Journal of Computer-Aided Molecular Design
https://doi.org/10.1007/s10822-020-00352-8

Application of target repositioning and in silico screening to exploit fatty acid binding proteins (FABPs) from *Echinococcus multilocularis* as possible drug targets

Julián A. Bélgamo[1,3] · Lucas N. Alberca[2,3] · Jorge L. Pórfido[1,4] · Franco N. Caram Romero[2] · Santiago Rodriguez[1,3] · Alan Talevi[2,3] · Betina Córsico[1,3] · Gisela R. Franchini[1,3]

Received: 20 May 2020 / Accepted: 9 October 2020

Abstract

Fatty acid binding proteins (FABPs) are small intracellular proteins that reversibly bind fatty acids and other hydrophobic ligands. In cestodes, due to their inability to synthesise fatty acids and cholesterol de novo, FABPs, together with other lipid binding proteins, have been proposed as essential, involved in the trafficking and delivery of such lipophilic metabolites. Pharmacological agents that modify specific parasite FABP function may provide control of lipid signalling pathways, inflammatory responses and metabolic regulation that could be of crucial importance for the parasite development and survival. *Echinococcus multilocularis* and *Echinococcus granulosus* are, respectively, the causative agents of alveolar and cystic echinococcosis (or hydatidosis). These diseases are included in the World Health Organization's list of priority neglected tropical diseases. Here, we explore the potential of FABPs from cestodes as drug targets. To this end, we have applied a target repurposing approach to identify novel inhibitors of *Echinococcus* spp. FABPs. An ensemble of computational models was developed and applied in a virtual screening campaign of DrugBank library. 21 hits belonging to the applicability domain of the ensemble models were identified, and 3 of the hits were assayed against purified *E. multilocularis* FABP, experimentally confirming the model's predictions. Noteworthy, this is to our best knowledge the first report on isolation and purification of such four FABP, for which initial structural and functional characterization is reported here.

Keywords Drug repurposing · Target repurposing · FABP · Virtual screening · *Echinococcus* spp. · Neglected tropical diseases

Introduction

Echinococcus multilocularis and *Echinococcus granulosus* are, respectively, the causative agents of alveolar and cystic echinococcosis (or hydatidosis). These diseases are included in the list of twenty neglected tropical diseases (NTDs) considered as priority by World Health Organization [1]. NTDs affect more than a billion people every year, mainly in poor regions with limited access to adequate sanitary conditions and living in close association with infectious vectors, domestic animals and livestock.

Echinococcosis is caused by the larval stage of *Echinococcus* spp., which generates cyst-like structures in the intermediate host's inner organs. Currently, the drug treatment for echinococcosis relies on benzimidazoles, mainly albendazole, which was reported to be ineffective in 40% of cystic echinococcosis cases [2]; furthermore, it is parasitostatic only and not tolerated by many patients [3]. Arising

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10822-020-00352-8) contains supplementary material, which is available to authorized users.

✉ Gisela R. Franchini
gfranchini@conicet.gov.ar

[1] Instituto de Investigaciones Bioquímicas de La Plata (INIBIOLP), Facultad de Ciencias Médicas, Universidad Nacional de La Plata (UNLP)-Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), La Plata, Argentina

[2] Laboratory of Bioactive Research and Development (LIDeB), Faculty of Exact Sciences, Universidad Nacional de La Plata (UNLP), Buenos Aires, Argentina

[3] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

[4] Present Address: Institut Pasteur Montevideo, Montevideo, Uruguay

Published online: 17 October 2020

of resistance to albendazole has been reported for many helminths [4]. The scarcity of available anthelmintic drugs and the emergence of resistant parasites make the discovery of new anthelmintic drugs mandatory.

The sequencing of tapeworm genomes [5], showed the absence of genes coding for de novo synthesis of fatty acids and cholesterol, confirming previous biochemical observations [6]. Additionally, an amplification and high expression of genes coding for lipid binding proteins (fatty acid binding proteins (FABPs) and Antigen B) was also observed. The encoded proteins could be involved in lipid acquisition from host tissues.

FABPs are small intracellular lipid binding proteins (around 15 kDa) that reversibly bind fatty acids and other hydrophobic ligands. They are usually highly expressed in cells with very active lipid metabolism [7, 8]. Members of this family have been found throughout the animal kingdom, but no counterparts have been found in plants or fungi [9]. Our group has functionally characterized a member of this family of proteins from *E. granulosus* (EgFABP1), [10]. Remarkably, EgFABP1, confirms the presence of all the usual FABPs structural characteristics and facilitates the search for specific compounds able to bind and interfere with their function.

These lipid-binding proteins have been suggested as possible drug targets and/or drug carriers against these parasites [11, 12]. By exploring the genomes of *E. multilocularis* and *E. granulosus*, six genes coding for FABPs were found in each organism. In the case of *E. granulosus*, all of them have different coding sequences, whereas in *E. multilocularis*, two of the genes code for the same protein [13].

Target repurposing is a relatively novel strategy to discover new hits against infectious agents, including NTDs [14]. Target repurposing begins with a defined parasitic target with an established homolog in another species (human or other). Drugs that target this second species (either approved drugs or clinical candidates), are then explored as a starting point to develop compounds that inhibit the pursued infectious agent. Hits identified using such strategy are later subjected to molecular optimization to gain selectivity. For instance, Nallan et al. tested the ability of a panel of inhibitors for mammalian protein farnesyl transferase in preclinical and clinical development to inhibit in vitro growth of *Plasmodium falciparum* parasites [15]. Similarly, other authors have repurposed human phosphodiesterase inhibitors against *Trypanosoma brucei* phosphodiesterases, as potential starting points for the development of new treatments against African trypanosomiasis [16, 17].

In this work, we explored the potential of cestode FABPs as drug targets. To this end, we have used a target repurposing approach to identify, through a virtual screening campaign based on an ensemble learning approximation, novel inhibitors of *Echinococcus* FABPs. Additionally, four *E. multilocularis* FABPs were purified for the first time and used in the experimental validation of the models' predictions.

Materials and methods

Materials

Oleic acid, sulindac, clotrimazole, cetylpyridinium chloride, urea, and buffers were acquired from Sigma-Aldrich (St. Louis, MO). 1-anilino naphthalene-8-sulfonic acid (ANS) was purchased from Molecular Probes (Eugene, Oregon, USA).

Dataset collection

Human FABP2 inhibitors were compiled from literature [18–31]. 108 compounds assayed through a fluorescence polarization assay and 159 compounds identified through 8-anilino-1-naphthalenesulfonic acid displacement assay were found, adding a total of 267 compounds. After removing duplicated compounds, we were left with a 258-compound dataset. Such compounds were labelled as either ACTIVE or INACTIVE according to their reported inhibitory data. Compounds with $IC_{50} < 20$ μM were allocated to the ACTIVE class; otherwise, they were allocated to the INACTIVE class. Considering such criteria, the dataset included 125 active compounds and 133 inactive compounds. The molecular representations of the dataset compounds were standardized using the standardization tool Standardizer v. 17.2.6.0 available in JCHEM Suite (Chemaxon). The molecular diversity of the whole dataset and within each category can be appreciated in Figure 1 of the Supplementary Material. The heatmap shows, for every compound pair, the Tanimoto distances computed using ECFP_4 molecular fingerprints. The heatmap was built using the Python library Plotly (https://plotly.com/python/) and Tanimoto distances were calculated using ScreenMD—Molecular Descriptor Screening v. 17.3.27.0 (ChemAxon). The dataset is included as Supplementary Data 1.

Dataset partition

The dataset was split into a representative training set (which was used to calibrate the models) and a test set (which was used to assess the performance of the models in an independent manner). It has been reported that representative splitting tends to produce models with better predictivity [32, 33]. Accordingly, we resorted to a representative sampling procedure to divide the dataset; such representative partition resulted from a sequential clustering procedure. In the first place, the hierarchical clustering method included

in LibraryMCS software (version 17.2.13.0–ChemAxon) was applied, which relies on the Maximum Common Substructure. A compound from each of the resulting clusters was then randomly chosen and used as a seed to perform a non-hierarchical clustering using the k-means algorithm, as implemented in Statistica 10 Cluster Analysis module (Statsoft). Hierarchical clustering allowed deciding on an initial partition of n molecules into k groups, and this preliminary clustering was then optimized through the non-hierarchical procedure, as suggested by Everitt et al. [34]. This combined approach has earlier been applied for representative dataset partitioning, with good results [35–37].

The previous clustering procedure was performed separately for the ACTIVE and INACTIVE classes. 75% of the elements in each cluster of the ACTIVE class were held for the training set (making a total of 93 training examples); an equal number of compounds were taken from the INACTIVE category clusters (74% of each INACTIVE cluster). In this way, a balanced training sample (comprising an identical proportion of active and inactive examples) was obtained and model bias toward predicting an overrepresented class was avoided. The remaining 32 active and 40 inactive compounds were allocated to the test set (72 compounds in total).

