Parasitology Research (2019) 118:3159–3171
https://doi.org/10.1007/s00436-019-06445-0

TREATMENT AND PROPHYLAXIS - ORIGINAL PAPER

Identification of adaptive inhibitors of *Cryptosporidium parvum* fatty acyl-coenzyme A synthetase isoforms by virtual screening

Somdeb Chattopadhyay[1] · Rajani Kanta Mahapatra[1]

Received: 7 February 2019 / Accepted: 27 August 2019 / Published online: 5 September 2019

Abstract

Cryptosporidiosis is a significant cause of gastroenteritis in both humans and livestock in developing countries. The only FDA-approved drug available against the same is nitazoxanide, with questionable efficacy in malnourished children and immunocompromised patients. Recent in vitro studies have indicated the viability of Triacsin C as a potential drug candidate, which targets the parasite's long-chain fatty acyl coenzyme A synthetase enzyme (LC-FACS), a critical component of the fatty acid metabolism pathway. We have used this molecule as a baseline to propose more potent versions thereof. We have applied a combined approach of substructure replacement, literature search, and database screening to come up with 514 analogs of Triacsin C. A virtual screening protocol was carried out which lead us to identify a potential hit compound. This was further subjected to a 100-ns molecular dynamics simulation in complex to determine its stability and binding characteristics. After which, the ADME/tox properties were predicted to assess its viability as a drug. The molecule R134 was identified as the best hit due to its highest average binding affinity, stability in complex when subjected to MD simulations, and reasonable predicted ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) properties comparable to those of the Triacsin C parent molecule. We have proposed R134 as a putative drug candidate against the *Cryptosporidium parvum* LC-FACS enzyme isoforms, following an in silico protocol. We hope the results will be helpful when planning future in vitro experiments for identifying drugs against Cryptosporidium.

Keywords Cryptosporidium · Fatty acyl coenzyme A synthetase · Cheminformatics · Virtual screening · Molecular dynamics simulation

Introduction

Protozoans of genus Cryptosporidium are opportunistic apicomplexan parasites infecting the ileal lumen of the small intestine in a majority of known vertebrates, including humans (Borowski et al. 2010). The parasite has been linked to multiple water-borne outbreaks of gastroenteritis (Hayes et al. 1989; Corso et al. 2003; Mac Kenzie et al. 1994) and has been recognized by the US Centers for Disease Control and Prevention (CDC) as a Category B Potential Bioterrorism Agent (Rotz et al. 2002), infecting the apical surface microvilli of epithelial cells in the ileum, starting a six-stage monoxenous life cycle (O'Donoghue 1995). The infection is largely self-limiting and asymptomatic in immunocompetent adults, with a mean duration of 13 days, although extreme cases can be accompanied by clinical symptoms such as diarrhea, abdominal pain, nausea, vomiting, and anorexia. The immune status of the host in particular seems to play a major role in deciding the severity of clinical symptoms (O'Donoghue 1995; Hunter and Nichols 2002). Immunocompromised individuals with SCID, hyper-IgM syndrome, or HIV patients, whose circulating CD4+ T cell count has fallen below 50 per cubic millimeters of blood, experience the highest morbidity and mortality with chronic cholera like diarrhea, vomiting, abdominal pain, pancreatitis, liver cirrhosis, and respiratory sinusitis accompanied by loss of weight. It can also lead to malnutrition and compromised cognitive development in children (Puleston et al. 2014; Guerrant 1997). The genus Cryptosporidium is unique among

Section Editor: Lihua Xiao

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00436-019-06445-0) contains supplementary material, which is available to authorized users.

✉ Rajani Kanta Mahapatra
rmahapatra@kiitbiotech.ac.in

[1] School of Biotechnology, KIIT Deemed to be University, Bhubaneswar, India

apicomplexans in that they lack a majority of the metabolic pathways which are otherwise common among other members of the phylum (Wanyiri and Ward 2006). With the exception of basic energy and lipid metabolism, the lack of major biosynthetic machinery means that the parasite is largely dependent upon the host organism for its nutritional requirements, which includes sequestration of fatty acids and amino acids. Due to this minimalist metabolism and a corresponding absence of common biochemical pathways, Cryptosporidium infections fail to respond favorably to chemotherapies which are otherwise effective for other apicomplexans. This work specifically focuses on finding adaptive inhibitors for the vulnerable lipid metabolism pathway in *Cryptosporidium parvum* IOWA-II (henceforth referred to as *C. parvum*) because of its ubiquitous presence and ability to infect both humans and livestock, thus having a greater potential for damage. Despite ongoing efforts (Tzipori and Ward 2002) to find chemotherapeutic interventions, currently available options for treatments are rather sparse (Chellan et al. 2017; Azam et al. 2015; Ryan and Hijjawi 2015). For immunocompetent individuals, the only FDA-approved molecule as of now is nitazoxanide, but its efficacy is inconclusive in malnourished children (Samie et al. 2015). Paromomycin, a broad spectrum antibiotic which is not approved by the FDA, is sometimes used to treat AIDS patients suffering from opportunistic Cryptosporidium infections. However, studies have put its efficacy into question by showing that its effects are comparable to placebo (Hewitt et al. 2000). The fatty acid metabolism of *C. parvum* is a particularly favorable target for new drugs, largely due to the fact that downstream pathways involve lipid synthesis and membrane remodeling, which are crucial to the organism's survival. One component of this pathway is the AMP-binding, Mg^{++}-dependent, long-chain fatty acyl coenzyme A synthetase (LC-FACS) [EC 6.2.1.3], catalyzing a two-step thioesterification reaction involving free fatty acid and coenzyme A as substrates, leading to the formation of fatty-acyl coenzyme A. This process of activation of fatty acids into acyl-CoA thioesters by LCFACS allows them to act as intermediates for various energy metabolism pathways such as beta-oxidation, protein modifications, cellular transport, signaling, and proliferation. The *C. parvum* genome encodes three different isoforms of LC-FACS: (Cryptosporidium parvum acyl coenzyme A synthetase) CpACS1, CpACS2, and CpACS3, each with different spatio-temporal expression patterns. CpACS3 has been observed to be expressed primarily during the oocyst and sporozoite stages of the parasite's life cycle. In contrast, CpACS1 is expressed in the anterior portion of the free sporozoites and merozoites during the early intercellular developmental stages. CpACS2 is speculated to have a largely housekeeping role, such as salvaging of fatty acids from the host, considering the fact that it is expressed almost constitutively throughout the lifecycle of the parasite and can be observed in high concentrations in its cytoplasm (Guo 2014). Potent CpACS inhibitors already exist. One such molecule is 2,4,7-undecatrienal nitrosohydrazone, also known by its common name, Triacsin C ($C_{11}H_{19}N_3O$, PubChem ID: 9576787), which incidentally also doubles as a potent vasodilator. Isolated as a secondary metabolite of the fungus Streptomyces aureofaciens (Hiroshi et al. 1987), this highly hydrophobic molecule resembles a polyunsaturated C11 fatty acid, with the carboxyl group at the alkenyl chain terminus substituted by an *N*-hydroxytriazene chemotype, giving the molecule an acidic character (Yoshida et al. 1982). Recent studies (Guo et al. 2013) have revealed that Triacsin C acts as a competitive inhibitor for CpACS1 and CpACS2, with IC50 values of 3.7 and 2.32 μmol L^{-1} respectively, which are 10X smaller than what would be the required dose for paromomycin and nitazoxanide, which are the current recommendatory drugs for the treatment of cryptosporidiosis. Also, the effective dose for treating mice with Triacsin C was approximately 8 to 15 mg kg^{-1} day^{-1}, significantly lower than that of nitazoxanide, with a very low risk of toxicity as a consequence (Guo et al. 2013). In this work, we attempt to identify putative functional analogs of Triacsin C, with the criteria that they are as follows: (1) more potent than the parent compound, (2) active against both CpACS1 and CpACS2 isoforms, (3) display favorable ADMET characteristics, and (4) easy to synthesize by medicinal chemists. To this end, we generated bioisosteres of Triacsin C by scaffold hopping, coupled with screening of databases and the literature to identify any pre-existing structural analogs to build up a screening library. Virtual screening was carried out to rank the molecules based on their predicted binding affinities, followed by repeated dockings to validate the docked poses of the six top ranking potential candidate ligands. We took the top three ranked molecules among them for ADMET predictions. Finally, we identified a common molecule among the top three candidates for both the CpACS isoforms, ran a 100-ns molecular dynamics simulation to observe its interactions and stability when bound to the proteins, and subsequently proposed it as an adaptive inhibitor. Figure 1 illustrates the structures of the identified molecules.

