

Combining proteomics and bioinformatics to explore novel tegumental antigens as vaccine candidates against *Echinococcus granulosus* infection

Sebastián Miles¹ | Madelón Portela² | Marek Cyrklaff³ | María Eugenia Ancarola⁴ | Friedrich Frischknecht³ | Rosario Durán² | Sylvia Dematteis¹ | Gustavo Mourglia-Ettlin¹ 

¹Área Inmunología, DEPBIO/IQB, Facultad de Química/Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

²Unidad de Bioquímica y Proteómica Analíticas, Institut Pasteur de Montevideo and IIBCE, Montevideo, Uruguay

³Integrative Parasitology, Center for Infectious Diseases, Heidelberg University, Heidelberg, Germany

⁴Instituto de Investigaciones en Microbiología y Parasitología Médica (IMPAM, UBA-CONICET), Facultad de Medicina, Universidad de Buenos Aires, Buenos Aires, Argentina

Correspondence

Gustavo Mourglia-Ettlin, PhD, Área Inmunología, DEPBIO/IQB, Facultad de Química/Facultad de Ciencias, Universidad de la República, Montevideo 11600, Uruguay.
Email: gmourglia@higiene.edu.uy

Abstract

Echinococcus granulosus is the parasite responsible for cystic echinococcosis (CE), an important worldwide-distributed zoonosis. New effective vaccines against CE could potentially have great economic and health benefits. Here, we describe an innovative vaccine design scheme starting from an antigenic fraction enriched in tegumental antigens from the protoscolex stage (termed PSEx) already known to induce protection against CE. We first used mass spectrometry to characterize the protein composition of PSEx followed by Gene Ontology analysis to study the potential Biological Processes, Molecular Functions, and Cellular Localizations of the identified proteins. Following, antigenicity predictions and determination of conservancy degree against other organisms were determined. Thus, nine novel proteins were identified as potential vaccine candidates. Furthermore, linear B cell epitopes free of posttranslational modifications were predicted in the whole PSEx proteome through colocalization of in silico predicted epitopes within peptide fragments identified by matrix-assisted laser desorption/ionization-TOF/TOF. Resulting peptides were termed “clean linear B cell epitopes,” and through BLASTp scanning against all nonhelminth proteins, those with 100% identity against any other protein were discarded. Then, the secondary structure was predicted for peptides and their corresponding proteins. Peptides with highly similar secondary structure respect to their parental protein were selected, and those potentially toxic and/or allergenic were discarded. Finally, the selected clean linear B cell epitopes were mapped within their corresponding 3D-modeled protein to analyze their possible antibody accessibilities, resulting in 14 putative peptide vaccine candidates. We propose nine novel proteins and 14 peptides to be further tested as vaccine candidates against CE.

KEYWORDS

Echinococcus granulosus, immunoinformatics, linear B cell epitopes, MALDI-TOF/TOF, tegument, vaccine

1 | INTRODUCTION

Cystic echinococcosis (CE) is an important zoonotic disease, currently listed as a Neglected Tropical Disease by the World Health Organisation (WHO). *Echinococcus granulosus*—the parasite responsible for CE—is a cestode that requires two mammalian hosts for completing its life cycle: canids (mainly dogs) act as definitive hosts, whereas farm or wild ungulates are usually the intermediate hosts. Humans are considered accidental intermediate hosts, and the WHO estimates that more than 1 million people worldwide are infected at any one time.¹ Intermediate hosts become infected by ingesting infective eggs containing oncospheres, which migrate to various visceral organs, such as liver and lungs, and develop into the metacestode stage (hydatid cyst). When hydatid cysts are fertile, they asexually produce protoscoleces (PSC) turning into the infective stage for definitive hosts.² Interestingly, if a fertile cyst accidentally ruptures within an intermediate host, spilled PSC have the developmental plasticity to develop into new cysts in a process called secondary infection.³

Efforts to mitigate CE worldwide have focused on controlling the parasite life cycle by using antihelminthic drugs, by promoting hygienic slaughter and by improving health education in rural communities. Reported costs for controlling CE is over US\$ 3 billion annually, and despite those efforts, the number of human cases is substantially increasing worldwide every year.⁴ In the last decades, efforts in designing effective vaccines against *E. granulosus* infection have increased, focusing on both intermediate and definitive hosts, but showing different and inconsistent success. Up to date, only one vaccine against CE (named EG95) has reached the market. EG95 contains a recombinant 17 kDa protein derived from *E. granulosus* oncospheres, and induces over 95% protection in sheep.^{5,6} However, the EG95 vaccine antigen—originally developed for the G1 parasite genotype—has been shown to be polymorphic and immunologically different in other parasite genotypes and/or isolates, raising the question whether the current EG95 vaccine would be effective against every *E. granulosus* genotype.^{7,8}

Currently, there are two main approaches for designing vaccines. Originally, most vaccines were created by immunization with the whole pathogen, and—when possible—using some purified antigens responsible for the overall protection. More recently, after the concept of reverse vaccinology was proposed, new vaccine candidates are selected from the whole pathogen genome/proteome by *in silico* tools.⁹ Both approaches have pros and cons, and currently approved vaccines have been designed following both strategies. Additionally, strategies for peptide-based (or epitope-based) vaccines can complement

both abovementioned approaches through finding the smallest part of an antigen retaining protective activity. The primary advantage of this approach is that peptides can be produced in nonbiological synthetic ways, reducing production costs enormously. Peptide mapping (or epitope mapping) can be performed *in vitro* or *in silico*, being the second option less accurate although virtually costless.¹⁰ In this regard, peptide-based vaccines for helminth-derived infections have also been reported.^{11–15} In the case of CE, two studies using different approaches to obtain peptide-based vaccines derived from the EG95 antigen have been reported. In the case of Woollard et al,¹⁶ they studied four different EG95-derived peptides, which all successfully induce anti-EG95 antibodies in immunized sheep, although no oncospheres-killing activity was observed *in vitro*. In the other study, Read et al¹⁷ worked with an EG95-derived mimotope (a linear peptide that mimics a conformational B cell epitope) named E-100, which showed high efficiency in inducing antibodies able to kill oncospheres *in vitro*.

In the present work, we identified novel protein candidates to be further tested as vaccine antigens, and we also proposed a new strategy for designing peptide-based candidates through the integration of proteomics, systems vaccinology, and bioinformatics. To that end, we first performed a proteomic characterization of a PSC-derived fraction composed of tegumental antigens, and then applied a systematic bioinformatic and immunoinformatic pipeline. After that, novel proteins were proposed as vaccine candidates against *E. granulosus*. Additionally, through an innovative bio-informatic workflow, we identified peptides for being further tested in peptide-based vaccines, which represent linear B cell epitopes most likely to be accessible to antibodies and free of posttranslational modifications within the native protein. Their sequences were studied in depth and peptides were selected for their specificity, for not being neither toxic nor allergenic, and for retaining a secondary structure similar to the native peptide within their parental protein, thus improving the likelihood of inducing protective cross-reacting responses after immunization.

2 | MATERIALS AND METHODS

2.1 | Parasites and antigens

E. granulosus PSC were collected by opening fertile hydatid cysts from natural bovine infections with a scalpel and they were washed several times with phosphate-buffered saline (PBS) pH 7.2 containing antibiotics (penicillin 60 µg/mL, streptomycin 100 µg/mL, and amphotericin-B 250 ng/mL). Tegumental proteins were extracted from highly viable PSC (viability ≥ 80%) using

an extracting solution consisting of the detergent MEGA-10 (1% w/v), EDTA (5 mM) and phenylmethylsulfonyl fluoride (2 mM) in PBS according to Hernández and Nieto.¹⁸ Briefly, 125,000 viable PSC/mL of extracting solution were incubated for 2 hours at room temperature with soft constant shaking. Then, PSC were allowed to settle down and the supernatant was removed and extensively dialyzed at room temperature against PBS through a dialysis tube cellulose membrane (MW cut-off: 12 000 Da). Obtained antigens (termed PSEx) were characterized in terms of lipid, protein, and carbohydrate content according to,¹⁹ and stored at -20°C until used. Treated PSC were washed three times with PBS, and their integrity was assessed by observation under a light microscope staining with eosin to improve contrast.

2.2 | Transmission electron microscopy

Aliquots of PSC pre- and post-extraction were fixed with 2.5% glutaraldehyde and 2% paraformaldehyde in PBS and stored at 4°C until use. After changing fixing buffer to cacodylate buffer (50 mM cacodylate, 50 mM KCl, 2.6 mM CaCl_2 , 2.6 mM MgCl_2 , and 2% sucrose), PSC samples were incubated with 1% osmium tetroxide in 100 mM cacodylate buffer for 60 minutes at room temperature, washed two times with 100 mM cacodylate buffer and two times with EM-grade water and incubated overnight at 4°C in 1% uranyl acetate in water. Samples were rinsed two times with water, dehydrated in a graded acetone series (from 30% to 100%) at room temperature before being embedded in epoxy resin, and polymerized for at least 48 hours at 60°C . Ultrathin sections of 70 nm were obtained by sectioning PSC with a Leica EM UC6 ultramicrotome (Leica Microsystems, Wetzlar, Germany) and a diamond knife. Sections were collected on EM grids, counterstained using lead citrate and uranyl acetate, and examined with a Zeiss 10C transmission electron microscope (80 kV; Carl Zeiss, Goettinger, Germany).