Molecular descriptor calculation and modelling procedure

3668 conformation-independent descriptors were computed with Dragon 6.0 software. A random subspace approximation was applied to sample 1000 descriptor subsets of 200 potential independent features each. In the random subspace approach, the molecular descriptors are randomly sampled, and each model is trained on one subset of the feature space [38, 39], resulting in models that do not over-focus on features that display high explanatory power for the calibration sample.

A dummy variable (class label) was used as a dependent variable. It was assigned observed values of 1 for compounds within the ACTIVE category and observed values of 0 for compounds in the INACTIVE one. Using a Forward Stepwise procedure and a semi-correlation approach [40], 1000 linear classifiers were so obtained, one from each random subset of features. To reduce the probability of overfitting, only one molecular descriptor every twelve training examples was allowed into each model. A maximal Variance Inflation Factor of 2 was tolerated to exclude highly correlated descriptor pairs from the models. R language and environment was used for all data analysis. The R package data table (https://cran.r-project.org/package=data.table) was employed to handle datasets.

The robustness and predictive ability of the models were firstly assessed through Y-randomization and Leave-Group-Out cross-validation tests. In the case of randomization, the class label was randomized across the compounds in the training set. The training set with the randomized dependent variable was then used to train new models, from the descriptor selection step. Such procedure was repeated 100 times for each descriptor subset, and the 95% confidence interval was built around the mean accuracy of the randomized models, under the expectancy that the randomized models will perform poorly compared to the real ones. Regarding the Leave-Group-Out cross-validation, random stratified subsets comprising 9 active compounds and 9 inactive compounds were removed from the training set in each cross-validation round, and the model was regenerated using the remaining compounds as training examples. The resulting model was used to predict the class label for the removed compounds. The procedure was repeated 100 times for each mode, with each of the training set compounds removed at least once. The results were informed as the average accuracy across the folds, and this was compared to the accuracy of the model for the original training set and also, as advised by Gramatica [41], to the No-Model error rate or risk (NOMER%), i.e., the error provided in absence of model.

The predictivity of each individual model was assessed through external validation, using the 72-compound test set that was already described in *Dataset Collection* section. A diversity of statistical parameters commonly used to assess the performance of classificatory models were estimated [41, 42] for both the training and test sets: sensitivity (*Se*, i.e., true positive rate), specificity (*Sp*, i.e., true negative rate), accuracy (*Acc*, i.e., overall percentage of good classifications), positive and negative predictivity and the F-measure.

Retrospective screening campaigns

For further validation of our models, a retrospective virtual screening experiment was implemented, seeding the 32 known active compounds from the test set among 1568 decoys obtained from the Directory of Useful Decoys Enhanced (DUD-E; [43]), a widely used benchmarking tool which allows obtaining putative inactive compounds.

Using rigorous simulated ranking experiments, Truchon and Bayly [44] have previously demonstrated that the Area Under the Receiver Operating Characteristic curve (AUC ROC) is dependent on the ratio of active compounds/inactive compounds, and the standard deviation of the metric converges to a constant value when small yield of actives (Ya) are observed in the screened library (Ya below 0.05 provides robust results). Furthermore, a reasonably small Ya ensures that the saturation effect is constant or absent. A high number of decoys (around 1000 or higher) and a small Ya contribute to a controlled statistical behaviour.

Ensemble learning

Classifier ensembles are known to enhance generalization and accuracy in comparison to individual (single model) classifiers [45, 46].

The best individual models were selected and combined using the area under the ROC curve metric (AUC ROC) in the retrospective screen as criterion of performance. To choose the ideal number of models to be included in the ensemble, systematic combinations of the 2 to 100 best performing individual models were considered (the two best-performing models were combined, then the three best-performing models, and so on up to a total of 100 models included in the ensemble). Four combination strategies were applied to obtain a combined score: MIN operator (which returns the minimum score among the individual scores of the combined models); Average Score; Average Ranking and; Average Voting. Voting was calculated according to the equation previously used by Zhang and Muegge [47]. AUC ROCs were obtained with the pROC package [48]; the Delong method was used to statistically compare the AUC ROCs. BEDROC and the enrichment factor (considering the 1% top-ranked compounds) (EF1%) were also computed. For that purpose, we resorted to the online open source application Rocker [49].

Use of positive predictive value surfaces to choose an adequate score threshold

A practical concern in any virtual screen involves estimating the probability that a predicted hit will confirm its activity when submitted to experimental testing (i.e., the Positive Predictive Value, PPV). A priori estimation of such probability is however not possible due to its dependency on the *Ya* of the screened library, which is not known beforehand:

$$PPV = \frac{Se \times Ya}{Se \times Ya + (1 - Sp) \times (1 - Ya)} \quad (1)$$

where *Se* represents the sensitivity associated to a given score threshold value and *Sp* represents the specificity. Equation (1) was applied to build PPV surfaces. In order to choose an optimal threshold value to select predicted hits in prospective virtual screening experiments, 3D plots displaying the interplay between PPV, the *Se/Sp* ratio and *Ya* were built for each individual model and for each model ensemble, as previously described [35]. Using the library built for the retrospective screening experiment, *Se* and *Sp* were computed in all the range of possible score threshold values. Since controlled statistical behaviour has been observed for libraries of 1000 compounds or more and *Ya* below 0.05, we can sensibly assume that the ROC curve and its derived metrics will be similar when applying the models to screen other chemical libraries with similar characteristics (size > 1000 compounds, $Ya < 0.05$). Having in mind that in real virtual screening applications *Ya* is ignored a priori but invariably low, this parameter was varied between 0.001 and 0.010. The R package plotly was used to obtain all the PPV graphs. Visual analysis of the resulting PPV surfaces allowed us to select a score threshold value associated to the desired PPV range.

Prospective virtual screening

Based on the visual inspection of the resulting PPV graphs, we have realized a prospective virtual screen, applying a 5-model ensemble (see description later, under "Results") using the MIN operator to combine individual classifiers. Based on PPV surface analysis, we chose a score threshold that provides a PPV ≥ 25% at $Ya = 0.01$.

We have used the 5-model ensemble to screen DrugBank 5.0.8 database, a public database containing extensive information about the US Food and Drug Administration (FDA) approved, experimental, illicit and investigational drugs [50], as well as approved and investigational drugs from other regulatory agencies, such as Health Canada or the European Medicines Agency. The molecular representations of the database compounds were standardized using Standardizer 16.10.10.0 (ChemAxon). The following actions were applied to obtain homogeneous representations of the molecular structure for the subsequent virtual screen: (1) strip salts; (2) remove solvents; (3) clear stereo; (4) remove absolute stereo; (5) aromatize; (6) neutralize; (7) add explicit hydrogens; and (8) clean 2D. Duplicated structures were removed using Instant JCHEM v. 17.2.6.0. Belonging to the applicability domain was checked using the leverage approach. Three hits were selected for experimental evaluation, using the following criteria: (a) no previous reports of activity on *Echinoccocus* spp.; (b) availability through local suppliers; (c) cost; (d) belonging to the applicability domain of the five models included in the best ensemble and; (e) molecular diversity: after clustering the in silico hits using Chemaxon's LibraryMCS (minimal maximum common substructure containing 12 atoms) the assayed candidates should not fall within the same cluster. Using these criteria, three hits, the anti-inflammatory sulindac, the antiseptic cetylpyridinium and the antifungal agent clotrimazole, were selected for experimental validation of the predictions.

Recombinant FABP expression and purification

FABP 1 to 4 from *E. multilocularis* were overexpressed in *Escherichia coli*, harboring the pET28a(+)-EmFABP1 to 4. All proteins were purified from soluble extracts except for EmFABP2 that was recovered from inclusion bodies (see

below). Cells were lysed by sonication and centrifuged at 30,000×*g* for 40 min at 4 °C. The His-tagged fusion proteins were purified by nickel affinity and gel filtration chromatography using a Superdex-75 (GE Healthcare). In each case a final delipidation step was performed using Lipidex-1000 (Sigma) hydrophobic chromatography. When necessary the removal of copurifying ligands from the bacterial expression system was achieved by an extra step using reverse-phase high performance liquid chromatography (RP-HPLC) with a C8 stationary phase and water/acetonitrile/trifluoroacetic acid mobile phase. Proteins were refolded in water and then a final buffer change step was performed to 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.8 mM KH_2PO_4 pH 7.4 (PBS). For the case of EmFABP2, inclusion bodies were isolated by centrifugation (at 30,000×*g* for 40 min at 4 °C) and washed with buffer 20 mM Tris–HCl, 500 mM NaCl, 5 mM imidazole and 1 mM DTT, pH 7.4, supplemented with urea up to 8 M concentration. After the sample was dissolved, it was further centrifuged at 30,000×*g* for 20 min at 4 °C and the supernatant applied to a nickel column (GE Healthcare) equilibrated with the same buffer in presence of urea. Refolding was achieved by eluting EmFABP2 in 20 mM Tris–HCl, 500 mM NaCl, 250 mM imidazole and 1 mM DTT, pH 7.4.