Methods

LC-FACS tertiary structure prediction and refinement

The amino acid sequences for CpACS1 and CpACS2 isoforms, having accession codes XM626649 and XM626248 respectively, were retrieved from GenBank in the FASTA format. They are monomeric enzymes, 683 and 685 amino acids long, encoded by single exon genes located in chromosomes 3 and 5 respectively for CpACS 1 and 2. The sequences were submitted to the NCBI PSI-BLAST server and run against all entries available in the PDB database, using the default

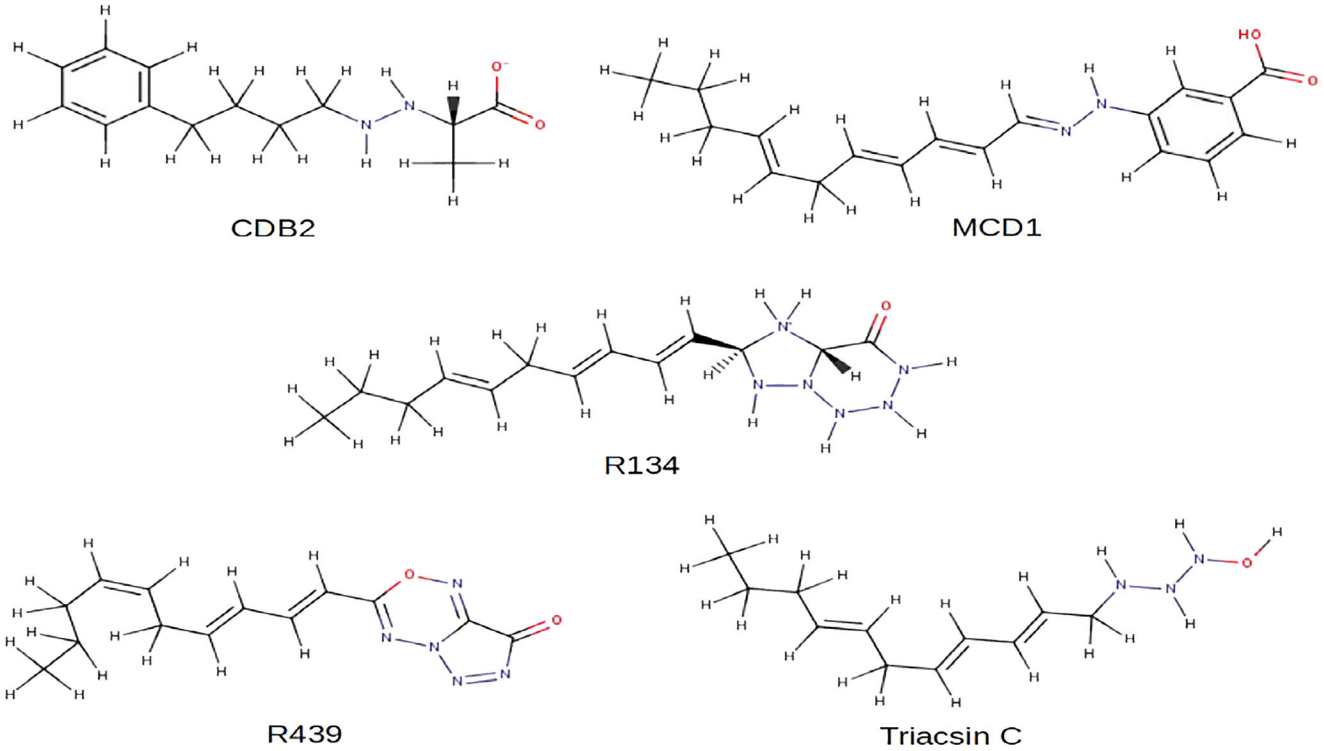

Fig. 1 Molecules identified as top hits against CpACS1 and CpACS2 after the screening process, along with the Triacsin C parent molecule for comparison. Please see Supplementary Tables S2 and S3 for more details