2.3 | 2D sample preparation and electrophoresis

PSEx antigens were resolved by 2D electrophoresis as described elsewhere.²⁰ Briefly, 300 μg of proteins were incubated overnight at -20°C in 300 μL of ice-cold acetone containing 20% of trichloroacetic acid and 0.07% dithiothreitol (DTT). After centrifugation for 15 minutes at 10 000g and 4°C , the supernatant was discarded and 300 μL of ice-cold acetone containing 20% dimethylsulfoxide and 0.07% DTT was added and incubated for 60 minutes at -20°C . Then, the sample was centrifuged

for 15 minutes at 10,000 g and 4°C , the supernatant was discarded and 300 μL of ice-cold acetone containing 0.07% DTT was added. This step was repeated twice. Finally, the pellet was lyophilized, resuspended in IPG buffer and frozen at -80°C for 24 hours to improve solubilization. For the first dimension, 7 cm linear pH gradient (pH 3-10) Immobiline DryStrips (GE Healthcare, Darmstadt, Germany) were rehydrated with the sample and run on an IPGphor isoelectric focusing system (9.5 hours run and a total of 35,500 Vh), and stored at -80°C until use. Strips were then equilibrated in equilibration buffer (50 mM, Tris-Cl pH 8.8; urea, 6 M; glycerol, 30%; sodium dodecyl sulfate (SDS), 2%; and traces of bromophenol blue) containing DTT (10 mg/mL) for 15 minutes, further equilibrated in equilibration buffer containing iodoacetamide (25 mg/mL) for 15 minutes, and sealed to 10% acrylamide gels using 0.5% agarose in standard Tris-glycine electrophoresis buffer. Second dimension SDS-PAGE was run at 50 V for the first 15 minutes and then raised to 150 V until ending. Finally, replicates of 2D gels were stained with either Coomassie blue dye or mass-spectrometry compatible silver nitrate staining.

2.4 | Mass spectrometry

Spots obtained from both Coomassie blue and silver nitrate-stained 2D gels were analyzed by matrix-assisted laser desorption/ionization (MALDI)-TOF/TOF mass spectrometry. First, spots were excised, faded and tryptic digestions were performed using sequencing-grade modified trypsin (Promega, Madison, WI). After gel extraction into 60% acetonitrile containing 0.1% TFA, the excess of acetonitrile was removed by speed vacuum. Peptide samples were then combined with an equal volume of matrix and spotted onto a MALDI sample plate and allowed to dry at room temperature. Mass spectra were acquired on a MALDI-TOF/TOF mass spectrometer (4800 MALDI TOF/TOF Mass Analyzer, Abi Sciex, Warrington, UK) operating in the positive ion reflector mode. Protein identifications were performed using an in-house Mascot v.2.3 search engine by searching a custom database that includes the full proteome of *E. granulosus* and *Echinococcus multilocularis*, composed of 20 787 sequences (10 310 548 residues) obtained from the Sanger Helminth Database (<http://www.sanger.ac.uk/resources/downloads/helminths/>). Additionally, every mass spectrum was also analyzed using NCBI database to discard possible host related proteins. The search criteria used were cysteine carbamidomethylation and methionine oxidation as variable modification, and mass deviation < 200 ppm with peptide fragment tolerance of 0.45 Da. Scores > 56 were considered significant ($P < 0.05$).

2.5 | Gene ontology analysis

A global comprehension of PSEx was obtained by analyzing the relationship between all protein-related Gene Ontology (GO). In the first place, GO terms were collected from UniProtKB database for each protein. Since protein annotation was incomplete in terms of GO codes, further terms were assigned to each protein using the software FFPred3, which predicts the functions based on the amino acid sequence and the relationship between biophysical attributes (secondary structure, transmembrane helices, intrinsically disordered regions, and signal peptides among other motifs) and protein functions.²¹ Only terms with high storage virtual machine (SVM) reliability and $P > 0.8$ for Biological Process and Molecular Function, and $P > 0.7$ for Cellular Component were included. If no terms reached the abovementioned thresholds, the highest probability term was included. Due to software limitations, only proteins less than 1500 amino acids in length were analyzed. Finally, the proteome GO analysis was performed by REVIGO²² and visualized as pie charts showing the top represented “super-clusters” of loosely related terms. For this, all the GO terms were submitted to REVIGO and set to use SimRel as the semantic similarity measure and allowing similarities of 0.9.

2.6 | In silico proteome characterization

An extensive in silico characterization of MALDI-TOF/TOF identified proteins was performed. First, each protein cellular sublocalization was predicted by DeepLoc software.²³ Then, individual and global antigenicity was assessed by predictions with ANTIGENpro software,²⁴ and abundance of antigenic regions (AAR) was predicted using the web server Secret-AAR.²⁵ Proteins with ANTIGENpro scores ≥ 0.5 were considered as highly antigenic.²⁶ Finally, we obtained a “uniqueness score” (US) of each protein by BlastP, where we defined the US as the product between query coverage and identity percentages of the highest max score match against all the proteins on the nonredundant NCBI database (excluding Cestoda, Trematoda, and Nematoda taxonomy). US values of 0.5 to 0.8 identified “highly divergent” proteins, while US values ≤ 0.5 corresponded to “unique” proteins.

2.7 | Clean linear B cell epitopes

Linear B cell epitopes were predicted in each protein using the algorithm Bepipred (version 2.0).²⁷ Continuous

antigenic segments more than or equal to six amino acids in length were considered as linear B cell epitopes. Epitopes were further colocalized on the matching peptides used to identify the corresponding protein by MALDI-TOF/TOF, assuming those regions are free of posttranslational modifications (PTMs). Linear B cell epitopes that completely overlapped with those regions or included the N-terminal Lys/Arg of the trypsin-cut peptides, were termed “clean linear B cell epitopes.”

2.8 | Peptide quality control

Sequences identified as clean linear B cell epitopes were evaluated in terms of safety before proposing them as peptide-based vaccine candidates. A pipeline was designed to avoid unwanted-cross reactivity, toxicity, and/or allergenicity. In addition, a study of secondary structure similarity (SSS) was performed to compare the possible secondary structure of synthetic peptides with the corresponding sequence on the native protein. For this, we first discarded every peptide exhibiting a 100% similarity (BlastP) with any protein on the nonredundant NCBI database (excluding Cestoda, Trematoda, and Nematoda taxonomy). Then, secondary structures of clean linear B cell epitopes and their corresponding proteins were predicted using SOPMA software,²⁸ and the percentage of amino acids in the same conformation (SSS) was calculated. Only peptides with $SSS \geq 80\%$ were selected. Finally, peptides predicted to be toxic and/or—if longer than 11 amino acids—allergenic were excluded from further analyses. Such predictions were performed using Toxinpred²⁹ and AlgPred³⁰ softwares, respectively.

2.9 | 3D modeling and epitope mapping

Clean linear B cells epitopes that successfully passed the above-mention pipeline were mapped within the 3D structure of their corresponding protein to analyze the antibody accessibility according to.²⁶ Briefly, we first performed a 3D structure modeling by low-homology protein threading using RaptorX software³¹ to every protein that exhibited at least one peptide that successfully passed the previous pipeline. Only 3D models with uGDT (unnormalize Global Distance Test) higher than 50 were taken into account, and for proteins with more than one predicted 3D model, the structure with the highest uGDT was chosen. Then, clean linear B cell epitopes were mapped within each selected structure using SWISS-PdbViewer,³² and their antibody accessibility was evaluated.