Purity was assessed in SDS-PAGE gels. Finally, pooled fractions containing the concentrated proteins were stored at − 80 °C. Protein concentration was assessed by spectrophotometer measurements at 280 nm with the corresponding extinction coefficients: 10,033 M/cm, 11,523 M/cm, 14,105 M/cm and 20,003 M/cm for EmFABP1 to 4, respectively. For the case of EgFABP1 was purified as described previously [10].

Circular dichroism

Spectra were recorded on a Jasco J-810 spectropolarimeter. Data in the far-UV (200–250 nm) region was collected using a 1 mm path cuvette. A scan speed of 20 nm/min with a time constant of 1 s was used. EmFABP 1 to 4 were dissolved in buffer PBS. Each spectrum was measured at least three times, and the data were averaged to reduce noise. Molar ellipticity was calculated as described elsewhere [51], using mean residue weight values of 13,355, 1155, 1153, 1126 for EmFABP1 to 4, respectively.

Equilibrium unfolding studies

Conformational transitions were monitored as a function of denaturant concentration by measuring the change in the intrinsic fluorescence intensity of proteins. Urea stock solutions were prepared on the same day of the experiment. Individual samples ranging in denaturant concentration from 0 to 8 M Urea were obtained by dilution of a fixed volume of protein stock solution in mixtures of buffer PBS and 8 M Urea. Spectra were measured after incubation for at least 1 h to ensure that the equilibrium had been reached. Fluorescence data were corrected for the background signal of buffer and denaturant and expressed in arbitrary units.

ANS binding

In this assay the binding was monitored by changes in the probe emission. The ligand was repeatedly added from a stock solution to each protein dissolved in PBS buffer. Incubation for 2 min at 25 °C ensured complete equilibration. Initially, nonlinear regression fitting to the binding data was achieved to a model considering a single class of binding sites using the software GraphPad Prism.

Competition experiments

Displacement of bound ANS (20 μM) from EmFABPs by oleic acid, a known FABPs ligand [52, 53] or the correspondent in silico hit was measured by the decrease of fluorescence intensity with increasing oleic acid or drug concentration. Protein concentration was 1uM in PBS buffer, pH 7.4. The spectra were recorded after equilibration for 2 min at 25 °C. The apparent dissociation constant (Kd_{app}) was calculated by fitting the following equation to the data:

$$\Delta F = \frac{\Delta F_0}{1 + \frac{[oleic\ acid]}{K_{dapp}}} + \Delta F_{res} \quad (2)$$

where ΔF represents the value of the observed fluorescence intensity subtracted from the contribution of the probe at each concentration of oleic acid assayed; ΔF_0 is the difference in fluorescence intensity measured in the absence or in the presence of an excess amount of oleic acid; and ΔF_{res} is a term that accounts for the remnant fluorescence at high oleic acid concentration. All fluorescence data were normalized by dividing each difference by ΔF_{max} ($= \Delta F_0 + \Delta F_{res}$).

Results

A ligand-based virtual screening approach was used to discover inhibitors of *Echinococcus* spp. FABPs'. 1000 individual linear classifiers were obtained by applying a random subspace approach on a pool of more than 3000 Dragon conformation-independent (0D–2D) molecular descriptors. The individual models were validated both internally and externally.

Results of the internal validation for the five best individual models are shown in Table 1. Regarding the Leave-Group-Out, results for each individual classifier are informed as the mean accuracy across the folds, which is in all cases

Table 1 Summarized results of the internal validation of the best individual classifiers

Model	Training set	Leave group out cross-validation		Fisher´s randomization test	
	Mean accuracy (%)	Mean accuracy (%)	Standard deviation	Confidence interval 95%	
M997	85.48	79.21	8.97	58.08	60.16
M177	82.26	77.56	9.25	57.87	59.71
M1	82.26	78.49	9.12	58.25	59.96
M556	83.87	80.09	8.56	57.52	59.29
M893	82.80	79.39	8.63	57.68	59.61

similar to 80% and close to the correspondent accuracy on the training set, suggesting the models are robust. Since the proportions of the active and inactive compounds in the training set are identical (as are in each of the Leave-Group-Out folds) the correspondent NOMER% (accuracy associated to random classification) is 50%, well below the performance of the models in the cross-validation, as expected.

Regarding the randomization results, Table 1 shows the 95% confidence interval around the mean accuracy of the randomized models. The accuracy of the randomized models is in all cases much below the accuracy of the true (non-randomized) models, and close to the NOMER%, suggesting a low probability of spurious correlations for the true models.

External validation was performed using the 72-compound independent test set. Table 2 summarizes the results. In general, no matter which parameter is taken into consideration, the best individual classifiers show a similar performance in the training and test set.

The best individual model (model 997) included the following features:

$$\begin{aligned}\text{Class} = {} & 0.05428 + 0.26464 * \text{NaasN} - 32.77330 * \text{JGI6} \\ & + 0.07801 * \text{J\_D/Dt} - 0.01735 * \text{ATSC4m} \\ & - 0.44701 * \text{F04[N–N]} + 0.11065 \\ & * \text{Eig15\_AEA(ed)} + 0.20216 * \text{F04[N–S]} \\ & - 0.15403 * \text{B06[N–O]}\end{aligned}$$

with Wilks' Lambda = 0.51, F = 20.970 and $p < 0.0000$. Dragon's nomenclature for the molecular descriptors has been kept in the previous expression, where the molecular descriptors correspond either to topological indices or atom pair counts or presence. NaasN is an atom-type E-state index representing the number of E State of the atom type aaS; JGI6 is the topological charge index of order 6; J_D/Dt is a Balaban-like index derived from the distance/detour matrix; ATSC4m is a Broto-Moreau autocorrelation of lag 4 weighted by atomic mass; F04[N–N] symbolizes the frequency of Nitrogen—Nitrogen atom pairs at a topological distance of 4; Eig15_AEA(ed) denotes the 15th eigenvalue from the augmented edge adjacency matrix weighted by edge degree; F04[N-S] indicates the frequency of Nitrogen–Sulphur atom pairs at a topological distance of 4 (2D Atom Pairs) and B06[N–O] denotes the presence of Nitrogen – Oxygen atoms at topological distance of 6.

The five best individual models and a brief description of the molecular descriptors in them have been listed as Supplementary Data 2.

For a more challenging and realistic simulation, the enrichment behavior of the individual models was studied through a retrospective virtual screen, where a small proportion of active compounds (32) was dispersed among a high number (1568) of presumed inactive compounds (decoys). Table 3 shows the enrichment metrics of the five best performing individual classifiers.

Whereas, the performance of the best individual classifiers was very satisfactory (AUC ROC above 0.94, with

Table 2 Comparison of the performance of the five best individual classifiers on the training and test set

Model	*Se*	*Sp*	Acc	F-measure	Positive predictivity	Negative predictivity
Training set						
M997	0.82	0.89	0.85	0.19	0.88	0.83
M177	0.77	0.87	0.82	0.22	0.86	0.79
M1	0.80	0.85	0.82	0.25	0.84	0.81
M556	0.81	0.87	0.84	0.22	0.86	0.82
M893	0.78	0.87	0.83	0.22	0.86	0.80
Test set						
M997	0.81	0.83	0.82	0.29	0.79	0.85
M177	0.78	0.88	0.83	0.22	0.83	0.83
M1	0.81	0.83	0.82	0.29	0.79	0.85
M556	0.78	0.88	0.83	0.22	0.83	0.83
M893	0.85	0.80	0.83	0.32	0.81	0.84

Table 3 Performance of the five best individual classifiers and the best model ensemble in the retrospective screen

	AUC ROC			BEDROC	EF_1.0
	Training set	Test set	Retrospective screen	Retrospective screen	Retrospective screen
M997	0.90	0.86	0.95	0.67	30
M177	0.89	0.88	0.95	0.76	43
M1	0.89	0.88	0.94	0.76	40
M556	0.91	0.87	0.94	0.67	27
M893	0.90	0.86	0.94	0.64	30
MIN_5 ensemble	–	–	0.99	0.88	50