parameters. Both the isoforms were found to have less than 35% sequence identity with the PDB entries. A fold recognition approach using the I-TASSER server (Roy et al. 2010) was thus used to generate tertiary structures for both the isoforms. Of the multiple models generated per submission, the best one was selected based on the lowest C score. The RAMPAGE server was used to generate the Ramachandran plots for both the models to assess their quality. Both had significant problems in their structures, with CpACS1 having 65% residues in the favorable region and 14.8% in the outliers, while CpACS2 favored slightly better with 80% and 7.2% residues in the favorable and outlier regions respectively. The models were thus submitted to the Galaxy Refine server (Heo et al. 2013) for molecular dynamics assisted structure perturbation and refinement. This process was repeated until quality of the structures failed to increase any further, as evidenced by the Ramachandran plots. YASARA minimization server (Krieger et al. 2002) was used to further relieve any internal structural stresses within the protein models. The structures were uploaded to the SAVESv5 meta-server for quality assessment, and portions of the structure reported to have errors above threshold were remodeled using aggressive refinement in MODELLER v9.19. The residues were then minimized using the AMBER ff14SB force field in a vacuum, using 200 steps of steepest descent minimization in UCSF Chimera (v1.12RC). The I-TASSER and 3DLigandSite (Wass et al. 2010) servers were used to predict the binding pocket, and the residues which were common to both the server predictions were considered. The Mg^{++} ion binding site for both isoforms were predicted using a combination of ProBIS (Konc and Janežič 2010) and 3DLigandSite servers, based on a consensus between the two. Using UCSF Chimera, residues within a 5-Å radius of the Mg^{++} ion were selected, and 200 steps of steepest descent energy minimization were carried out using the AMBER ff14SB force field with AM1BCC charges. The PDB files were then manually edited to fix the ion in place using available metal bonds. The resulting structures have been made available for perusal. The long-chain fatty acyl-coenzyme A synthetase family of proteins contain two conserved motifs: an ACS signature motif (TGDIxxxxxxGxxxIxDRxK) and an ATP/AMP binding motif (TSGTTGxPK). The EMBOSS Water server using the Smith-Waterman local sequence alignment algorithm was used to align and identify the positions of these two motifs in both the isoforms (Supplementary Figs. S1 and S2).

Preparation of screening library

The screening library ($n = 515$, including Triacsin C itself) was prepared using a combination of Triacsin C analogs obtained from literature ($n = 7$), a similarity search on PubChem ($n = 3$), bioisosteres generated using scaffold hopping ($n = 477$), and a few molecules created manually by mimicking the Triacsin C pharmacophoric features ($n = 7$). In addition, molecules with similar properties to Triacsin C were obtained from the ChEMBL(v23) and DrugBank(v5.0.1) databases

($n = 20$). The Spark (v10.5) software package from Cresset was used to perform scaffold hops on the *N*-hydroxytriazene chemotype of Triacsin C. This is because a majority of the reported analogs maintained the unsaturated and highly hydrophobic alkenyl chain intact (Kim et al. 2012), while mutating on the terminal moiety. Spark performs bioisosteric replacements on a selected substructure of the parent molecule by searching a database of fragments derived from the curated ChEMBL and VEHICLe chemical structure databases. The fragments are ranked internally in Spark based on how often they appear in the literature or in commercially available compounds. In this case, all fragments from the available databases in Spark were used, ranging in ranks from "Very Rare" to "Very Common." A total of 1000 bioisosteres were generated and ranked based on their 3D pharmacophore alignment scores in comparison to the minimized Triacsin C parent molecule. The molecules with chemotypes such as Michael acceptors or thiols were automatically filtered out, resulting in a reduced set of 808 molecules. Their synthetic accessibility score (SA score) (Ertl and Schuffenhauer 2009) was calculated using the SAScorer.py Python script from the RDKit cheminformatics library. The SA score is a mathematical function based on the structural features of a molecule, which puts a quantitative measure, from a range of 1 to 10, on how difficult it might be for a medicinal chemist to synthesize the same. In a further attempt at refinement, the R cheminformatics packages ChemmineR (v3.32.1) and ChemmineOB (v1.18.0) from Bioconductor were used to perform a SMARTS pattern based substructure search on the molecules for potentially unwanted functional groups (Walters and Murcko 2002) as seen in Table 1. The resulting 477 molecules which passed the filters were converted to 3D, minimized with 1000 steps of steepest descent using the General AMBER force field (GAFF), protonated to pH 7.4, and exported as SD files using OpenBabel (v2.3.2). In order to identify Triacsin C analogs from ChEMBL and DrugBank, both these databases were downloaded in their entirety and concatenated into a single SD file (1,735,864 molecules). The CDK (Chemistry Development Toolkit) cheminformatics nodes for the KNIME (Konstanz Information Miner) data analytics platform (v3.5) (Beisken et al. 2013) were then used to calculate their molecular descriptors ($n = 48$), excluding the descriptors "BCUT" and "Charged Partial Surface Area (3D)" due to multiple runtime errors. The data was exported as a CSV file (set A). Please refer to Supplementary Figs. S11a and S13 for the KNIME workflow and an overview of the entire process. A similar method was followed to calculate the descriptors for the generated bioisosteres and Triacsin C analogs obtained from the literature and PubChem searches (set B). A filter was implemented in R, which calculated the mean and standard deviation of each of the descriptors (columns) in set B, excluding those with zero variance ($n = 3$). A molecule from set A, represented as a 43-dimensional vector, was allowed to pass through only if each of its properties fell within 1.7 × SD from the mean of each of the corresponding properties from set B. This process was repeated for each of the molecules in set A, which resulted in twenty of them passing the filters. These were converted to 3D from SMILES strings, minimized, protonated, and exported as SDFs. The final screening library contained a total of 514 molecules analogous to the Triacsin C parent (Supplementary Fig. S13). For each molecule in the screening library, the standard fingerprint was calculated using the KNIME CDK nodes. The Tanimoto coefficient measures the divergence between two molecules (a and b) represented as binary bit vectors (fingerprints) as per the following function:

$$\text{Tanimoto}(a,b) = \frac{|a \cap b|}{|a| + |b| - |a \cap b|}$$

which gives an indication as to the diversity of the screening library with respect to Triacsin C. Supplementary Figure S11b shows the KNIME workflow used to calculate the same.