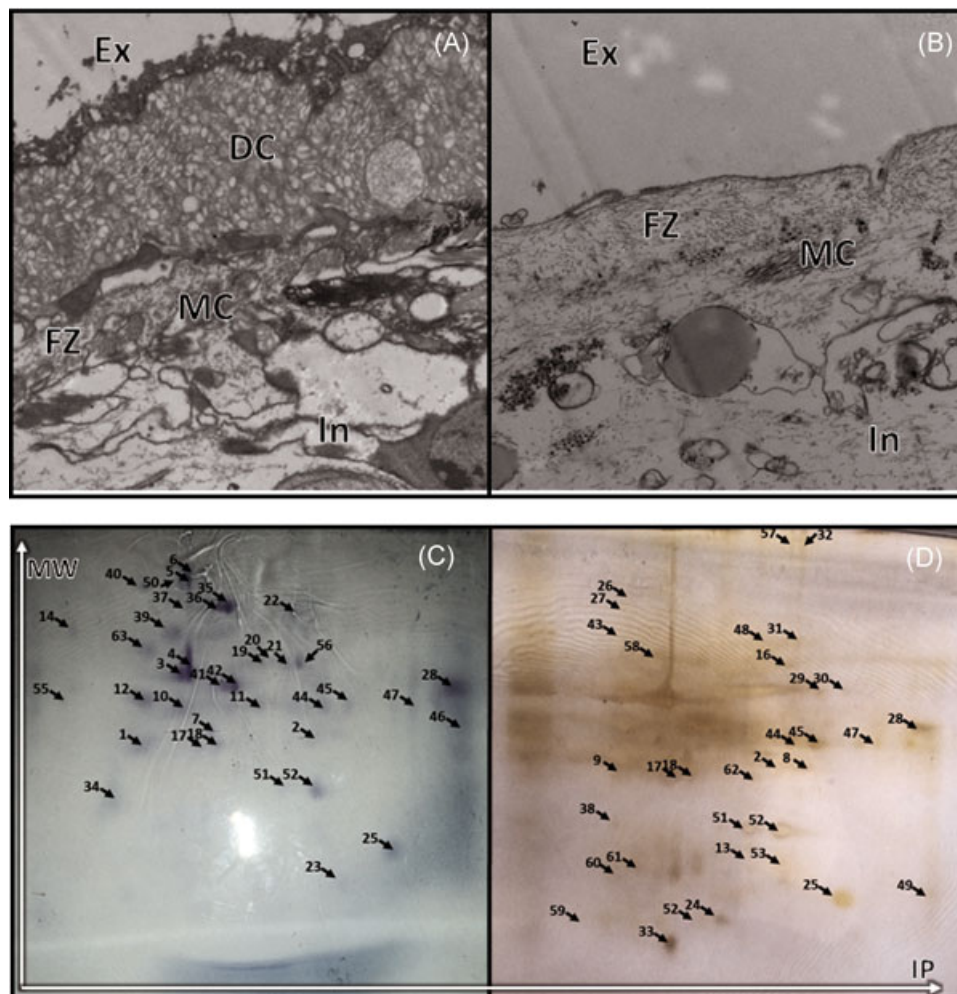


FIGURE 1 Tegumental ultrastructure and 2D-electrophoresis. Micrographs (top row) depict TEM images of pre-PSEx-extracted (A) and post-PSEx-extracted (B) PSC at $\times 6000$. The Ex, FZ, MC, and In of the parasite are shown in (A) and (B), as well as the DC in (A). 2D-PAGE (bottom row) show 2DE of PSEx stained with Coomassie blue (C) or silver nitrate (D). Gels show proteins typically less than 120 kDa with IP values in the linear range 3 to 10 (from left to right). Arrows in gels indicate localization of proteins identified by mass spectrometry, and numbering correspond to the description in Table 1. DC, distal cytoplasm; Ex, exterior; FZ, fibrous zone; In, interior; MC, muscle cell; PAGE, polyacrylamide gel; PSE, protoscolex; TEM, transmission electron microscopy

3 | RESULTS

3.1 | Structural morphology of PSC

Tegument morphology in PSC pre-extraction (Figure 1A) and post-extraction (Figure 1B) was analyzed in 70 nm thick sections by transmission electron microscopy (TEM). A comparative study between them showed that after protein extraction, the general parasite structure remained mostly intact, while a complete disruption of tegumental distal cytoplasm was observed (Figure 1B). In fact, while distal cytoplasm and tegumental spines almost completely disappeared, the fibrous zone, muscle cells and inner structures were still observed with little or no changes. This suggests that the PSEx fraction was composed mostly of proteins

present in classical tegumental structures, with few internal parasite proteins as contaminants.

3.2 | PSEx proteome

Biochemical composition of PSEx showed protein-to-carbohydrates and protein-to-lipids mass ratios of 7.2 and 1.3, respectively; suggesting that PSEx is an antigenic fraction enriched in proteins and lipids. Therefore, an extensive proteomic analysis of PSEx was performed by 2D gel electrophoresis followed by mass spectrometry (MALDI-TOF/TOF). To that end, after removal of lipids and highly hydrophobic proteins from PSEx, 2D SDS-PAGE (linear pH 3%-10% and 10% acrylamide) were run and stained with either Coomassie blue (Figure 1C) or silver nitrate (Figure 1D). After that, 96 spots were

TABLE 1 Proteomic characterization of PSEx

Number	Spot	Protein name	Accession number	MW (Da)	IP	Length (aa)	MASCOT score	Fragmented peptides	Consensus coverage, %	Previously identified
1	1	14-3-3 Protein	Q56J98	27 928	4.91	247	155	3	72	
2	2	3-Oxoacyl acyl carrier protein reductase	A0A068WAT3	34 833	5.63	326	145	2	42	37
3	3/4/5/6/7	Actin	W6UMH9	41 718	5.30	393	105	2	64	–
4	3/4/5/6	Actin cytoplasmic A3	U6JLF5	41 789	5.31	375	138	2	64	37
5	3/4/5/6/7	Actin, cytoplasmic 2	W6UFA4	40 189	5.66	377	105	2	64	–
6	3/4/5/6/7	Actin-2	Q03341	41 856	5.37	376	104	2	37	–
7	8	Aldo keto reductase family 1 member B4	A0A068WU90	42 259	8.58	375	185	2	22	–
8	9	Ankyrin	A0A068WVL6	89 159	8.11	823	66	1	20	–
9	10	Annexin	A0A068WJF9	36 659	5.17	323	329	4	54	36,37
10	10	Annexin	A0A068WRH5	34 805	5.38	310	77	3	59	37
11	11	Annexin	A0A068WUR8	38 733	4.82	348	315	3	58	37
12	12	Annexin	U6JIM9	37 280	5.89	331	75	1	11	–
13	13	Basement membrane specific heparan sulfate	A0A068X0X0	859 349	5.01	7809	69	1	24	35
14	14	Calreticulin	A5YTY7	45 339	4.48	395	95	1	39	36
15	15	Cathepsin d lysosomal aspartyl protease	A0A068WAA1	47 109	5.59	423	57	1	13	–
16	16	Citrate synthase	U6JOK5	51 378	8.26	465	231	3	56	35,36
17	17/18	Endophilin B1	A0A068WMU2	28 782	5.83	252	245	3	67	20,34,35,37
18	19/20/21/22	Enolase	D0VLV3	46 532	6.48	433	126	2	59	33
19	23	Expressed protein	A0A068WLY9	22 367	9.55	192	96	2	32	37,39
20	24	Fatty acid binding protein FABP2	A0A068WLT8	23 901	8.41	213	64	1	62	34,37
21	25	Fatty acid binding protein FABP2	U6JF28	15 361	5.64	133	86	1	23	–
22	26/27	Filamin	A0A068WYB5	276 623	5.37	2558	62	1	16	–
23	28	Fructose-bisphosphate aldolase	U6IXV3	39 641	8.31	363	210	3	77	
24	29/30	Gelsolin	U6IX85	41 998	7.63	364	101	2	47	36
25	31/32	Glucose-6-phosphate isomerase	A0A068WLH0	61 566	6.48	547	61	1	20	35
26	33	Glyceraldehyde-3-phosphate dehydrogenase	A0A068WKQ2	37 739	8.55	352	108	2	46	34,35,37

(Continues)

TABLE 1 (Continued)

Number	Spot	Protein name	Accession number	MW (Da)	IP	Length (aa)	MASCOT score	Fragmented peptides	Consensus coverage, %	Previously identified
27	34	Heat shock 70 kDa protein	A0A068X4W2	28 849	5.80	252	61	1	23	–
28	35/36/37/38	Heat shock 70 kDa protein 4	A0A068WHV0	72 510	5.59	664	208	3	62	–
29	35/36	Heat shock 70 kDa protein 4	U6JIC8	71 039	5.65	651	60	1	75	34,36,38
30	39	Heat shock protein 60	A0A068WSG0	60 575	5.37	568	261	3	59	37
31	35/36	Heat shock protein 71 kDa protein	A0A068X560	72 526	5.52	665	204	3	63	–
32	40	Heat shock protein HSP 90α	W6UM50	73 829	5.03	738	180	2	40	36
33	41/42	Major egg antigen p40	U6JBW8	35 825	5.81	314	99	1	71	34,36,37
34	5/43	Major vault protein	A0A068WUZ9	97 000	5.33	869	64	1	27	37
35	44/45/46	Malate dehydrogenase	A0A068WYB8	35 519	8.82	338	63	1	17	35,37
36	44/47	Malate dehydrogenase	A0A068X1L3	36 547	7.62	332	74	1	24	34,35,37
37	48	Mechanosensory protein 2	A0A068WM93	53 528	8.21	484	59	1	32	–
38	49	Nucleoside diphosphate kinase	A0A068W1L30	16 676	10.16	149	101	2	21	40
39	5/6/50	Paramyosin	A0A068WA67	98 882	5.29	863	69	1	60	34,37
40	51/52/53	Peptidyl-prolyl cis-trans isomerase	P14088	17 343	6.41	162	232	3	35	20,33,37,38
41	54	Peptidylprolyl isomerase	A0A068X276	7885	9.98	65	64	1	33	–
42	55	Phosphatase 2a inhibitor i2pp2a	A0A068WLN2	34 168	4.35	300	86	1	19	–
43	56/57	Phosphoenolpyruvate carboxykinase	A0A068X1L0	71 188	6.93	635	57	1	58	34,35,37
44	58	Protein disulfide-isomerase	A0A068W1L3	81 413	7.86	716	64	1	41	39
45	59	SPRY domain containing protein	A0A068WNG3	136 955	5.92	1223	154	2	14	–
46	48	Structural maintenance of chromosomes protein	A0A068WEW7	149 088	6.21	1347	71	1	9	36
47	54	Superoxide dismutase [Cu-Zn]	W6Y988	40 189	5.66	177	100	2	33	34,35,37
48	60/61	Thioredoxin peroxidase	Q8T6C4	21 392	5.78	193	79	1	55	19,31-34,41
49	2/62	Triosephosphate isomerase	U6JN02	27 122	6.60	250	54	1	78	32-34,41
50	63	Tubulin β chain	A0A068WEL5	49 630	4.79	443	93	1	33	–
51	63	Tubulin β chain	A0A068WLA0	52 862	4.88	472	68	1	37	–
52	63	Tubulin β chain	A0A068WQC9	57 714	5.36	509	71	1	41	–
53	63	Tubulin β chain	U6JIN6	49 802	4.79	445	100	2	49	–
54	63	Tubulin β chain	U6JM68	49 584	4.79	444	104	2	41	–

(Continues)

TABLE 1 (Continued)

Number	Spot	Protein name	Accession number	MW (Da)	IP	Length (aa)	MASCOT score	Fragmented peptides	Consensus coverage, %	Previously identified
55	63	Tubulin $\beta/\beta 2$ chain	A0A068WML9	49 841	4.75	445	293	4	67	36,37
56	60	Uncharacterized protein	U6JRE4	50 538	9.32	444	71	1	33	—
57	48	UTP--glucose-1-phosphate uridylyltransferase	W6UT22	57 047	6.31	533	123	2	27	38
58	59	Vesicular amine transporter	A0A068X307	49 068	4.99	443	124	2	36	35,39

Abbreviations: aa, amino acid; IP, isoelectric point.; MW, molecular weight; IP, isoelectric point.