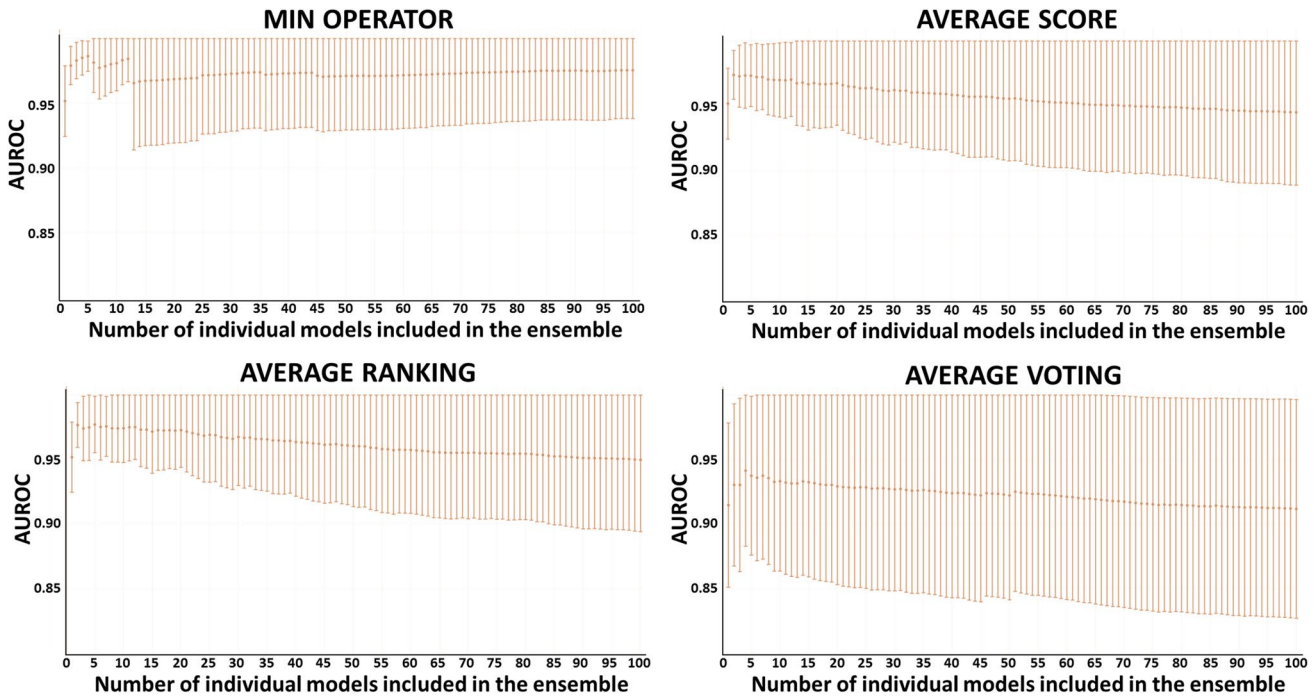

Fig. 1 AUC ROC obtained in the retrospective screen as a function of the number of combined models. **a** Minimum score; **b** average score; **c** average ranking; **d** average voting

1.00 corresponding to a perfect classifier, and enrichment factor up to 43), we explored ensemble learning approaches to obtain meta-classifiers with improved accuracy and a more robust behaviour. Figure 1 shows the AUC ROC values obtained in the retrospective screen when systematically combining between the 2 and 100 best performing individual models, using the four combination strategies: MIN, AVE, RANK and VOT.

The expectations on the ensemble approach were confirmed, since several combinations were found statistically outperforming the best individual models from a statistical viewpoint. The MIN operator consistently outperformed the other combination strategies, both in terms of AUC ROC in the retrospective screen and dispersion around the mean AUC ROC. The MIN combination of the 5 best models seemed to provide, in terms of AUC ROC, the best results in the retrospective screen (0.99, $p = 0.0011$ in comparison to the best individual model). The EF1% was also increased to 50. The enrichment metrics for the best ensembles are also shown in Table 3.

After selecting the MIN combination of the five best individual models (MIN-5), we resorted to PPV surface analysis to choose the optimal score threshold for prospective virtual screening applications. With the help of PPV surfaces, the evolution of the most relevant metric for our purposes, the PPV, i.e., the actual probability that a predicted hit will confirm activity when experimentally tested, can be visually optimized as a function of the (*Se/Sp*) ratio across a range of *Ya* values. For this analysis, we have considered that the association between the *Se/Sp* and the score values

of the MIN-5 model ensemble (observed in the retrospective screening campaign) would hold when performing screens on other libraries. This is a strong assumption that is not necessarily true. Nevertheless, since the AUC ROC values obtained for the retrospective screen are clearly high (0.99 for the best model ensemble) while on the other hand the Ya ratio (0.02) and size (> 1000 compounds) speak of a controlled statistical behaviour [44], we believe it is a valid assumption in the current setting.

Using PPV surfaces (Fig. 2), we have chosen 0.4075 as the MIN-5 ensemble score threshold to be used in our prospective virtual screening campaign. This score is associated to a *Se/Sp* ratio of 0.8199 and a PPV value above 0.25 for a *Ya* of 0.01. This means that, according to this theoretic analysis, if *Ya* in the real virtual screen was 0.01, we would have to submit about four predicted hits to experimental testing in order to find one confirmed hit.

The prospective virtual screen of DrugBank 5.0 using the previous score cut-off value resulted in 25 hits, with 21 of them belonging to the applicability domain of the five models included in the ensemble. Two of these 21 compounds had reports on inhibitory effects on human FABPs and belonged to the training set, thus being excluded from further analysis or testing. Based on the selection criteria previously listed under the "Methods" section and our funding availability, we acquired and submitted three hits to experimental testing: the anti-inflammatory agent sulindac, the antifungal clotrimazole and the antiseptic cetylpyridinium (Fig. 3).

For that purpose, recombinant EmFABPs 1 to 4 were successfully expressed and purified from *E. coli* cultures. To control for structural and functional integrity, the following

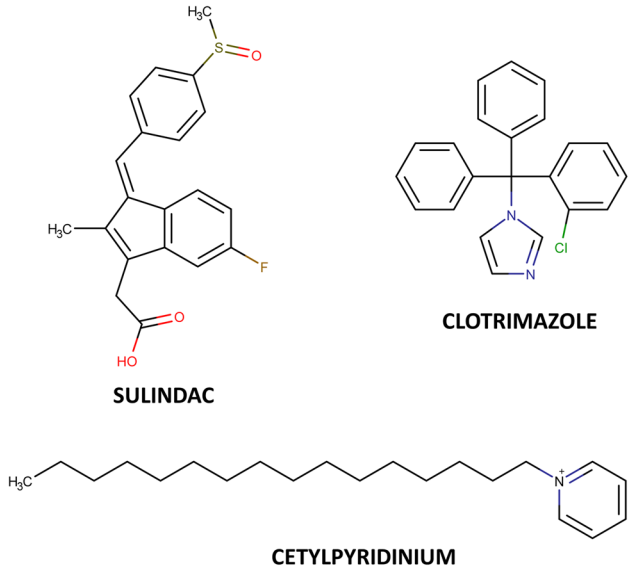

Fig. 3 Chemical structure of the in silico hits submitted to experimental validation

methods were employed: (a) circular dichroism spectroscopy, (b) unfolding transitions, and (c) oleic acid binding using displacement assays.

The far-UV circular dichroism (CD) spectra described for all FABPs show two characteristic features: a minimum at approximately 216 nm and a strong positive band occurring below 200 nm. This is in agreement with their overall structure that contains approximately 80% content of β-sheet and 20% α-helices [54, 55]. In the case of FABPs from *E. multilocularis* (Fig. 4), all of them showed minima

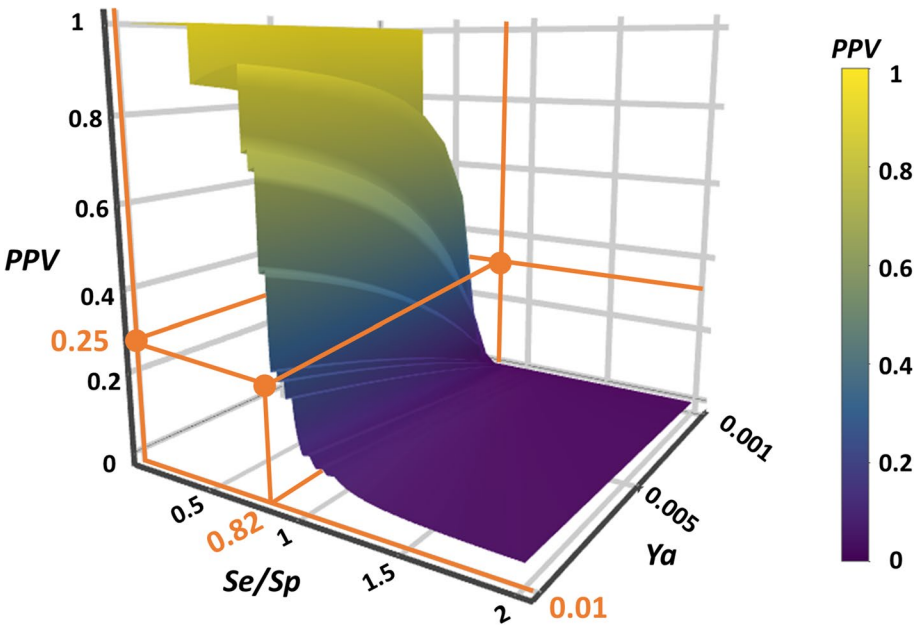

Fig. 2 PPV surface for the retrospective screen using the best 5-model ensemble. The *Se/Sp* ratio correspondent to the chosen score threshold value and the associated PPVs within the Ya 0.001–0.010 range are indicated