Virtual screening

MGLTools (v1.5.6) from the Scripps Research Institute was used to prepare the CpACS model structures for docking. The CpACS models were protonated and saved in the AutoDock PDBQT file format. These files were manually edited to assign a +2 charge to the magnesium ion in the binding pocket. For CpACS1, the grid box center coordinates were set to $x = 71.75$, $y = 78.286$, and $z = 72.323$ in Cartesian space. The box dimensions in x, y, and z axes were set to 24, 22, and 22 units respectively. For CpACS2, the box center (x, y, z) coordinates were set to 78.559, 78.696, and 73.749, while the box sizes were set

Table 1 Rejected functional groups from the screening library and their SMARTS patterns

Type	SMARTS
Sulfonyl halide	S(=O)(=O)[F,Cl,Br,I]
Nitro group	[N;10,11;$(N(=O)~[O;H0;20,21])]
Aldehyde	[HC]=O
Primary alkyl halide	[Cl,Br,I][CH2]
Epoxide or aziridine	C1[O,N]C1
Sulfonate ester	[[6]S(=O)(=O)O[(O'Donoghue 1995)
Phosphonate ester	[#6]OP(=O)O[#6]
Peroxide	OO
1,2-Dicarbonyl	C(=O)C(=O)
Acid halide	C(=O)[Cl,Br,I]

Please refer to paper by Walters and Murcko (Walters and Murcko 2002)

to 24, 24, and 22 in x, y, and z directions respectively. The scale factor was set to 1. The grid box size was selected to ensure that it fully encompasses the predicted binding site residues. Virtual screening was performed using a custom Python script, using AutoDock Vina (v1.1.2) (Trott and Olson 2010) as the docking engine. The 515 molecules in the screening library were minimized with 1000 steps of steepest descent using the GAFF force field, protonated to pH 7.4, applied appropriate torsions and converted to PDBQT using OpenBabel. To achieve a balance between docking speed and accuracy, the virtual screening was carried out at an exhaustiveness setting of 10. Once the screening concluded, the top six ranked molecules in terms of binding affinity were isolated, and re-docked, a total of 5 times, at an exhaustiveness setting of 25. This was done primarily to observe the variation in the resulting binding poses and affinities reported by Vina during repeated dockings, which would give an idea about the most favorable binding pose, should there be any variation. The docked pose with the highest binding affinity was extracted from the docking results and visualized using PyMOL (v1.8). Based on the predicted binding affinities (in kcal mol^{-1}), the ligand efficiency (LE), Ki (in μM), pKi (calculated from Ki in molars), binding efficiency index (BEI), and surface binding efficiency index (SEI) were calculated using the following equations:

$$\mathrm{Ki} = \exp\left(\frac{\Delta G \times 1000}{1.986 \times 298.15}\right) \times \frac{1}{10^{-6}}$$

$$\mathrm{LE} = \frac{-\Delta G}{\mathrm{HAC}}$$

$$\mathrm{pKi} = -\log_{10}(\mathrm{Ki})$$

$$\mathrm{BEI} = \frac{\mathrm{pKi}}{\mathrm{MW(kDa)}}$$

$$\mathrm{SEI} = \frac{100 \times \mathrm{pKi}}{\mathrm{PSA}}$$

where HAC and PSA represent the heavy atoms count and topological polar surface area respectively. Due to potential drawbacks of the ligand efficiency parameter, the indices BEI and SEI as proposed by the Abbott group (Abad-Zapatero and Metz 2005) were used, aiming to capture the potency per unit molecular weight (in kDa) and unit exposed polar surface area (normalized to 100) respectively. The PSA values (in angstrom) were calculated using OpenBabel. In order to analyze the ligand interactions with the residues in the binding pocket, the "Protein Ligand Interaction Profiler" (PLIP) server (Salentin et al. 2015) was used, which gave a detailed report on the hydrophobic interactions, hydrogen bonds, salt bridges, and pi-stackings. The default thresholds were left unchanged: 4.0 Å for hydrophobic interactions, 4.1 Å for hydrogen bonds (between donor and acceptor), and 5.5 Å for salt bridges. Minimum hbond donor angle was 100°.

Molecular dynamics simulations

The ligand common to both CpACS1 and CpACS2 and ranked within the top three in terms of binding affinities was isolated after the virtual screening run, complexed with the two isoforms, and subjected to a molecular dynamics simulation for 100 ns to explore their binding properties and overall stability. The simulation was carried out using GROMACS (v2018.3) (Abraham et al. 2015). Instead of a cubic simulation box, a rhombic dodecahedron periodic unit cell was used for performance reasons, with the following dimensions: $9.6 \times 9.6 \times 9.6$ (nm) for CpACS1 and $10.65 \times 10.65 \times 10.65$ (nm) for CpACS2. A 0.7-nm gap was maintained between the proteins and the edge of the unit cell. The TIP3P three-point water model was used as the solvent. The unit cells were solvated with 17,086 (CpACS1) and 24,394 (CpACS2) molecules of the same, followed by the addition of 4 and 3 chloride counter ions respectively to neutralize the systems. Prior to production MD, the two complexes were subjected to energy minimization using 50,000 steps of steepest descent. After the energies converged, 300 ps equilibrations using both NVT and NPT ensembles were carried out, with the ligand restrained with a force constant of 1000 kJ mol^{-1} nm^2. The system temperature was brought up to 300 K and the pressure to 1 bar.

The ligand was modeled based on the GAFF force field, and AM1-BCC partial charges were assigned using the Antechamber module available in AmberTools (v18). The ACPYPE script (da Silva and Vranken 2012) was used as a wrapper on top of Antechamber. For production MD, the time step was set to 2 fs, and bonds were constrained using the LINCS algorithm. The AMBER99SB force field was used to model the protein, and temperature and pressure were maintained using the Berendsen thermostat and Parrinello-Rahman barostat respectively. The PME (Particle Mesh Ewald) summation method was used to model non-bonded long-range electrostatics using a Fourier grid spacing of 0.16 nm, while the values in between grid points were calculated using 4th-order cubic interpolation. The van der Waal's force cutoff distance was set at 1.2 nm.

The trajectories for the running simulations were saved at 10 ps intervals and later analyzed using the tools provided within the GROMACS distribution. To assess the bonding characteristics between the CpACS isoforms and the ligand, the PLIP server was used to analyze the interactions at 10 ns intervals over a total period of 100 ns. The root mean squared deviation (RMSD) between the starting structures and their successive snapshots were calculated based on the C-alpha atom positions using the "super" structural alignment algorithm in PyMOL (v1.8) and the extra_fit pymol script.

ADMET predictions

Computational prediction of in vivo absorption/distribution/metabolism/toxicity of a drug candidate allows us to eliminate lead compounds with potentially unfavorable pharmacokinetic/pharmacodynamic properties early in the drug discovery pipeline. To that end, the top three most favorable inhibitors for the CpACS isoforms, as identified by virtual screening, were taken for ADME/tox predictions. A combination of admetSAR (v2.0) (Yang et al. 2019), SwissADME (Daina et al. 2017), and SwissTargetPrediction (Gfeller et al. 2014) servers was used, which employs a multitude of machine learning and QSPR regression models to predict some of the most common ADMET end points for a given chemical structure. In this study, for a given molecule, the physicochemical, pharmacokinetic, structural alerts (if any), and drug/lead likeness values were predicted using SwissADME. Toxicity properties were determined using admetSAR and non-specific binding using SwissTargetPrediction server. The molecules were submitted to the servers as SDFs or SMILES strings after conversion by OpenBabel (O'Boyle et al. 2011).