The full PSEx proteome was obtained through MALDI-TOF/TOF. "Spot" column shows the corresponding spot from 2DE gels (Figure 1C and 1D). "Accession number" refers to the UniProtKB IDs. "MASCOT score" corresponds to the lowest score for proteins identified in several spots. "Fragmented peptides" corresponds to the lowest number of matched MS/MS peptides. "Consensus coverage" is the percentage of the sum of all peptides used to identify the same protein in several spots.

excised from gels (50 spots from Coomassie blue stained gels and 46 spots from silver stained gels), and then they were faded and trypsin-digested for mass spectrometry analysis. Finally, 58 different proteins were successfully identified within 63 spots (Table 1). Among these, 27 proteins were identified only in Coomassie blue-stained gels, 20 proteins only in silver-stained gels, and 11 proteins were found in spots excised from both gels. In some spots, more than one protein—exhibiting similar molecular weights and isoelectric points—were identified. Since, in that case, it is virtually impossible to differentiate such proteins in a 2D gel, all of them were included in the final proteome. At the end, 22 new proteins were described for the first time in this study, while the remaining 36 have been already reported in previous *E. granulosus* proteomic analyses (Table 1). Interestingly, no host-derived proteins (ie *Bos taurus*) were identified.

3.3 | PSEx Gene Ontology

For each identified protein, GO terms were first obtained from their corresponding annotation in UniProtKB database. Then, the prediction of complementary terms was performed for each protein using FFPred3 software. Due to software limitations, only proteins less than 1500 amino acids long were analyzed. From annotated proteins in UniProtKB database, 197 GO terms were obtained, which did not cover the three categories (Biological Process, Molecular Function, and Cellular Component) for most proteins. After GO prediction using FFPred3 software, 544 additional terms were obtained. Finally, by combining both GO terms sources—and removing duplicates between annotated and predicted terms—711 terms corresponding to 56 proteins were obtained. Among them, 184 unique GO terms were found, which were reduced into a two-level hierarchy of GO terms resulting in 67, 79, and 33 unique Biological Processes, Molecular Functions, and Cellular Components, respectively. Pie charts show top represented loosely related terms joined into "superclusters."²² From this analysis, PSEx components could globally be described as mostly vesicle-related proteins involved in microtubule-based processes with diverse metabolic functions (Figure 2).

3.4 | In silico PSEx characterization

To obtain a protein-by-protein characterization of PSEx, an extensive in silico study of its identified proteome was performed (Table 2). Prediction of cellular localization showed a predominance (73%) of cytoplasmic proteins (42 in 57), followed by 11% characterized as mitochondrial

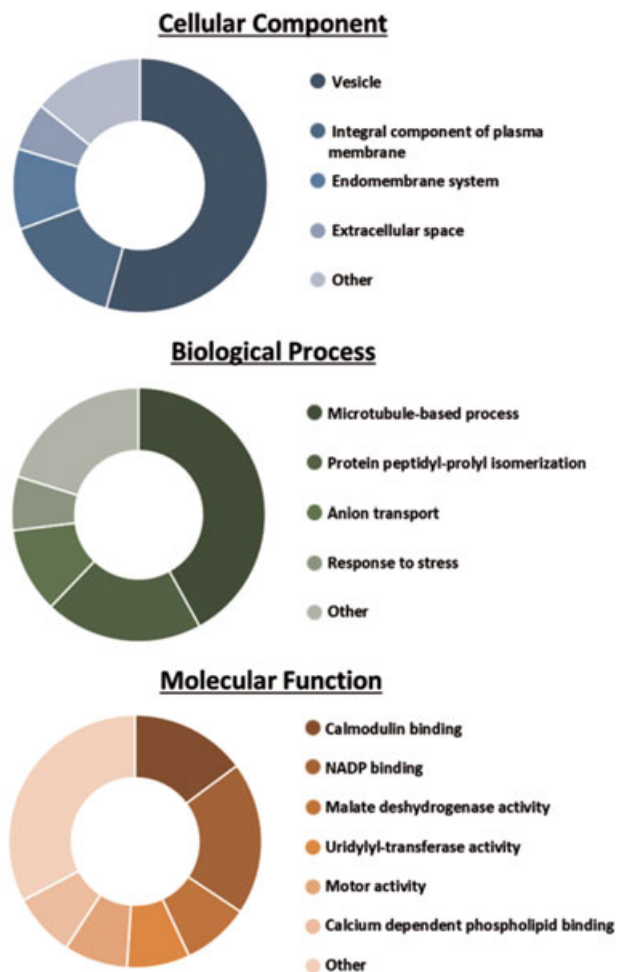


FIGURE 2 Gene Ontology study of PSEx proteome. GO analysis of PSEx proteins is shown for Cellular Component, Biological Process, and Molecular functions. For each category, the unique terms were reduced into a two-level hierarchy of GO terms, and visualized as pie-charts showing top represented loosely related terms joined into “superclusters.” GO, Gene Ontology; PSEx, protoscolex

proteins (6 in 57), and the remaining 16% classified as belonging to other cellular or extracellular compartments (9 in 57). Additionally, the overall antigenicity of PSEx proteome was determined. In this sense, ANTIGENpro obtained a mean antigenicity score of 0.67, while a mean AAR value of 41.0 was calculated through Secret-AAR. Proteins with ANTIGENpro scores > 0.50 are usually considered highly antigenic,²⁶ and according to this threshold, 74% of proteins within PSEx proteome (43 in 58) belonged to that category (Table 2). Finally, a US was calculated for each protein, resulting in 40% of “highly divergent” proteins (23 in 58), and 34% of “unique” proteins (20 in 58) within PSEx proteome (Table 2). Interestingly, similar values were obtained when considering only first time-identified proteins: 0.68 and 40.0 mean values for antigenicity score and AAR, respectively.

Furthermore, 27% of such proteins (5 in 22) were categorized as “highly divergent”, while 36% (8 in 22) were “unique” (Table 2).

3.5 | Identification of clean linear B cell epitopes

Proteins within PSEx proteome were scanned for linear sequences potentially recognizable by antibodies (eg, linear B cell epitopes). Due to software limitations, only proteins less than 2500 amino acids in length (56 out of 58) were analyzed. Results showed 578 lineal B cell epitopes within the PSEx proteome (Figure 3), with a mean value of 10 epitopes per protein. Every protein showed at least two linear B cell epitopes, and two proteins (#37 and #46) exhibited 33 epitopes, the largest number observed. Then, in the interest of identifying clean linear B cell epitopes, for example those PTMs-free, we evaluated the colocalization of every obtained epitope over the sequence fragments used to identify proteins by mass spectrometry. Since the mass of those fragments is practically equal to the overall sum of their constituent amino acids, we can assume that such regions are free of PTMs. After colocalization, 146 clean linear B cell epitopes were predicted in 46 proteins within PSEx proteome (Figure 3). Those epitopes ranged from six to 46 amino acids in length, with a mean value of 12 amino acids per epitope. Then, aiming to avoid unwanted cross-reactivity, the 146 sequences were scanned against all nonhelminth proteins by BlastP, and those peptides matching with 100% identity to any protein were discarded (Figure 3). Interestingly, only peptides ranging from 9 to 46 amino acids in length were identified among the 50 epitopes selected, which belonged to 29 different proteins (Figure 3). After that, and in pursuance of maximizing the likelihood of selecting peptides able to induce cross-reacting antibodies against their parental native proteins, SSS values were calculated for those 50 selected epitopes. Results showed that 17 peptides—belonging to 15 different proteins—exhibited SSS scores $\geq 80\%$ (Figure 3), being 16 and 1 of them mostly random coil and α -helix in secondary structure, respectively. Finally, those sequences were further analyzed for safety considerations, removing peptides predicted to be toxic and/or allergenic. Only one peptide was discarded in this step, leaving 16 peptide candidates from 15 different proteins present in PSEx (Figure 3). Summing up, from the initially 578 peptides identified (derived from an input of 56 proteins), only 16 of them (roughly 3%) belonging to 15 different proteins successfully passed through our rationally designed pipeline. Interestingly, four of these proteins were identified for the first time in this study (#3, #8, #45, and #53),