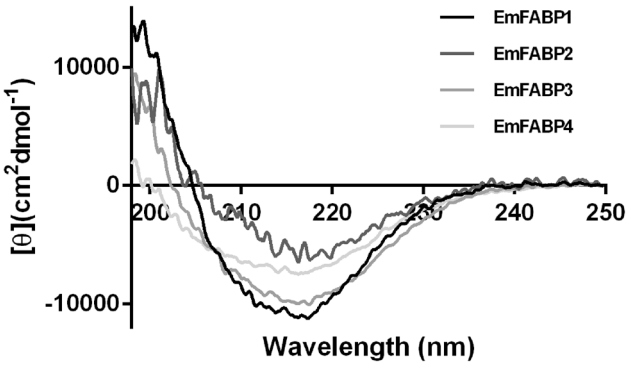

Fig. 4 The far-UV CD spectra of EmFABPs. CD spectra were collected in the far UV region 190–250 nm for EmFABP 1 to 4. Proteins (20–40 uM) were dissolved in PBS buffer (pH7.4) and spectra were recorded at 25 °C. Average of three spectra is shown for each EmFABP

at 216 nm but only EmFABP1-3 presented a strong positive band centered at 200 nm. For the case of EmFABP4 no signal was observed below 200 nm suggesting the lack of a helical content. This result could be due to the fact that EmFABP4 was the only EmFABP that behaved as a stable dimer in solution.

All EmFABPs except EmFABP2 unfold through a cooperative transition suggesting stable conformations (Fig. 5). Transitions were centered at 6.1, 6.35 and 4.9 M urea for EmFABP 1, 3 and 4, respectively. EgFABP1 was included as control and it presented a transition midpoint of 5.5 M urea. These results are in good agreement with previously reported midpoints for other vertebrates FABPs [56]. Since EmFABP2 showed to be highly unstable no further experiments were conducted on this protein.

In order to assess ligand binding capacity of the recombinant EmFABPs, binding affinity to oleic acid was measured with a competition assay. Briefly, increasing amounts

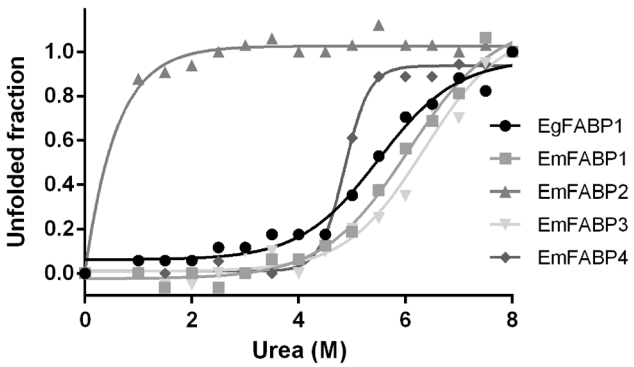

Fig. 5 Equilibrium unfolding transitions of EmFABPs 1 to 4 monitored by the change in fluorescence emission as a function of urea concentration. Data were plotted as the unfolded fraction of each protein vs urea concentration

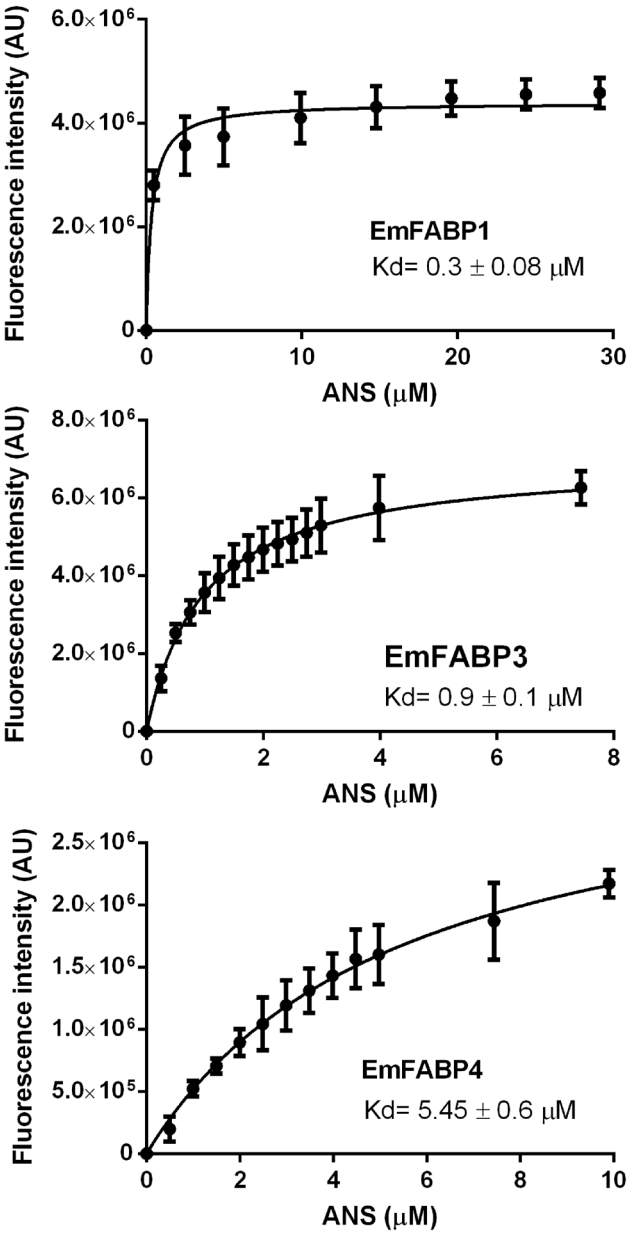

Fig. 6 Binding of ANS to EmFABPs. EmFABP1, 3 and 4 (1 μM) were titrated with the fluorescent probe ANS. Increasing amounts of the ligand were added to the protein solution, and the fluorescence intensity was measured after equilibration of the sample for 2 min. Excitation wavelength was 350 nm, and the emission was monitored at 380–600 nm for ANS, respectively. The solid lines correspond to the optimal curve fitting of the data by nonlinear regression to one site binding model

of oleic acid were added to a preformed protein-ANS complex. As a first step Kd values for ANS were determined (Fig. 6) in order to define further experimental conditions. Binding isotherms for ANS to EmFABP1, 3 and 4 yielded Kd values of 0.3 ± 0.08, 0.9 ± 0.1, and 5.4 ± 0.6 μM, respectively. All the EmFABPs showed Kd values similar

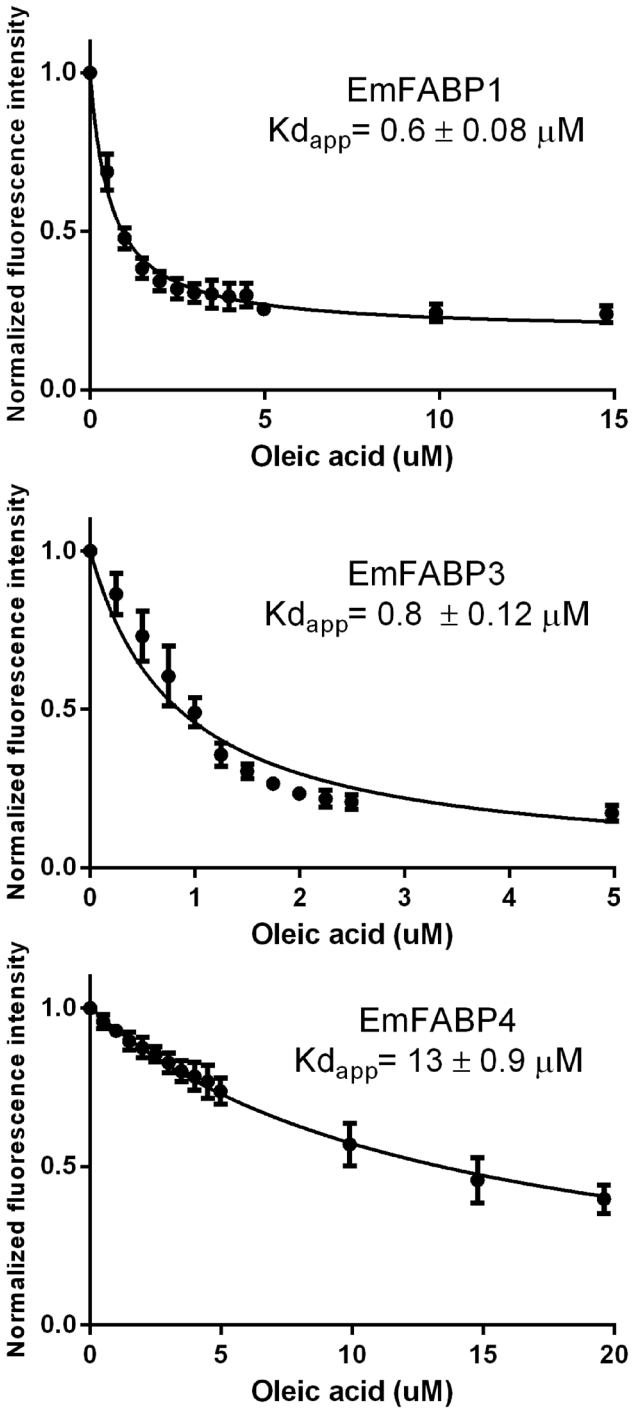

Fig. 7 Competition of ANS bound to EmFABPs by oleic acid. ANS bound to EmFABP 1, 3 and 4 was incubated with increasing amounts of oleic acid at 25 °C. The relative fluorescence intensity was plotted. Excitation wavelength was set at 350 nm. The continuous lines correspond to the fitting of Eq. (2)

to those previously reported for IFABP (Kd: 1–10 μM, see also [57, 58].