Results

Protein model validation

The initial models generated by I-TASSER with 14.8% and 7% residues in the outlier region of the CpACS1 and CpACS2 models respectively. These were improved by using repeated refinement steps, followed by manual residue level rebuilding using MODELLER (Eswar et al. 2006). The quality of both the CpACS models after the refinement stages is seen in Table 2. The structure analysis by ProSA server ascertained that the models were within the range of what would be expected from experimental X-ray crystal structures (Supplementary Figs. S14a and S14b). Although some residues in both the models are in the outlier region, care was taken so that no residue in the binding pocket had any improper torsion. Please refer to Supplementary Table S6 for the full SAVESv5 meta-server report on model quality. The binding pocket residues for both the isoforms were identified using a consensus between I-TASSER and 3DLigandSite servers (Supplementary Fig. S6). For CpACS1, the predicted residues were the following: THR256, ALA419, ALA420, PRO421, GLU440, GLY441, PHE442, GLY443, MET444, SER445, ASP525, ILE537, ARG540, and LYS664. For CpACS2, the residues were: HIS301, PHE303, GLY409, ALA410, ALA411, PRO412, GLU431, GLY432, TYR433, GLY434, MET435, THR436, ASP517, ILE529, ARG532, and LYS655. The ACS signature motifs for both the isoforms were observed close to the binding pocket along with the AMP binding motif. For CpACS1, the signature motif was from residue 523 to 542 with a 45% sequence identity, while for CpACS2 the ACS motif was from residue 515 to 534, also with a 45% sequence identity. The AMP binding motifs were found to be from residue 256 to 264 at 88.9% identity for CpACS1 and from residues 251 to 259 at 88.9% identity again for CpACS2.

Predicted inhibitors

For both isoforms, the top six molecules in terms of binding affinity were taken and re-docked repeatedly at a high exhaustiveness setting in Vina; the resulting top three molecules with the highest affinity (Supplementary Fig. S5) were considered as putative functional analogs of Triacsin C. Please refer to Table 3 for the virtual screening results and properties of the same.

The molecules with IDs R134, R439, and CDB2 were found to be most favorable in case of CpACS1, while the molecules R134, R38, and MCD1 were predicted to be potent against CpACS2 (Supplementary Tables S2 and S3). The binding modes for the top three inhibitors for both the isoforms were investigated using the PLIP server (Fig. 2). Overall, the compounds in the screening library displayed a slightly lower ΔG_{bind} for CpACS2 compared to CpACS1 as evidenced from the binding energy distributions (Supplementary Fig. S4). This is in concordance with experimental data (Guo et al. 2013), wherein Triacsin C displayed a lower Ki value for CpACS2 compared to CpACS1. Interestingly, R134 was a common molecule which was observed to be consistently present among the top three most potent inhibitors of the two CpACS isoforms. For CpACS1, ALA420 and LYS664 were common residues which had hydrophobic interactions with R134 and CDB2. PRO421 and PHE442 showed hydrophobic interactions with CDB2 and Triacsin C. CDB2 was also involved in pi-pi interactions with PHE442 and was engaged in hydrogen bonding and hydrophobic interactions with ARG540 and ARG667 respectively. Both ILE665 and ARG667 formed hydrogen bonds with CDB2 and Triacsin C. In case of CpACS2, no pi-stacking interactions with any of the residues were observed. The

Table 2 Model refinement results from SAVES v5 meta server

Isoform	% res* allowed	% res* outlier	ProSA (*z* score)	ERRAT quality
CpACS1	88.6	1.6	−7.76	87.72
CpACS2	91.0	2.9	−9.61	89.43

Table 3 Virtual screening results showing properties of the top three predicted ligands compared to Triacsin C. HAC represents heavy atoms count, BEI and SEI represent binding, and surface binding efficiency indices respectively. LE denotes ligand efficiency

Ligand	Receptor	ΔG^{bind}_{VINA}	MW (Da)	HAC	Ki (uM)	pKi	LE	B.E.I.	S.E.I.
R134	CpACS1	− 7.20 ± 0.00	280.36	20	5.23	5.28	0.36	18.836	6.124
R134	CpACS2	− 8.08 ± 0.16	280.36	20	1.185	5.926	0.40	21.136	6.872
CDB2	CpACS1	− 7.70 ± 0.07	234.29	17	2.251	5.648	0.45	24.106	9.155
MCD1	CpACS2	− 8.14 ± 0.05	298.38	22	1.07	5.97	0.37	20.008	9.67
R439	CpACS1	− 7.30 ± 0.00	273.29	20	4.42	5.35	0.36	19.59	6.17
R38	CpACS2	− 7.86 ± 0.05	278.37	20	1.718	5.765	0.39	20.71	7.896
TriacsinC	CpACS1	− 5.54 ± 0.11	207.27	15	86.43	4.063	0.36	19.60	7.549
TriacsinC	CpACS2	− 6.26 ± 0.05	207.27	15	25.62	4.591	0.41	22.15	8.53

residue MET435 was involved in hydrogen bonding with the ligands R134 and Triacsin C. THR251 was involved with both hydrogen bonding and hydrophobic interactions with the ligand MCD1 and so was the residue THR653. ALA410 was involved in hydrogen bonds and hydrophobic interactions with R134 and MCD1 respectively. LEU156 was involved in hydrophobic interactions with the inhibitors R38 and MCD1, while ARG532 was involved in hydrogen bonding with R134 and salt bridge formation with MCD1. Except MET435, Triacsin C had no common interacting residues with the other ligands in case of CpACS2. For CpACS1, the molecule CDB2 had the most favorable mean binding affinity of -7.7 ± 0.07 kcal mol^{-1}, followed by R439 with -7.3 kcal mol^{-1} and R134 with an affinity of -7.2 kcal mol^{-1}.