TABLE 2 Immunobioinformatic characterization of PSEx

Number	Protein name	Accession number	Localization	Antigenicity	AAR	US
1	14-3-3 Protein	Q56J98	Cytoplasm	0.76	41.3	0.62
2	3-Oxoacyl acyl carrier protein reductase	A0A068WAT3	Peroxisome	0.10	65.4	0.40
3	Actin	W6UMH9	Cytoplasm	0.43	32.8	0.93
4	Actin cytoplasmic A3	U6JLF5	Cytoplasm	0.39	47.0	0.91
5	Actin, cytoplasmic 2	W6UFA4	Cytoplasm	0.35	34.4	0.97
6	Actin-2	Q03341	Cytoplasm	0.35	47.1	0.90
7	Aldo keto reductase family 1 member B4	A0A068WU90	Cytoplasm	0.63	62.7	0.40
8	Ankyrin	A0A068WVL6	Cytoplasm	0.76	37.5	0.28
9	Annexin	A0A068WJF9	Cytoplasm	0.50	46.3	0.43
10	Annexin	A0A068WRH5	Cytoplasm	0.85	31.1	0.46
11	Annexin	A0A068WUR8	Cytoplasm	0.72	24.9	0.42
12	Annexin	U6JIM9	Cytoplasm	0.30	47.4	0.34
13	Basement membrane specific heparan sulfate	A0A068X0X0	NP	NP	27.9	0.05
14	Calreticulin	A5YTY7	ER	0.90	49.5	0.51
15	Cathepsin d lysosomal aspartyl protease	A0A068WAA1	Vesicle	0.26	42.4	0.50
16	Citrate synthase	U6J0K5	Mitochondrion	0.43	35.8	0.65
17	Endophilin B1	A0A068WMU2	Mitochondrion	0.94	42.2	0.24
18	Enolase	D0VLV3	Cytoplasm	0.53	31.0	0.88
19	Expressed protein	A0A068WLY9	Extracellular	0.74	21.4	0.12
20	Fatty acid binding protein FABP2	A0A068WLT8	Cytoplasm	0.52	42.8	0.61
21	Fatty acid binding protein FABP2	U6JF28	Cytoplasm	0.67	33.5	0.74
22	Filamin	A0A068WYB5	Cytoplasm	NP	25.8	0.52
23	Fructose-bisphosphate aldolase	U6IXV3	Cytoplasm	0.75	33.1	0.73
24	Gelsolin	U6IX85	Cytoplasm	0.82	33.2	0.32
25	Glucose-6-phosphate isomerase	A0A068WLH0	Cytoplasm	0.80	49.8	0.67
26	Glyceraldehyde-3-phosphate dehydrogenase	A0A068WKQ2	Cytoplasm	0.80	29.4	0.71
27	Heat shock 70 kDa protein	A0A068X4W2	Cytoplasm	0.96	36.1	0.67
28	Heat shock 70 kDa protein 4	A0A068WHV0	Cytoplasm	0.92	35.0	0.82
29	Heat shock 70 kDa protein 4	U6JIC8	Cytoplasm	0.90	34.3	0.84
30	Heat shock protein 60	A0A068WSG0	Mitochondrion	0.66	33.5	0.79
31	Heat shock protein 71 kDa protein	A0A068X560	Cytoplasm	0.93	35.1	0.82
32	Heat shock protein HSP 90α	W6UM50	Cytoplasm	0.73	38.9	0.74
33	Major egg antigen p40	U6JBW8	Mitochondrion	0.74	31.5	0.28
34	Major vault protein	A0A068WUZ9	Cytoplasm	0.42	39.5	0.56
35	Malate dehydrogenase	A0A068WYB8	Mitochondrion	0.43	37.7	0.63
36	Malate dehydrogenase	A0A068X1L3	Cytoplasm	0.45	41.6	0.63
37	Mechanosensory protein 2	A0A068WM93	Cell membrane	0.58	69.3	0.40
38	Nucleoside diphosphate kinase	A0A068WL30	Cytoplasm	0.24	150	0.43
39	Paramyosin	A0A068WA67	Cytoplasm	0.87	34.6	0.73
40	Peptidyl-prolyl cis-trans isomerase	P14088	Cytoplasm	0.92	32.6	0.77
41	Peptidylprolyl isomerase	A0A068X276	Cytoplasm	0.89	22.0	0.58
42	Phosphatase 2a inhibitor i2pp2a	A0A068WLN2	Nucleus	0.85	50.2	0.38
43	Phosphoenolpyruvate carboxykinase	A0A068X1L0	Cytoplasm	0.60	35.3	0.61
44	Protein disulfide-isomerase	A0A068WIL3	Cytoplasm	0.80	37.7	0.27
45	SPRY domain containing protein	A0A068WNG3	Cytoplasm	0.94	35.0	0.11

(Continues)

TABLE 2 (Continued)

Number	Protein name	Accession number	Localization	Antigenicity	AAR	US
46	Structural maintenance of chromosomes protein	A0A068WEW7	Nucleus	0.66	42.1	0.41
47	Superoxide dismutase [Cu-Zn]	W6V988	Cytoplasm	0.86	35.6	0.58
48	Thioredoxin peroxidase	Q8T6C4	Cytoplasm	0.67	64.7	0.65
49	Triosephosphate isomerase	U6JN02	Cytoplasm	0.62	35.9	0.81
50	Tubulin β chain	A0A068WEL5	Cytoplasm	0.83	37.0	0.97
51	Tubulin β chain	A0A068WLA0	Plastid	0.84	43.0	0.86
52	Tubulin β chain	A0A068WQC9	Cytoplasm	0.73	42.5	0.78
53	Tubulin β chain	U6JJN6	Cytoplasm	0.84	37.2	0.97
54	Tubulin β chain	U6JM68	Cytoplasm	0.78	37.1	0.96
55	Tubulin $\beta/\beta 2$ chain	A0A068WML9	Cytoplasm	0.76	37.2	0.96
56	Uncharacterized protein	U6JRE4	Mitochondrion	0.46	37.1	NH
57	UTP-glucose-1-phosphate uridylyltransferase	W6UT22	Cytoplasm	0.69	59.3	0.58
58	Vesicular amine transporter	A0A068X307	Cell membrane	0.72	24.7	0.17

Abbreviations: AAR, abundance of antigenic region; ER, endoplasmic reticulum; NH, no hit protein obtained by BLASTp; NP, not performed due to software limitations; US, uniqueness score.

The full PSEx proteome was characterized through several immunobioinformatic tools. “Localization” column shows the predicted subcellular localization of each protein. “Antigenicity” and “AAR” columns show the predicted values for antigenicity and abundance of antigenic regions, respectively. “US” column shows the uniqueness score calculated for each protein. NH, no hit protein obtained by BLASTp (indicating query coverage and identity percentage smaller than 25%), thus a US value of 0.00 was assigned. Details of software settings are described in Section 2.

7 corresponded to “unique” proteins (#8, #11, #17, #33, #44, #45, and #58), and four were “highly divergent” proteins (#1, #14, #16, and #23). Moreover, two of them (#8 and #45) were “unique” as well as highly antigenic,

and were identified for the first time in this study (Tables 1 and 2).

3.6 | 3D modeling and epitope mapping

3D models of those 15 proteins exhibiting at least one clean linear B cell epitope that overcame our previous analysis, were obtained using RaptorX software. In this regard, among the 15 obtained models not a clear predominant tertiary structure was observed (Figure 4). Then, to analyze epitope surface accessibility, selected peptides were mapped in their 3D-modeled parental protein as described elsewhere.²⁶ Only models with uGDT ≥ 50 were selected, resulting in 14 good models (Figure 4). Regarding “SPRY domain containing protein” (#45 in our set), its 3D model did not reach the uGDT threshold and it was discarded from accessibility analysis (Figure 4P). Then, by means of SWISS-PdbViewer, we assessed the surface accessibility to antibodies, obtaining 14 accessible epitopes. One of them, corresponding to “Heat shock 70 kDa protein 4” (#29 in our set), although mapped on a good 3D model (uGDT = 470), it located on a highly disordered portion of the predicted model, and therefore it was discarded (Figure 4O). Finally, we rationally identified 14 accessible clean linear B cell epitopes—belonging to 13 different proteins present in PSEx—which deserve further testing as vaccine candidates against *E. granulosus* infection (Table 3). Interestingly, among the parental proteins of such epitopes, 3 of them belonged to the set of 22 proteins identified for the

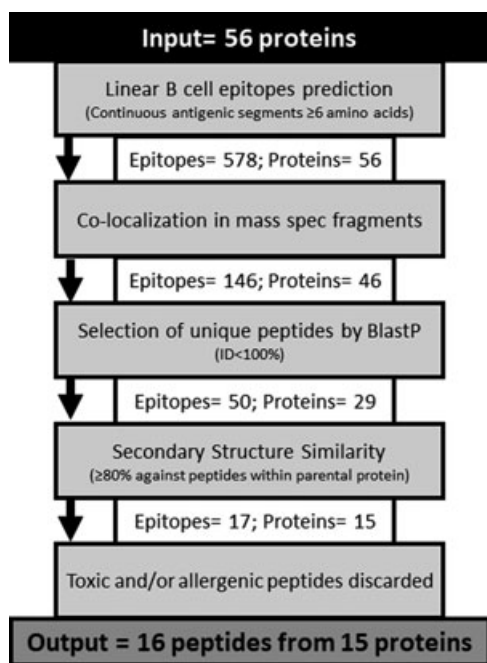


FIGURE 3 In silico prediction of peptide-based vaccine candidates. Bioinformatic workflow designed and used for predicting antigenic peptides derived from experimentally identified proteins within PSEx. Details on software settings are described in Section 2