Oleic acid was able to displace ANS from EmFABPs cavities at different degrees. Kd_{app} values for EmFABPs 1, 3 and 4 were 0.6 ± 0.08, 0.8 ± 0.12 and 13 ± 0.9 μM, respectively (Fig. 7).

EmFABPs capacity to bind the three hits selected from the in silico screen was evaluated using competition assays. Increasing amounts of the hit compounds were added to a preformed protein-ANS complex. Figure 8 shows the degree of displacement of ANS exerted by different assayed drugs. At 15 μM concentration, sulindac displaced ANS from the EmFABPs 1, 3 and 4.

Discussion and conclusions

The current chemotherapy for echinococcosis is based on the use of benzimidazoles, particularly mebendazole and albendazole. This treatment has no lethal effects on parasites but severely affects proliferation resulting in the need of long lasting administration of drugs in order to control the cyst growth [3, 59]. Additionally, some patients present side effects and hepatotoxicity and in those cases no other alternative treatments are available [60].

Recent work reported the existence of six FABP encoding genes on the available *E. multilocularis* genome [13]. Two of these genes encode proteins with identical primary sequence thus they were named *emfabp1.1* and *emfabp1.2*. Regarding the other four genes, one is orthologous to *egfabp2* [61], whereas the other three represented novel genes (named *emfabp3 to 5*). Recently published transcriptomic information of *E. multilocularis,* show that all FABPs are transcribed at different stages of *E. multilocularis* [5, 62, 63]. According to the different sets of data, *emfabp1.1* and *emfabp1.2* are the most highly expressed FABP genes in *E. multilocularis*, it is important to recall that these two genes are identical [13].

In this work we have successfully purified in a recombinant form EmFABP 1 to 4 for the first time. All four EmFABPs showed a CD spectrum compatible with a high β-sheet content in their secondary structure and in agreement with the in silico predictions already analysed [13]. For EmFABP4, the most striking feature is that the predicted sequence is much longer (176 amino acids) than that for a regular FABP (around 130 amino acids). In this case, a typical FABP fold is predicted but no specific structure is assigned to the C terminus of the protein [13]. Surprisingly, EmFABP4 was purified as a dimer and this state seemed to be stable in solution. Whether EmFABP4 exists as a dimer in its biological environment remains to be elucidated. It is noticeable that FABPs with C-terminal extensions from *Fasciola* spp. have been also reported recently [12]. Except for the case of EmFABP2, all EmFABPs showed to be stable in solution presenting unfolding midpoints similar to other well characterised vertebrate FABPs [53–55].

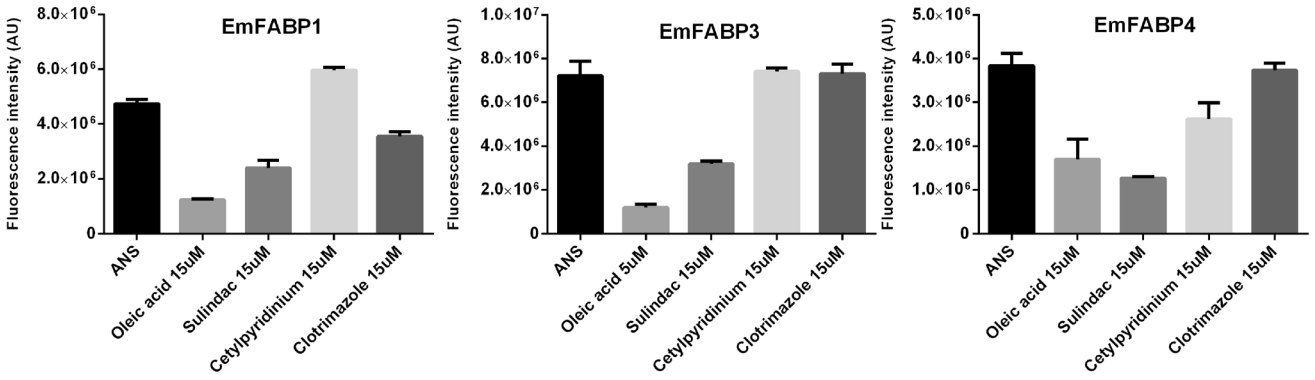

Fig. 8 Drug binding assay. Recombinant EmFABPs were incubated with 20 μM ANS (black bar). The displacement of ANS is shown as a loss of fluorescence intensity. Natural ligand (Oleic acid) was used as a positive control of displacement. Compounds were tested in a range of 0–30 μM. The figure shows a representative concentration. Two drugs (sulindac and clotrimazole) were diluted in DMSO. DMSO does not make a significant contribution at the intensity drop in the range used (data not shown)

As in vertebrate's FABPs, the sequence identity between different FABPs from invertebrates varies significantly but they still share their overall 3D protein structure. When compared to mammalian FABPs, they show higher similarity to those belonging to group IV [7, 64]. They all invariantly contain the P2 ligand-binding motif (an Arg… Arg-x-Tyr motif involved in ligand binding) [65, 66], with the sole exception of EmFABP4 that has a Tyr instead of an Arg in the first position. Noticeably, EmFABP1, 3 and 4 proved to bind oleic acid showing submicromolar affinities for EmFABP1 and 3, in agreement with reports for other FABPs from vertebrates and invertebrates [52, 53, 67, 68]. EmFABP4 showed considerably lower affinity in comparison with EmFABP1 and 3, which might be related to the fact that EmFABP4 exists as a dimer in vitro.

Vertebrates FABPs have drawn attention as drug targets for many diseases like obesity, atherosclerosis, diabetes and metabolic syndrome [69–71]. In this sense several compounds have been synthesized in order to achieve specificity and many patents were registered [72].

Using an ensemble learning approximation, we have performed a ligand-based virtual screening campaign to identify new *Echinococcus* spp. FABPs inhibitors as potential new treatments against echinococcosis. This is, to our best knowledge, the first virtual screening application to identify new inhibitors of such potential drug targets in invertebrates.

Although the individual models obtained through the random subspace approximation showed very good performance at a retrospective virtual screening experiment, we resorted to ensemble learning to achieve improved behaviour.

The three hits that emerged from the prospective in silico screen and were submitted to experimental assays confirmed the model ensemble predictions, as all of them displaced, at 15 μM concentrations, oleic acid from at least one of the four recombinant EmFABPs. The observed Positive Predictive Value (100% of confirmed predictions) exceeds our conservative theoretic estimation of a PPV of 25% for a hypothetic proportion of active compounds of 0.01 in the screened library, which confirms the utility of PPV surface analysis to optimize the score cut-off value that will be used in prospective screening applications.

Though the previous findings confirm the general validity of our strategy, contribute to the validation of the reported model ensemble and provide possible starting points to design new therapies against *Echinococcus*, it should be underlined that not all the three hits are equally attractive for further drug development, nor are they equally interesting as direct drug repurposing candidates. In particular, whereas cetylpyridinium has been widely used in mouthwashes, toothpaste, shampoos, deodorants, soaps and others, at low concentrations, as conservative and antiseptic, most undiluted quaternary ammonium salts are acutely toxic when given orally to mammals ([73] and refs therein), thus precluding their development as systemic medications.

Sulindac acts as a prodrug, that is converted in vivo to an active sulphide compound with anti-inflammatory activity by liver enzymes. Interestingly, the peak steady state plasma levels observed for sulindac and its sulphide at therapeutic doses are similar to the active concentrations reported here as EmFABPs inhibitor [74]. In contrast, the lower therapeutic doses in which clotrimazole is administered for the treatment of systemic mycoses, its poor oral absorption and extensive first pass metabolism result in much lower plasma concentrations, incompatible with the active concentrations reported here [75–77].

All in all and having in mind that it has proven effective in the displacement of ANS from the three recombinant FABPs assayed, sulindac seems to be the most interesting confirmed hit for further development of new treatments against echinococcosis.

Acknowledgements JAB, LNA, JLP, FNCR, SR, AT, BC and GRF, thank to National Research Council (CONICET). This work was supported by the CONICET (PIP 0739), the National Agency for Scientific and Technological Promotion (ANPCYT; PICT 2017-0643; PICT 2013-2121) and the National University of La Plata (UNLP).