For CpACS2, the molecule R134 and MCD1 were among the most favorable inhibitors with a predicted binding affinity of -8.08 ± 0.16 kcal mol^{-1}, and -8.14 ± 0.05 kcal mol^{-1} respectively. The third most favorable was R38 with an affinity

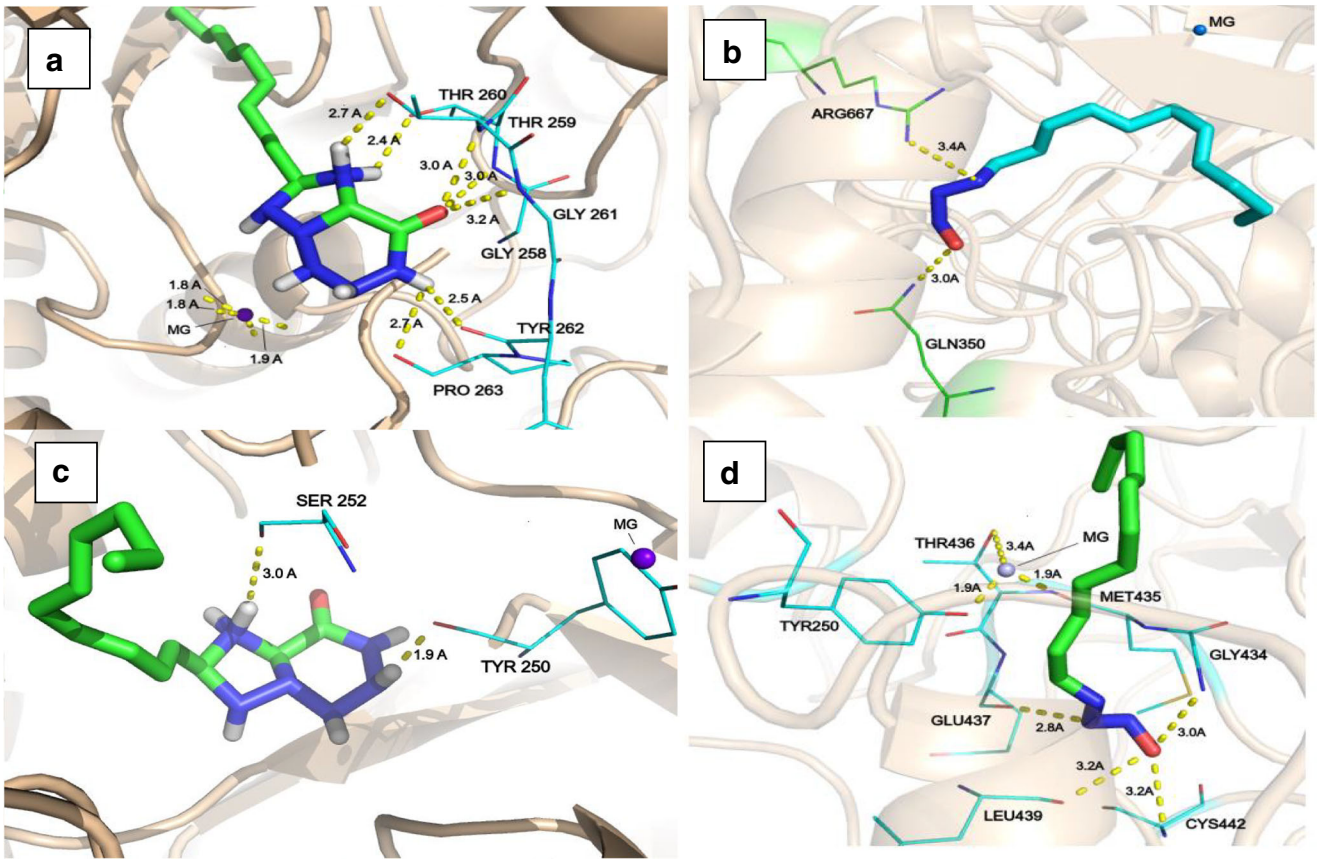

Fig. 2 Docking poses for **a** R134 with CpACS1 ($\Delta G = -7.2$ kcal/mol), **b** Triacsin C with CpACS1 ($\Delta G = -5.7$ kcal/mol), **c** R134 with CpACS2 ($\Delta G = -8.2$ kcal/mol), and **d** Triacsin C with CpACS2 ($\Delta G = -6.3$ kcal/mol). Refer to Supplementary Figs. S7 and S8 for all the poses and Supplementary Figs. S9 and S10 for variations in binding pose per inhibitor

Table 4 Binding interaction details of top three ligands after docking by AutoDock Vina, compared to Triacsin C

Ligand ID	Receptor	Hydrophobic interactions	HBonds	Salt bridges	Pi-stackings
R134	CpACS1	(Phe442x2), Val463, Ile537, Arg540	Gly261, Thr256, Thr259, (Thr260x3), Pro263, Glu446	Glu446	N/A
CDB2	CpACS1	Ala420, Pro421, Phe442, Val463, (Ile537x2), Lys664	(Asp539x2), Arg540, Ile665, Arg667	Arg540, Arg667	Phe442
R439	CpACS1	Thr256, Lys264	Phe255, Ser257 Ile307, Phe308 Ser445, Lys664	N/A	N/A
Triacsin C	CpACS1	(Phe442x2), Pro421 Val463, Ile537	Ile665, Arg667	N/A	N/A
R134	CpACS2	Val155, Leu156, Thr651, (Val657x2)	Tyr250, (Ser252x2), His301	N/A	N/A
R38	CpACS2	Val155, Leu156, (Val657x2)	His249, Ser252	N/A	N/A
MCD1	CpACS2	Leu152, Tyr153, (Leu156x2), Thr251, Ala300, (Phe303x2) Ala410, Thr651 Pro652, Thr653	Thr251, Thr651, Thr653	Arg532	N/A
Triacsin C	CpACS2	Tyr250, Phe303, Lys655	Met435, Leu438, Cys442	N/A	N/A

of -7.86 ± 0.5 kcal mol^{-1}. Table 4 details the interactions of the top three ligands with both the CpACS isoforms after docking. Supplementary Figure S15 displays the relative binding orientation of the top three predicted molecules for both the isoforms.

MD simulation results

Structural stability of the two complexes throughout the simulation time frame was assessed using the RMSD metric (Fig. 3a), which was calculated on the protein backbone. The CpACS isoform 1 appeared to attain stability around 15 ns into the simulation by settling on an RMSD of about 0.45 nm, which it maintained till around 40 ns before abruptly increasing and finally plateauing off at an average of 0.56 nm. Isoform 2 was observed to attain stability slightly after 40 ns, after which it maintained a mean RMSD of 0.5 till the end.

The RMS fluctuations (Fig. 3e) of the residues for both the isoforms were quite high for those in the central regions and at the N and C terminals, however, the binding pocket residues were reasonably stable for both. For isoform 1 (CpACS1), the average RMSF of binding site residues was 0.15 nm, with a maximum fluctuation of 0.26 nm observed for ARG540. Similarly, binding pocket residues of CpACS2 had an average RMSF of 0.11 nm, with a maximum of 0.14 nm observed in GLY432.

The radius of gyration (Fig. 3b) was computed to observe the overall compactness of the proteins throughout the course of the simulation. The value for isoform 1 was observed to increase and settle into a relatively stable average of 2.69 nm after about 40 ns, which suggests changes to overall protein structure after ligand binding. However, isoform 2 displayed no abrupt structural changes and maintained an average radius of gyration of 2.70 nm all through, albeit with minor shifts.