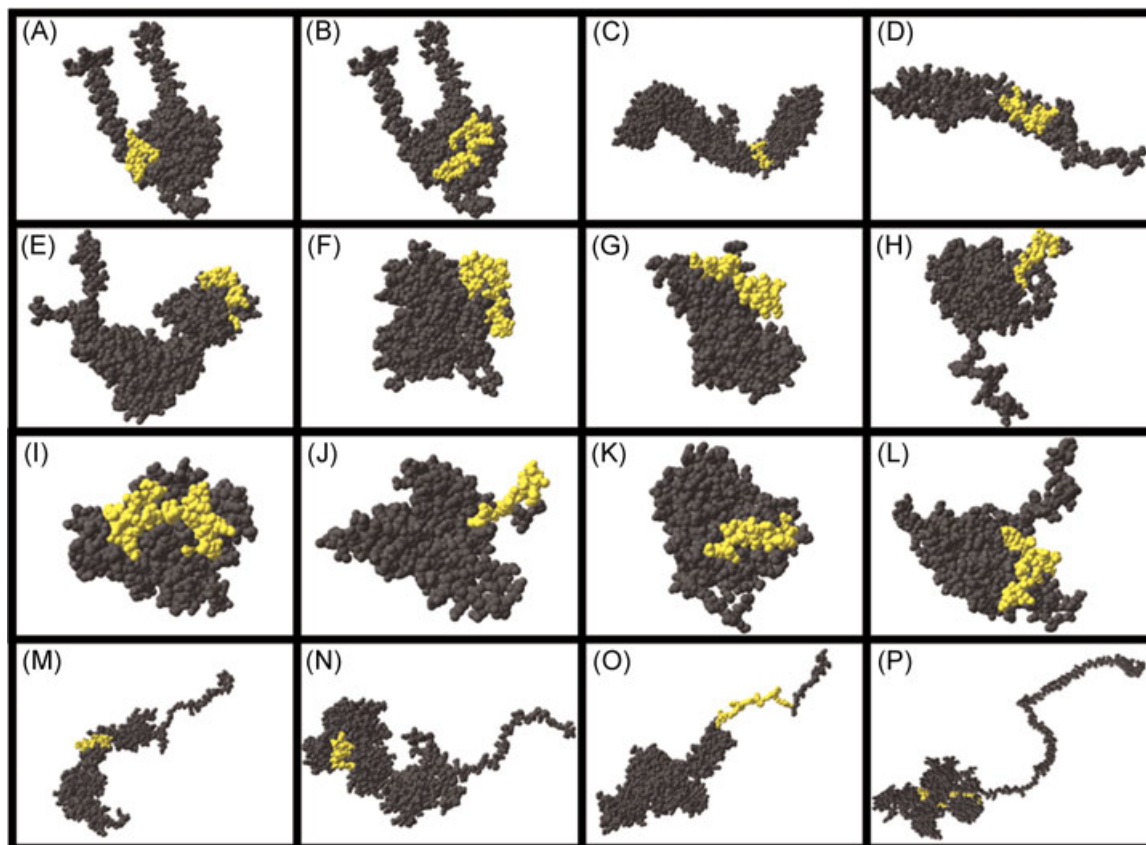


FIGURE 4 Mapping of clean linear B cell epitopes in 3D-modeled protein structures. For every protein that exhibited at least one selected peptide, a 3D-modeled structure was predicted by means of RaptorX software, and the clean linear B cell epitopes were visualized in the 3D models using DeepView. For visualization purposes, 3D overall structure are represented in gray, while mapped epitopes are depicted in yellow. The corresponding proteins are: A,B, calreticulin (#14); C, ankyrin (#8); D, endophilin B1 (#17); E, heat shock protein 60 (#30); F, actin (#3); G, annexin (#11); H, citrate synthase (#16); I, tubulin β chain (#53); J, 14-3-3 protein (#1); K, fructose-bisphosphate aldolase (#23); L, major egg antigen p40 (#33); M, vesicular amine transporter (#58); N, protein disulfide-isomerase (#44); O, heat shock 70 kDa protein 4 (#29); P, SPRY domain containing protein (#45)

first time in this study (#3, #8, and #53). Additionally, 4 (#1, #14, #16, and #23) and 6 (#8, #11, #17, #33, #44, and #58) proteins were classified as “highly divergent” or “unique”, respectively (Table 2).

4 | DISCUSSION

Due to the increasingly high economic and health costs of CE, new effective vaccines are needed for primary and/or secondary infections. The only currently marketed vaccine was designed for primary infection of sheep against parasite G1 genotype, but its performance against different *E. granulosus sensu lato* genotypes has to be analyzed.^{7,8,42} Several studies have shown promising results for new vaccines against CE, although a robust alternative has not been proposed so far.⁴ In this sense, an antigenic fraction composed of tegumental antigens from *E. granulosus* PSC (termed PSEx) has been previously shown to induce specific antibodies in dogs⁴³

as well as in sheep⁴⁴; and to induce protection against secondary infection in CD1 immunized mice.¹⁸ Although PSEx is assumed to contain only tegumental antigens, no extensive characterization has been performed so far. Additionally, since tegumental components would be the first ones to interact with the host immune system, identifying putative protective antigens within this fraction is a logical approach for designing new vaccines against *E. granulosus*. In the present study, we aimed at identifying—and designing—novel vaccine candidates against *E. granulosus* based on PSEx characterization through a multidisciplinary approach that combined TEM, proteomics, and bio-immuno-informatics.

First, we characterized some biochemical properties of PSEx. In this sense, we initially measured its protein, carbohydrate and lipid content, showing that PSEx is an antigenic fraction mainly enriched in proteins and lipids. Then, to test whether PSEx is primarily composed of tegumental antigens, we performed TEM studies on pre-PSEx-extracted and post-PSEx-extracted PSC. In this

TABLE 3 Epitope-based vaccine candidates

Peptide candidate	Sequence	Length (aa)	Theoretical mass (Da)	Theoretical IP	Parental protein	Parental position	SSS (%)
#1	TEGNIDKWQKSKYDESKGLGC	21	2442.73	5.88	A5YTY7	27-47	90
#2	HAKPKGDFDDRED	13	1529.59	4.75	A5YTY7	49-61	92
#3	ADFNIRNWR	9	1191.31	9.64	A0A068WVYL6	190-198	100
#4	GELVNKNEKTSYPTRTSD	18	2039.19	6.18	A0A068WMU2	32-49	88
#5	EADMYKLEDAQLQDLGR	17	1995.19	4.11	A0A068WVG0	324-340	82
#6	CYVLDFEQEMATAASSSLEKSYELPDG	28	3071.33	3.77	W6UMH9	218-245	92
#7	RHEPTEQQLKDIA NRGIDSIIDKKAAEADAQK	32	3589.97	5.61	A0A068WUJR8	175-207	90
#8	LPIERP KSLSTDGLRT	16	1783.06	8.75	U6J0K5	445-461	82
#9	PGTMDSVRAGPFGQIFRPDNFVFGQSGAGNNWAKGHYTEGAELVDS	46	4886.31	4.83	U6JIN6	70-115	80
#10	LWNSDAGDTDAEP	14	1461.46	3.37	Q56J98	230-243	85
#11	DLGVVPLGGTADECTTQGLDNL	22	2188.39	3.37	U6IXV3	110-131	81
#12	APNYKAIKLTPEKGLAIQPTAEQERQLAVTNKEGLEIVTADDGS	44	4709.29	4.77	U6JBW8	189-232	97
#13	MDLSGARELEALPDEKP	17	1871.09	4.18	A0A068X307	213-229	94
#14	RSPPATEILDESRYKELF	19	2208.46	4.87	A0A068WIL3	358-376	94

Abbreviations: aa, amino acid; IP, isoelectric point; SSS, secondary structure similarity.

The table shows each identified peptide sequence, as well as their length (in amino acids), theoretical mass, and theoretical IP. "Parental protein" and "parental position" columns show protein accession number from which the peptide derives, as well as the position within that protein (from N-terminal amino acid), respectively. Finally, SSS score for each peptide is shown in the last column.

regard, we observed that PSEx extraction completely disrupted the external tegumental layer of PSC, without any remarkable morphological changes in inner parasite structures. In fact, while distal cytoplasm and tegumental spikes disappeared almost completely after PSEx extraction, the fibrous zone, muscle cells, and inner PSC structures showed little, if any, modifications (Figure 1A and 1B). Therefore, we concluded that PSEx mostly contains antigens derived from PSC distal cytoplasm and tegumental spikes, which might include cytoplasmic, vesicular, and/or mitochondrial proteins.

Second, 2D electrophoresis was applied to PSEx to the extent of performing proteomic analyses by mass spectrometry. A complete list of the 58 identified proteins in PSEx is shown in Table 1. It is worth mentioning that some proteins were found in multiple spots—either in both gels or multiple times within the same gel—and, therefore, their MASCOT scores correspond to the lowest one and the sequence coverage correspond to a reconstruction made with the peptides obtained from every matching spot (maximum combined sequence coverage).