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by JAB, LNA, JLP, FNCR and SR. The first draft of the manuscript was written by GRF, BC and AT and all authors commented on subsequent versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was financed by ANPCyT PICT 2017-0643, ANPCyT PICT 2013-2121, CONICET PIP 0739, Incentivos UNLP and UNLP Young Scholar's grants.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This work has been conducted in accordance with all the ethical standards.

References

1. WHO (2020) World Health Organization. https://www.who.int/neglected_diseases/diseases/en/. Accessed 14 May 2020
2. Stojkovic M, Zwahlen M, Teggi A et al (2009) Treatment response of cystic echinococcosis to benzimidazoles: a systematic review. PLoS Negl Trop Dis. https://doi.org/10.1371/journal.pntd.0000524
3. Gottstein B, Stojkovic M, Vuitton DA et al (2015) Threat of alveolar echinococcosis to public health—a challenge for Europe. Trends Parasitol 31:407–412
4. Čudeková P, Várady M, Dolinská M, Königová A (2010) Phenotypic and genotypic characterisation of benzimidazole susceptible and resistant isolates of *Haemonchus contortus*. Vet Parasitol 172:155–159. https://doi.org/10.1016/j.vetpar.2010.04.022
5. Tsai IJ, Zarowiecki M, Holroyd N et al (2013) The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496:57–63. https://doi.org/10.1038/nature12031
6. Frayha GJ (1971) Comparative metabolism of acetate in the taeniid tapeworms *Echinococcus granulosus*, *E. multilocularis* and *Taenia hydatigena*. Comp Biochem Phys B 39:167–170. https://doi.org/10.1016/0305-0491(71)90264-1
7. Smathers RL, Petersen DR (2011) The human fatty acid-binding protein family: evolutionary divergences and functions. Hum Genom 5:1–22
8. Storch J, Corsico B (2008) The emerging functions and mechanisms of mammalian fatty acid-binding proteins. Annu Rev Nutr 28:73–95. https://doi.org/10.1146/annurev.nutr.27.061406.093710
9. Haunerland NH, Spener F (2004) Fatty acid-binding proteins—insights from genetic manipulations. Prog Lipid Res 43:328–349. https://doi.org/10.1016/j.plipres.2004.05.001
10. Pórfido JL, Alvite G, Silva V et al (2012) Direct interaction between EgFABP1, a fatty acid binding protein from *Echinococcus granulosus*, and phospholipid membranes. PLoS Negl Trop Dis 6:e1893
11. Becker MM, Kalinna BH, Waine GJ, McManus DP (1994) Gene cloning, overproduction and purification of a functionally active cytoplasmic fatty acid-binding protein (Sj-FABPc) from the human blood fluke *Schistosoma japonicum*. Gene 148:321–325. https://doi.org/10.1016/0378-1119(94)90706-4
12. Morphew RM, Wilkinson TJ, MacKintosh N et al (2016) Exploring and expanding the fatty-acid-binding protein superfamily in fasciola species. J Proteome Res 15:3308–3321. https://doi.org/10.1021/acs.jproteome.6b00331
13. Pórfido JL, Herz M, Kiss F et al (2020) Fatty acid-binding proteins in *Echinococcus* spp.: the family has grown. Parasitol Res 119:1401–1408. https://doi.org/10.1007/s00436-020-06631-5
14. Klug DM, Gelb MH, Pollastri MP (2016) Repurposing strategies for tropical disease drug discovery. Bioorg Med Chem Lett 26:2569–2576
15. Nallan L, Bauer KD, Bendale P et al (2005) Protein farnesyltransferase inhibitors exhibit potent antimalarial activity. J Med Chem 48:3704–3713. https://doi.org/10.1021/jm0491039
16. Bland ND, Wang C, Tallman C et al (2011) Pharmacological validation of *Trypanosoma brucei* phosphodiesterases B1 and B2 as druggable targets for African sleeping sickness. J Med Chem 54:8188–8194. https://doi.org/10.1021/jm201148s
17. Amata E, Bland ND, Hoyt CT et al (2014) Repurposing human PDE4 inhibitors for neglected tropical diseases: design, synthesis and evaluation of cilomilast analogues as *Trypanosoma brucei* PDEB1 inhibitors. Bioorg Med Chem Lett. https://doi.org/10.1016/j.bmcl.2014.07.063
18. Barf T, Lehmann F, Hammer K et al (2009) N-benzyl-indolo carboxylic acids: design and synthesis of potent and selective adipocyte fatty-acid binding protein (A-FABP) inhibitors. Bioorg Med Chem Lett 19:1745–1748. https://doi.org/10.1016/j.bmcl.2009.01.084
19. Beniyama Y, Matsuno K, Miyachi H (2013) Structure-guided design, synthesis and in vitro evaluation of a series of pyrazole-based fatty acid binding protein (FABP) 3 ligands. Bioorg Med Chem Lett 23:1662–1666. https://doi.org/10.1016/j.bmcl.2013.01.054
20. Ringom R, Axen E, Uppenberg J et al (2004) Substituted benzylamino-6-(trifluoromethyl)pyrimidin-4(1H)-ones: a novel class of selective human A-FABP inhibitors. Bioorg Med Chem Lett 14:4449–4452. https://doi.org/10.1016/j.bmcl.2004.06.058
21. Sulsky R, Magnin DR, Huang Y et al (2007) Potent and selective biphenyl azole inhibitors of adipocyte fatty acid binding protein (aFABP). Bioorg Med Chem Lett 17:3511–3515. https://doi.org/10.1016/j.bmcl.2006.12.044
22. Van Dongen MJP, Uppenberg J, Svensson S et al (2002) Structure-based screening as applied to human FABP4: a highly efficient alternative to HTS for hit generation. J Am Chem Soc 124:11874–11880. https://doi.org/10.1021/ja017830c
23. Zhou Y, Nie T, Zhang Y et al (2016) The discovery of novel and selective fatty acid binding protein 4 inhibitors by virtual screening and biological evaluation. Bioorg Med Chem 24:4310–4317. https://doi.org/10.1016/j.bmc.2016.07.022
24. Wang YT, Liu CH, Zhu HL (2016a) Fatty acid binding protein (FABP) inhibitors: a patent review (2012–2015). Expert Opin Ther 26:767–776
25. Cai H, Yan G, Zhang X et al (2010) Discovery of highly selective inhibitors of human fatty acid binding protein 4 (FABP4) by virtual screening. Bioorg Med Chem Lett 20:3675–3679. https://doi.org/10.1016/j.bmcl.2010.04.095
26. Cai HY, Wang T, Zhao JC et al (2013) Benzbromarone, an old uricosuric drug, inhibits human fatty acid binding protein 4 in vitro and lowers the blood glucose level in db/db mice. Acta Pharmacol Sin 34:1397–1402. https://doi.org/10.1038/aps.2013.97