The solvent accessible surface area (SASA) (Fig. 3c) of CpACS1 increased till roughly 15 ns before attaining an average value of 358 nm^2. For CpACS2, the SASA was observed to decrease and maintain an average value of 342 nm^2. All this seems to be indicative of the fact that both the isoforms underwent shifts in their conformation upon ligand binding but ended up reasonably stabilizing within around 40 ns. The average distance between the center of masses of both the proteins and the ligand (R134) was calculated (Fig. 3d), which decreased till around 17 ns for isoform 2 and remained stable with an average distance of 1.09 nm. For isoform 1, the distance between it and the ligand started increasing at about the 25 ns mark and attained a maximum of 1.8 nm at the 57 ns mark, before slowly decreasing to about 1.3 nm during the final stages of the simulation.

Figure 4 shows the superimposed ensemble of 10 ns snapshots of CpACS1 (Fig. 4a) and CpACS2 (Fig. 4c), ligand bound, throughout the 100-ns trajectory. Figure 4b (CpACS1) and Fig. 4d (CpACS2) show the changes in the binding conformation of the ligand and its interacting residues at 10 ps (green) and at 100 ns (magenta). The RMSD between these initial and final protein structures was calculated on the C-alpha atom positions and was observed to be 3.72 nm for CpACS1 and 3.29 nm for CpACS2. For isoform 2, the binding conformation of the ligand varied little throughout the course of the simulation (Fig. 4d), suggesting that the molecular docking performed by AutoDock Vina was reliable. In contrast, the ligand bound to CpACS1 was considerably more mobile (Fig. 4b), albeit within a confined region, which is suggestive of larger conformational changes within the CpACS1 structure to accommodate the bound ligand (Table 5).

The residues of both the isoforms and their types of interaction with the bound ligand are detailed in the Table 6 (CpACS1) and Table 7 (for CpACS2). For magnitude of hydrogen bonding between the ligand (R134) and the two isoforms over 100 ns, please refer to Supplementary Fig. S12.

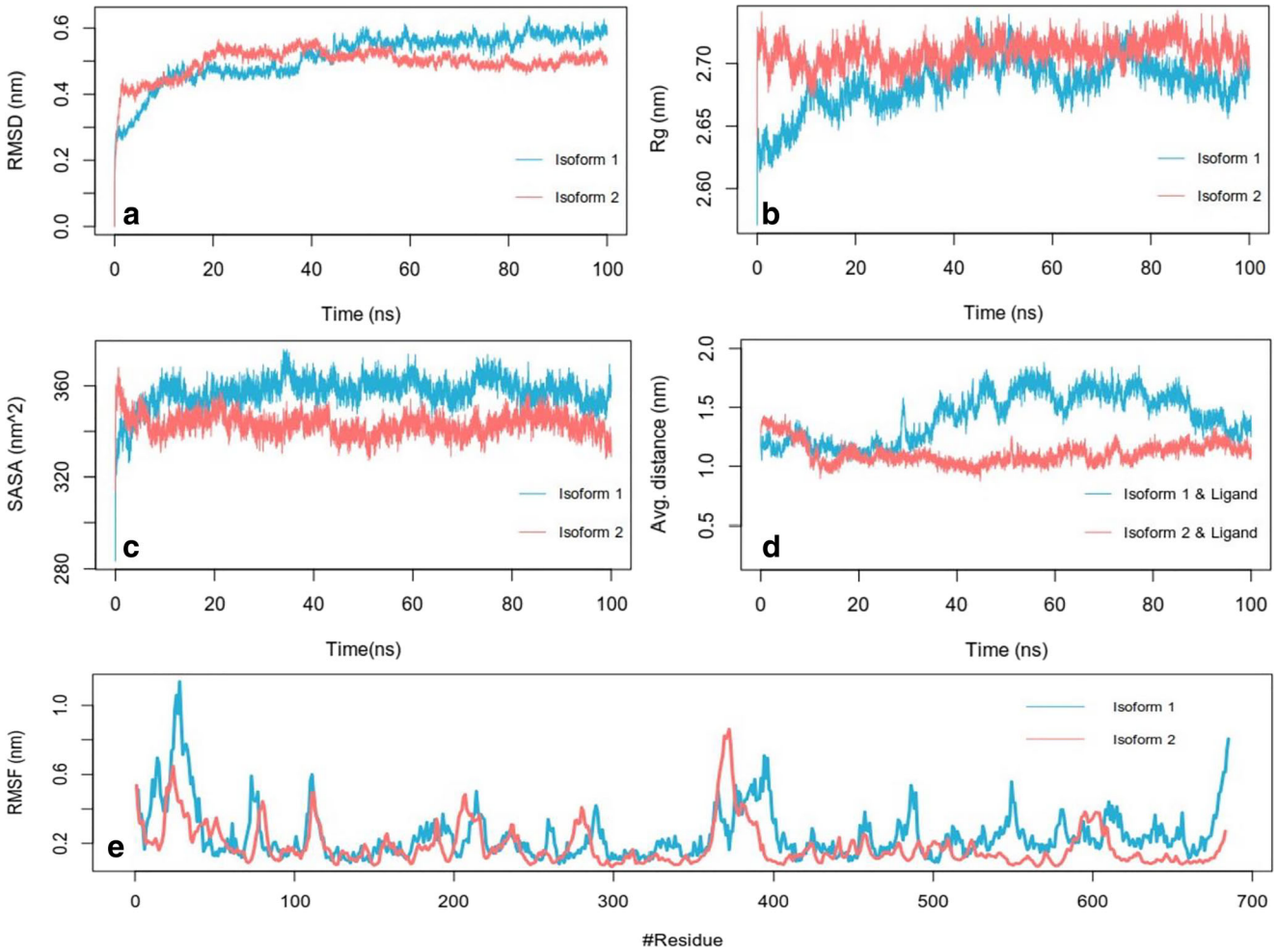

Fig. 3 Plots showing changes in various parameters of the two ligand (R134)-bound CpACS isoforms throughout the course of the simulation. **a** RMSD. **b** Radius of gyration. **c** Solvent accessible surface area. **d** Average distance of ligand from binding pocket. **e** RMSF. All units of length in nanometers