Several proteomic analyses of different *E. granulosus* stages or antigenic fractions have been reported up to date, including PSC excretion/secretion products,³³ somatic antigens of PSC^{20,34,37,40} somatic antigens of adult worms,³⁴ hydatid fluid from fertile, and infertile cysts,^{35,38,39} as well as extracellular vesicles obtained from fertile hydatid fluid.³⁶ Among the 58 proteins we reported for PSEx, 36 of them were previously identified in at least one of the abovementioned studies, while 22 proteins were described for the first time in the present work (Table 1). Furthermore, several of the here-identified proteins have been proposed as vaccine candidates against *E. granulosus*.^{4,45,46} Among them, Eg.P29—also known as endophilin B1 (#17 in our set)—revealed significant protective activity against secondary infection in mice,⁴⁷ as well as against primary infection in sheep,⁴⁸ inducing in both cases roughly 95% protection levels. According to our classification, this protein is “unique” (US: 0.24) and highly antigenic (ANTIGENpro score: 0.94), showing an AAR value (42.2) similar to the one obtained for the whole PSEx proteome.

Third, we performed an extensive immunobioinformatic characterization of PSEx proteome, including global as well as a protein-by-protein analyses. Regarding PSEx global analysis, we performed a GO terms study. It is worth mentioning that since PSEx proteome was identified via 2DE followed by MALDI-TOF/TOF, it is impossible to know which protein(s) is more abundant. Thus, unlike GO enrichment analysis—where the frequency a term is found is taken into consideration—our GO analysis only included unique terms describing the whole fraction characteristics. Therefore, our results

show that proteins within PSEx are mostly involved in microtubules-associated processes related to vesicles structures, and exhibit diverse metabolic functions (Figure 2). These results are in accordance with the already known functions of PSC tegument.⁴⁹ As seen in Figure 1A, PSC tegument (particularly its distal cytoplasm) is heavily composed of vesicular structures, which can be arranged in tegumental spikes (probably through microtubule rearrangement). Additionally, PSC tegument is extremely important in several aspects of parasite metabolism, like nutrient absorption and waste disposal.⁴⁹ Our GO analysis also suggested that PSC tegument might act as a defensive barrier against the host attack, as it is involved in stress response processes (eg, removal of superoxide radicals). Also, our results showed a predominance of cytoplasmic and, to a lesser extent, mitochondrial proteins within PSEx (Table 1), and although such results might seem contradictory with GO results, it is important to bear in mind that PSC tegument is a cellular syncytium whose outer layer is the distal cytoplasm, where mitochondria are usually found.⁴⁹

Although the information provided by our GO and subcellular analyses is valuable, it does not allow to identify novel putative protective antigens. Therefore, we further studied other characteristics of PSEx proteome, such as protein-by-protein intrinsic antigenicity and conservancy values. Regarding antigenicity, most proteins within PSEx (74%) were predicted to be highly antigenic and displaying relatively low AAR values (mean of PSEx proteome = 41.0). AAR value is an inverse measure of antigenicity, mostly used to measure the antigenicity of a set of proteins. Interestingly, the mean AAR value of PSEx proteome is slightly higher than the value predicted for a set of secreted proteins from parasitic cestodes (mean AAR = 24.5) proposed as highly antigenic candidates useful for serodiagnosis.²⁶ Then, when analyzing proteins’ “uniqueness score”—a measure of protein divergence obtained by BLASTp against nonhelminth organisms—we observed that over 40% of PSEx proteins can be assessed as “highly divergent” (US values in the 0.5–0.8 range) and 34% as “unique” proteins (US < 0.5). On the basis of these results, and bearing in mind that PSEx is able to induce protection after immunization,¹⁸ we propose nine new proteins to be further tested as vaccine candidates against *E. granulosus* infection. Those proteins were selected from the set of 22 proteins identified for the first time in the present study since they are: (i) either “highly divergent” or “unique”, and (ii) highly antigenic. According to our protein set, candidate proteins are: (#7) aldo keto reductase family 1 member B4, (#8) ankyrin, (#21) fatty acid binding protein FABP2, (#27) heat shock 70 kDa protein,

(#37) mechanosensory protein 2, (#41) peptidylprolyl isomerase, (#42) phosphatase 2a inhibitor i2pp2a, (#45) SPRY domain containing protein, and (#52) tubulin β chain. Unfortunately, since no additional information is available on possible polymorphisms in such proteins, their potential as universal vaccine candidates needs to be further studied.

Additionally, we also predicted and selected PSEx-derived peptides to be further tested as epitope-based vaccines. To that end, we designed and applied a novel working pipeline following a step-by-step reduction to select linear B cell epitopes free of PTMs on the native protein, which are not present in any other non-helminth organism, exhibiting similar secondary structure to the native peptide within the parental protein and displaying neither toxicity nor allergenicity potential (Figure 3). Additionally, surface antibody accessibility to the peptides within the native protein was also assessed (Figure 4). Here, it is worth mentioning that when deciding to use peptides as vaccine antigens in a carrier-conjugated format, some considerations must be taken into account. In this sense, although peptide synthesis can be performed in a nonbiological way, greatly reducing manufacturing costs, only a half of them, when conjugated to a carrier protein, have shown the ability to induce cross-reactive antibodies against the native protein.⁴¹ In this sense, our workflow aimed at maximizing this cross-reactivity potential by performing SSS analyses in early developmental stages. In addition, predicting their surface accessibility to antibodies would probably increase their success.²⁶ Another key-step in our pipeline is selecting peptides only found in helminth species. This step, in combination with the “uniqueness” analysis would be useful for two main reasons. On the one hand, if peptide sequences are present in host proteins, no antibody induction would probably occur because of tolerance processes or, if produced, self-reacting antibodies could display autoimmunity-inducing potential. On the other hand, if peptide sequences are present in other pathogens, then induced antibodies could interfere in diagnostic tests deriving in false-positive results in clinical settings. An additional consideration that must be taken into account when designing peptide-based vaccines is the need for protein carriers, which possess T cell epitopes useful for T-dependent antibody production. Although this seems to be a drawback for peptide-based vaccines, it may actually be an opportunity for success since “unique” epitopes could be used even when the whole protein shows low US. For example, by attaching unique epitopes like those described for actin (#3, US: 0.93) or tubulin β chain (#50, US: 0.97) to a carrier protein, specific antibodies against the peptide could be

obtained, with cross-reactive binding to the native protein, bypassing possible host tolerance mechanisms. However, this hypothesis has still to be experimentally confirmed.

Finally, it is important to highlight that as far as our knowledge extends, the present study is the first one to perform linear B cell epitope predictions within mass spectrometry identified protein fragments to identify clean linear B cell epitopes, that is, free of PTMs. Such a strategy, in combination with additional bioinformatic tools, constitutes a novel rational approach for identification of potential peptide-based vaccine candidates. In this regard, we finally suggested 14 novel peptides derived from 13 different proteins—ranging from 9 to 46 amino acids in length—to be further tested as vaccine candidates against *E. granulosus* infection (Table 3). Interestingly, one of those peptides (#3 in Table 3) belonged to one of the proteins also proposed as new vaccine candidates (#8 in Table 1). We are currently designing the effectiveness testing of the suggested vaccine candidates. To that end, in vivo immunogenicity will be firstly assessed in mice, and those effectively immunogenic will be further studied for their ability to induce antibodies with in vitro parasite-killing activity. Finally, promising candidates will be tested in the murine model of secondary CE,⁵⁰ and protective antigens might be later assayed in experimental infections using relevant live-stock species.

ACKNOWLEDGEMENTS

Authors are grateful to the Electron Microscopy Center Facility of Heidelberg University for the introduction and use of their microscopes. MSc Paula Arbildi is acknowledged for valuable suggestions on figures' esthetics. MEA was recipient of a short-term grant funded by United Nations University-Biotechnology for Latin America and the Caribbean (UNU-BIOLAC). CSIC-Universidad de la República (Uruguay), PEDECIBA-Química (Uruguay) and ANII (Uruguay) are acknowledged for their general financial support.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Conceptualization: SM and GM-E. Funding acquisition: SM, SD, RD, FF, and GM-E. Investigation: SM, MP, MC, MEA, and GM-E. Writing manuscript: SM and GM-E.

