxun, Z., Yimiti, D., Cao, C.B., Ma, H.M., Li, Y.J., Zhou, X.T., Zhu, M., Ma, X.M., Wen, H., Ding, J.B., 2013. Construction and expression of the Echinococcus granulosus recombinant BCG-EgG1Y162. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi 31, 110–113.
Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D., Hochstrasser, D.F., 1999. Protein identification and analysis tools in the ExPASy server. Methods Mol. Biol. 112, 531–552.
Yao, B., Zhang, L., Liang, S., Zhang, C., 2012. SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. PLoS One 7, e45152.
Zhang, J., Shang, Z., Zhang, X., Zhang, Y., 2011. Modeling and analysis of Schistosoma Argonaute protein molecular spatial conformation. Asian Pac. J. Trop. Biomed. 1, 275–278.
Zhang, F., Ma, X., Zhu, Y., Wang, H., Liu, X., Zhu, M., Ma, H., Wen, H., Fan, H., Ding, J., 2014. Identification, expression and phylogenetic analysis of EgG1Y162 from Echinococcus granulosus. Int. J. Clin. Exp. Pathol. 7, 5655–5664.
Zhang, F., Li, S., Zhu, Y., Zhang, C., Li, Y., Ma, H., Pang, N., An, M., Wang, H., Ding, J., 2018. Immunization of mice with egG1Y162-1/2 provides protection against Echinococcus granulosus infection in BALB/c mice. Mol. Immunol. 94, 183–189.
Zhu, J., Wang, S., Bu, D., Xu, J., 2018. Protein threading using residue co-variation and deep learning. Bioinformatics 34, i263–i273.