27. Cai H, Liu Q, Gao D et al (2015) Novel fatty acid binding protein 4 (FABP4) inhibitors: virtual screening, synthesis and crystal structure determination. Eur J Med Chem 90:241–250. https://doi.org/10.1016/j.ejmech.2014.11.020
28. Kaczocha M, Rebecchi MJ, Ralph BP et al (2014) Inhibition of fatty acid binding proteins elevates brain anandamide levels and produces analgesia. PLoS ONE. https://doi.org/10.1371/journal.pone.0094200
29. Lehmann F, Haile S, Axen E et al (2004) Discovery of inhibitors of human adipocyte fatty acid-binding protein, a potential type 2 diabetes target. Bioorg Med Chem Lett 14:4445–4448. https://doi.org/10.1016/j.bmcl.2004.06.057
30. Liu X, Huang X, Lin W et al (2011) New aromatic substituted pyrazoles as selective inhibitors of human adipocyte fatty acid-binding protein. Bioorg Med Chem Lett 21:2949–2952. https://doi.org/10.1016/j.bmcl.2011.03.063
31. Lan H, Cheng CC, Kowalski TJ et al (2011) Small-molecule inhibitors of FABP4/5 ameliorate dyslipidemia but not insulin resistance in mice with diet-induced obesity. J Lipid Res 52:646–656. https://doi.org/10.1194/jlr.M012757
32. Martin TM, Harten P, Young DM et al (2012) Does rational selection of training and test sets improve the outcome of QSAR modeling? J Chem Inf Model 52:2570–2578. https://doi.org/10.1021/ci300338w
33. Leonard JT, Roy K (2006) On selection of training and test sets for the development of predictive QSAR models. QSAR Comb Sci 25:235–251. https://doi.org/10.1002/qsar.200510161
34. Everitt BS et al (2011) In: Everitt BS (ed) Cluster analysis, 5th Edn. Wiley, Hoboken
35. Alberca LN, Sbaraglini ML, Morales JF et al (2018) Cascade ligand- and structure-based virtual screening to identify new trypanocidal compounds inhibiting putrescine uptake. Front Cell Infect Microbiol. https://doi.org/10.3389/fcimb.2018.00173
36. Alberca LN, Chuguransky SR, Álvarez CL et al (2019) In silico guided drug repurposing: discovery of new competitive and non-competitive inhibitors of falcipain-2. Front Chem. https://doi.org/10.3389/fchem.2019.00534
37. Gantner ME, Peroni RN, Morales JF et al (2017) Development and validation of a computational model ensemble for the early detection of BCRP/ABCG2 substrates during the drug design stage. J Chem Inf Model 57:1868–1880. https://doi.org/10.1021/acs.jcim.7b00016
38. Yu G, Zhang G, Domeniconi C et al (2012) Semi-supervised classification based on random subspace dimensionality reduction. Pattern Recognit 45:1119–1135. https://doi.org/10.1016/j.patcog.2011.08.024
39. El Habib DM, Amine Chikh M (2015) Combining bootstrapping samples, random subspaces and random forests to build classifiers. J Med Imaging Heal Inform 5:539–544. https://doi.org/10.1166/jmihi.2015.1423
40. Toropova AP, Toropov AA (2017) CORAL: binary classifications (active/inactive) for drug-induced liver injury. Toxicol Lett 268:51–57. https://doi.org/10.1016/j.toxlet.2017.01.011
41. Gramatica P (2013) On the development and validation of QSAR models. Methods Mol Biol 930:499–526. https://doi.org/10.1007/978-1-62703-059-5_21
42. Roy K, Mitra I (2011) On various metrics used for validation of predictive QSAR Models with applications in virtual screening and focused library design. Comb Chem High Throughput Screen 14:450–474. https://doi.org/10.2174/138620711795767893
43. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J Med Chem 55:6582–6594. https://doi.org/10.1021/jm300687e
44. Truchon JF, Bayly CI (2007) Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. J Chem Inf Model 47:488–508. https://doi.org/10.1021/ci600426e
45. Carbonneau MA, Granger E, Raymond AJ, Gagnon G (2016) Robust multiple-instance learning ensembles using random subspace instance selection. Pattern Recognit 58:83–99. https://doi.org/10.1016/j.patcog.2016.03.035
46. Min S-H (2016) A genetic algorithm-based heterogeneous random subspace ensemble model for bankruptcy prediction. Int J Appl Eng Res 11:2927–2931
47. Zhang Q, Muegge I (2006) Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring. J Med Chem 49:1536–1548. https://doi.org/10.1021/jm050468i
48. Robin X, Turck N, Hainard A et al (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. https://doi.org/10.1186/1471-2105-12-77
49. Lätti S, Niinivehmas S, Pentikäinen OT (2016) Rocker: open source, easy-to-use tool for AUC and enrichment calculations and ROC visualization. J Cheminform. https://doi.org/10.1186/s13321-016-0158-y
50. Wishart DS, Feunang YD, Guo AC et al (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. https://doi.org/10.1093/nar/gkx1037
51. Schmid F (1989) Spectral methods of characterizing protein conformation and conformational changes. In: Creighton TE (ed) Protein structure: a practical approach. IRL Press, Oxford, p 251
52. Richieri GV, Ogata RT, Kleinfeld AM (1994) Equilibrium constants for the binding of fatty acids with fatty acid- binding proteins from adipocyte, intestine, heart, and liver measured with the fluorescent probe ADIFAB. J Biol Chem 269:23918–23930
53. Zimmerman AW, Veerkamp JH (2002) New insights into the structure and function of fatty acid-binding proteins. Cell Mol Life Sci 59:1096–1116. https://doi.org/10.1007/s00018-002-8490-y
54. Curto LM, Caramelo JJ, Delfino JM (2005) Δ98Δ, a functional all-β-sheet abridged form of intestinal fatty acid binding protein. Biochemistry 44:13847–13857. https://doi.org/10.1021/bi051080s
55. Franchini GR, Curto LM, Caramelo JJ, Delfino JM (2009) Dissection of a β-barrel motif leads to a functional dimer: the case of the intestinal fatty acid binding protein. Protein Sci 18:2592–2602. https://doi.org/10.1002/pro.273
56. Ropson IJ, Gordon JI, Frieden C, Gordon JI (1990) Folding of a predominantly β-structure protein: rat intestinal fatty acid binding protein. Biochemistry 29:9591–9599. https://doi.org/10.1021/bi00493a013
57. Arighi CN, Rossi JPFC, Delfino JM (2003) Temperature-induced conformational switch in intestinal fatty acid binding protein (IFABP) revealing an alternative mode for ligand binding. Biochemistry 42:7539–7551. https://doi.org/10.1021/bi020680d
58. Pastukhov AV, Ropson IJ (2003) Fluorescent dyes as probes to study lipid-binding proteins. Proteins Struct Funct Genet 53:607–615. https://doi.org/10.1002/prot.10401
59. Hemphill A, Stadelmann B, Rufener R et al (2014) Treatment of echinococcosis: albendazole and mebendazole—what else ? Parasite. https://doi.org/10.1051/parasite/2014073
60. Reuter S, Buck A, Manfras B et al (2004) Structured treatment interruption in patients with alveolar echinococcosis. Hepatology 39:509–517. https://doi.org/10.1002/hep.20078
61. Esteves A, Portillo V, Ehrlich R (2003) Genomic structure and expression of a gene coding for a new fatty acid binding protein from *Echinococcus granulosus*. Biochim Biophys Acta - Mol Cell Biol Lipids 1631:26–34. https://doi.org/10.1016/S1388-1981(02)00321-9
62. Zheng Y, Blair D, Bradley JE (2013) Phyletic distribution of fatty acid-binding protein genes. PLoS ONE. https://doi.org/10.1371/journal.pone.0077636

63. Huang F, Dang Z, Suzuki Y et al (2016) Analysis on gene expression profile in oncospheres and early stage metacestodes from *Echinococcus multilocularis*. PLoS Negl Trop Dis 10:e0004634. https://doi.org/10.1371/journal.pntd.0004634
64. Esteves A, Ehrlich R (2006) Invertebrate intracellular fatty acid binding proteins. Comp Biochem Physiol C 142:262–274
65. Jones TA, Bergfors T, Sedzik J, Unge T (1988) The three-dimensional structure of P2 myelin protein. EMBO J 7:1597–1604
66. Jakobsson E, Alvite G, Bergfors T et al (2003) The crystal structure of *Echinococcus granulosus* fatty-acid-binding protein 1. Biochim Biophys Acta 1649:40–50. https://doi.org/10.1016/S1570-9639(03)00151-1
67. Angelucci F, Johnson KA, Baiocco P et al (2004) Schistosoma mansoni fatty acid binding protein: specificity and functional control as revealed by crystallographic structure. Biochemistry 43:13000–13011. https://doi.org/10.1021/bi048505f
68. Alvite G, Di Pietro SM, Santomé JA et al (2001) Binding properties of *Echinococcus granulosus* fatty acid binding protein. Biochim Biophys Acta - Mol Cell Biol Lipids 1533:293–302. https://doi.org/10.1016/S1388-1981(01)00164-0
69. Furuhashi M, Tuncman G, Görgün CZ et al (2007) Treatment of diabetes and atherosclerosis by inhibiting fatty-acid-binding protein aP2. Nature 447:959–965. https://doi.org/10.1038/nature05844
70. Furuhashi M, Hotamisligil GS (2008) Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. Nat Rev Drug Discov 7:489–503
71. Roden M (2007) Blocking fatty acids' mystery tour: a therapy for metabolic syndrome? Cell Metab 6:89–91. https://doi.org/10.1016/j.cmet.2007.07.008
72. Wang YT, Liu CH, Zhu HL (2016b) Fatty acid binding protein (FABP) inhibitors: a patent review (2012–2015). Expert Opin Ther Pat 26:767–776
73. Lin GHY, Voss KA, Davidson TJ (1991) Acute inhalation toxicity of cetylpyridinium chloride. Food Chem Toxicol 29:851–854. https://doi.org/10.1016/0278-6915(91)90113-L
74. Swanson BN, Boppana VK, Vlasses PH et al (1982) Sulindac disposition when given once and twice daily. Clin Pharmacol Ther 32:397–403. https://doi.org/10.1038/clpt.1982.178
75. Utz JP (1975) New drugs for the systemic mycoses: flucytosine and clotrimazole. Bull N Y Acad Med 51:1103
76. Groll AH, Piscitelli SC, Walsh TJ (1998) Clinical pharmacology of systemic antifungal agents: a comprehensive review of agents in clinical use, current investigational compounds, and putative targets for antifungal drug development. Advances in pharmacology. Academic Press Inc., Cambridge, pp 343–500
77. Shord SS, Chan LN, Camp JR et al (2010) Effects of oral clotrimazole troches on the pharmacokinetics of oral and intravenous midazolam. Br J Clin Pharmacol 69:160–166. https://doi.org/10.1111/j.1365-2125.2009.03559.x