ORCID

Gustavo Mourglia-Ettlin  <http://orcid.org/0000-0002-2418-2908>

REFERENCES

- World Health Organization, 2018. Neglected Tropical Diseases. Echinococcosis. www.who.int/en/news-room/fact-sheets/detail/echinococcosis
- Thompson RCA, Lymbery AJ. Echinococcus and hydatid disease. Wallingford: CAB International; 1995.
- Heath DD. The development of *Echinococcus granulosus* larvae in laboratory animals. *Parasitology*. 1970;60(3):449-456.
- Pourseif MM, Moghaddam G, Saeedi N, Barzegari A, Dehghani J, Omid Y. Current status and future prospective of vaccine development against *Echinococcus granulosus*. *Biologicals*. 2018; 51:1-11.
- Heath DD, Jensen O, Lightowlers MW. Progress in control of hydatidosis using vaccination: a review of formulation and delivery of the vaccine and recommendations for practical use in control programs. *Acta Trop*. 2003;85(2):133-143.
- Lightowlers MW, Jensen O, Fernandez E, et al. Vaccination trials in Australia and Argentina confirm the effectiveness of the EG95 hydatid vaccine in sheep. *Int J Parasitol*. 1999;29(4): 531-534.
- Pan W, Chen DS, Lu YJ, et al. Genetic diversity and phylogenetic analysis of EG95 sequences of *Echinococcus granulosus*: Implications for EG95 vaccine application. *Asian Pac J Trop Med*. 2017;10(5):524-527.
- Álvarez-Rojas CA, Romig T, Lightowlers MW. *Echinococcus granulosus sensu lato* genotypes infecting humans—review of current knowledge. *Int J Parasitol*. 2014;44(1):9-18.
- Mahmoud A. New vaccines: challenges of discovery. *Microb Biotechnol*. 2016;9(5):549-552.
- Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res*. 2006;2:2.
- de Aluja AS, Villalobos NM, Nava G, et al. Therapeutic capacity of the synthetic peptide-based vaccine against *Taenia solium* cysticercosis in pigs. *Vaccine*. 2005;23(31): 4062-4069.
- Jezek J, El Ridi R, Salah M, et al. *Fasciola gigantica* cathepsin L proteinase-based synthetic peptide for immunodiagnosis and prevention of sheep fasciolosis. *Biopolymers*. 2008;90(3):349-357.
- Robinson K, Bellaby T, Chan WC, Wakelin D. High levels of protection induced by a 40-mer synthetic peptide vaccine against the intestinal nematode parasite *Trichinella spiralis*. *Immunology*. 1995;86(4):495-498.
- Rojas-Caraballo J, López-Abán J, Pérez del Villar L, et al. In vitro and in vivo studies for assessing the immune response and protection-inducing ability conferred by *Fasciola hepatica*-derived synthetic peptides containing B- and T-cell epitopes. *PLoS One*. 2014;9(8):e105323.
- Arnon R, Tarrab-Hazdai R, Steward M. A mimotope peptide-based vaccine against *Schistosoma mansoni*: synthesis and characterization. *Immunology*. 2000;101(4):555-62..
- Woollard DJ, Gauci CG, Lightowlers MW. Synthetic peptides induce antibody against a host-protective antigen of *Echinococcus granulosus*. *Vaccine*. 1999;18(9-10):785-794.
- Read AJ, Casey JL, Coley AM, et al. Isolation of antibodies specific to a single conformation-dependant antigenic determinant on the EG95 hydatid vaccine. *Vaccine*. 2009;27(7):1024-1031.
- Hernández A, Nieto A. Induction of protective immunity against murine secondary hydatidosis. *Parasite Immunol*. 1994;16(10):537-544.
- Míguez M, Baz A, Nieto A. Carbohydrates on the surface of *Echinococcus granulosus* protoscoleces are immunodominant in mice. *Parasite Immunol*. 1996;18(11):559-569.
- Mourglia-Ettlin G, Miles S, Velasco-De-Andrés M, et al. The ectodomains of the lymphocyte scavenger receptors CD5 and CD6 interact with tegumental antigens from *Echinococcus granulosus sensu lato* and protect mice against secondary cystic echinococcosis. *PLoS Negl Trop Dis*. 2018;12(11):e0006891.
- Cozzetto D, Minneci F, Currant H, Jones DT. FFPred 3: feature-based function prediction for all Gene Ontology domains. *Sci Rep*. 2016;6:31865.
- Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011;6(7):e21800.
- Almagro Armentero JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics*. 2017;33(21): 3387-3395.
- Magnan CN, Zeller M, Kayala MA, et al. High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics*. 2010;26(23):2936-2943.
- Cornejo-Granados F, Hurtado-Ramírez JM, Hernández-Pando R, Ochoa-Leyva A. Secret-AAR: a web server to assess the antigenic density of proteins and homology search against bacterial and parasite secretome proteins. *Genomics*. 2018.
- Miles S, Navatta M, Dematteis S, Mourglia-Ettlin G. Identification of universal diagnostic peptide candidates for neglected tropical diseases caused by cestodes through the integration of multi-genome-wide analyses and immunoinformatic predictions. *Infect Genet Evol*. 2017;54:338-346.
- Jespersen MC, Peters B, Nielsen M, Marcanti P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*. 2017;45:W24-W29. <https://doi.org/10.1093/nar/gkx352>
- Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*. 1995;11(6): 681-684.
- Gupta S, Kapoor P, Chaudhary K, et al. In silico approach for predicting toxicity of peptides and proteins. *PLoS One*. 2013;8(9):e73957.
- Saha S, Raghava GP. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res*. 2006;34(web server issue):202-209.
- Källberg M, Wang H, Wang S, et al. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*. 2012;7(8):1511-1522.
- Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997;18(15):2714-23..
- Virginio VG, Monteiro KM, Drumond F, et al. Excretory/secretory products from in vitro-cultured *Echinococcus granulosus* protoscoleces. *Mol Biochem Parasitol*. 2012;183(1):15-22.

34. Cui SJ, Xu LL, Zhang T, et al. Proteomic characterization of larval and adult developmental stages in *Echinococcus granulosus* reveals novel insight into host-parasite interactions. *J Proteomics*. 2013;84:158-175.
35. Ahn CS, Kim JG, Han X, Kang I, Kong Y. Comparison of *Echinococcus multilocularis* and *Echinococcus granulosus* hydatid fluid proteome provides molecular strategies for specialized host-parasite interactions. *Oncotarget*. 2017;8(57):97009-97024.
36. Siles-Lucas M, Sánchez-Ovejero C, González-Sánchez M, et al. Isolation and characterization of exosomes derived from fertile sheep hydatid cysts. *Vet Parasitol*. 2017;236:22-33.
37. Lorenzatto KR, Kim K, Ntai I, et al. Top Down proteomics reveals mature proteoforms expressed in subcellular fractions of the *Echinococcus granulosus* preadult stage. *J Proteome Res*. 2015;14(11):4805-4814.
38. Zeghir-Bouteldja R, Polomé A, Bousbata S, Touil-Boukoffa C. Comparative proteome profiling of hydatid fluid from Algerian patients reveals cyst location-related variation in *Echinococcus granulosus*. *Acta Trop*. 2017;171:199-206.
39. dos Santos GB, Monteiro KM, da Silva ED, Battistella ME, Ferreira HB, Zaha A. Excretory/secretory products in the *Echinococcus granulosus* metacestode: is the intermediate host complacent with infection caused by the larval form of the parasite? *Int J Parasitol*. 2016;46(13-14):843-856.
40. Hidalgo C, García MP, Stoore C, et al. Proteomics analysis of *Echinococcus granulosus* protoscoleces stage. *Vet Parasitol*. 2016;218:43-45.
41. Hancock DC, O'Reilly NJ. Synthetic peptides as antigens for antibody production. *Methods Mol Biol*. 2005;295:13-26.
42. Álvarez-Rojas CA, Gauci CG, Lightowlers MW. Antigenic differences between the EG95-related proteins from *Echinococcus granulosus* G1 and G6 genotypes: implications for vaccination. *Parasite Immunol*. 2013;35(2):99-102.
43. Carol H, Nieto A. A mucosal IgA response, but no systemic antibody response, is evoked by intranasal immunisation of dogs with *Echinococcus granulosus* surface antigens iscoms. *Vet Immunol Immunopathol*. 1998;65(1):29-41.
44. Valizadeh M, Haghpanah B, Badirzadeh A, Roozintan E, Fallahi S, Raeghi S. Immunization of sheep against *Echinococcus granulosus* with protoscoleces tegumental surface antigens. *Vet World*. 2017;10(8):854-858.
45. Li ZJ, Zhao W. Analysis of protoscoleces-specific antigens from *Echinococcus granulosus* with proteomics combined with Western blot. *Biomed Environ Sci*. 2012;25(6):718-723.
46. Wang F, Ye B. In silico cloning and B/T cell epitope prediction of triosephosphate isomerase from *Echinococcus granulosus*. *Parasitol Res*. 2016;115(10):3991-3998.
47. Shi Z, Wang Y, Li Z, et al. Cloning, expression, and protective immunity in mice of a gene encoding the diagnostic antigen P-29 of *Echinococcus granulosus*. *Acta Biochim Biophys Sin*. 2009;41(1):79-85.
48. Wang H, Li Z, Gao F, et al. Immunoprotection of recombinant Eg.P29 against *Echinococcus granulosus* in sheep. *Vet Res Commun*. 2016;40:73-79.
49. Morseth DJ. Fine structure of the hydatid cyst and protoscoleces of *Echinococcus granulosus*. *J Parasitol*. 1967;53(2):312-325.
50. Mourglia-Ettlin G, Cucher M, Arbildi P, Rosenzvit M, Dematteis S. Natural and induced antibodies contribute to differential susceptibility to secondary cystic echinococcosis of Balb/c and C57Bl/6 mice. *Immunobiology*. 2016;221(1):103-115.

How to cite this article: Miles S, Portela M, Cyrklaff M, et al. Combining proteomics and bioinformatics to explore novel tegumental antigens as vaccine candidates against *Echinococcus granulosus* infection. *J Cell Biochem*. 2019;1-17. <https://doi.org/10.1002/jcb.28799>