ADMET predictions

The molecule R134 was predicted to have reasonable ADMET properties, with the highest oral bioavailability (74%), no PAINS (pan assay interference) structural alerts, and topological polar surface area (TPSA) of 81.66 Å, suggesting reasonable intestinal absorption but no blood-brain barrier (BBB) permeation, no predicted carcinogenicity or hepatotoxicity, high solubility in water at 1.17 mg ml^{-1}, no inhibition of cytochrome P450 isoforms (1A2,2C19,2C9,2D6,3A4), relatively easy to synthesize with a synthetic accessibility score of 4.61, lead-like, and only one violation of the Pfizer (Lipinski's) rule of five drug-likeness criteria (no. of NH or OH groups > 5). Also, the SwissTargetPrediction server did not identify any small molecule receptors for R134 in the human body. Triacsin C was predicted to be an inhibitor of CYP2C9 and CYP2C19, although it had better synthetic accessibility (3.84), and a clean record in the Pfizer, Amgen, GSK, Bayer, and Pharmacia drug-likeness filters. However, it had four Brenk structural alerts (compared to R134's two), failure in the lead-likeness category (MW < 250, rotatable bonds > 7), and a higher propensity for BBB permeation due to lower TPSA (53.82 Å). Please refer to Supplementary Table S1 for a detailed report. In conjunction to the server-based ADMET predictions, efficiency of the molecules was also predicted using the BEI-SEI plot as is seen in Fig. 5. The entire screening library appeared to be optimized in terms of potency per unit molecular weight, but not quite so in terms of total polar surface area. Overall, the plots are in concordance with results obtained by the Abbott group (Abad-Zapatero and Metz 2005) on a set of marketed oral drugs. The most efficient molecules occupy the BEI-SEI diagonal plane (have a BEI/SEI ratio ≈ 1). For CpACS1, the most efficient molecules were R1, R27, and R217, and for CpACS2, they were R344, R120, and R162 (Supplementary Tables S4 and S5). These molecules lack the desired potency in spite of their high efficiency and are liable for further optimizations. Table 5 gives an overview of the ADMET properties of the top three hits against CpACS1 and 2.

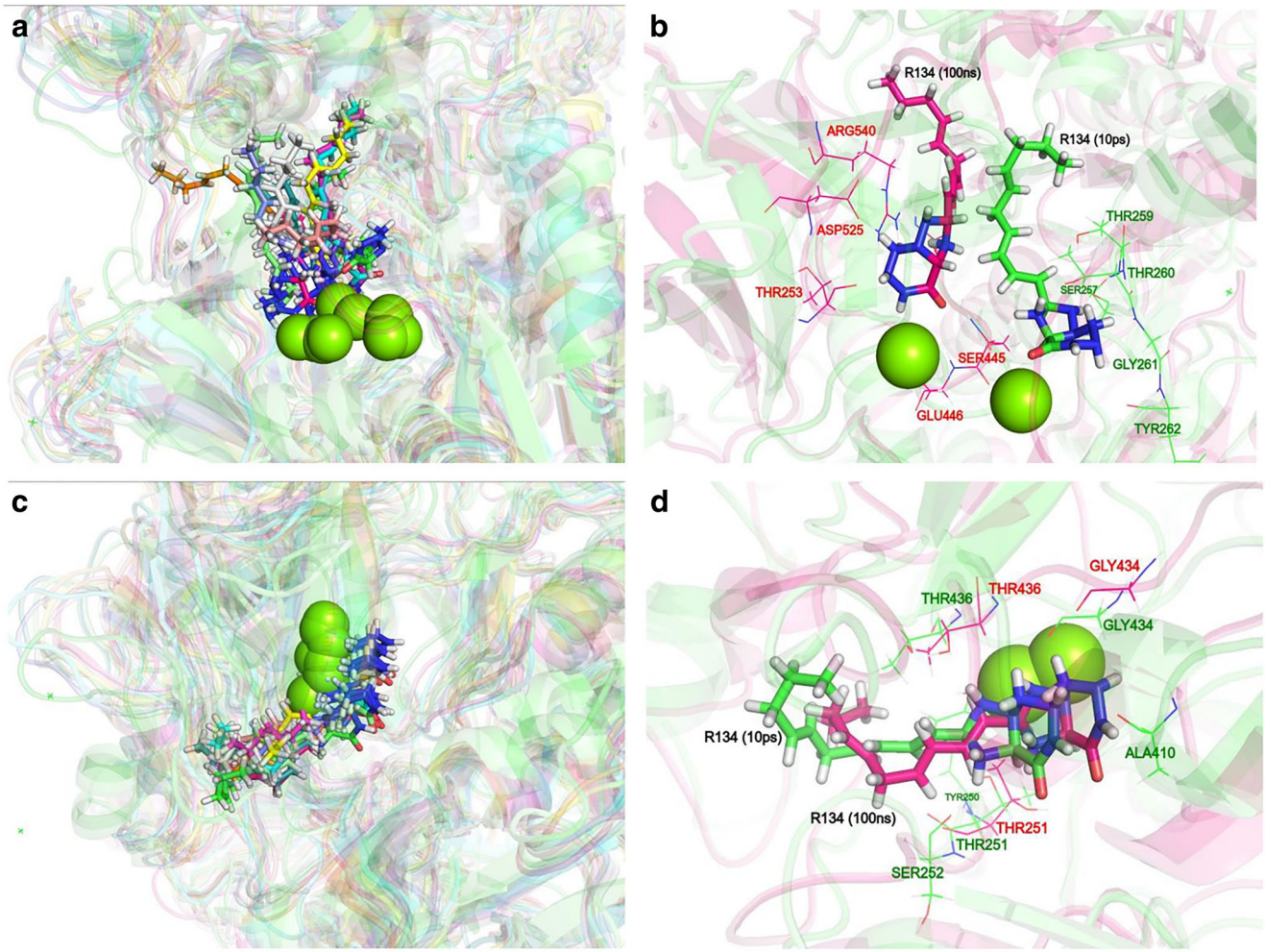

Fig. 4 Superimposed ensemble structures of both the CpACS isoforms over the 100 ns trajectory variation in binding conformations. The green spheres represent Mg^{++} ions. **a** Ten-nanosecond snapshots of ligand bound CpACS1-binding site. **b** Ligand conformations in CpACS1-binding site at 10 ps (green) and 100 ns (red). **c** Ten-nanosecond snapshots of CpACS2-binding site (ligand bound). **d** 10 ps (green) and 100 ns (red) snapshots of CpACS2-binding site showing variation in ligand binding conformations

Table 5 Server-based ADME/tox prediction results of the top three highest affinity ligands

Property	R134	R439	CDB2	R38	MCD1	Triacsin C
MW	279.36	273.29	234.29	278.37	298.38	207.27
HB donor	5	0	2	5	2	1
HB acceptor	5	6	3	5	3	3
Heavy atoms count	20	20	17	20	22	15
TPSA	81.66	86.70	61.69	69.63	61.69	53.82
PAINS	0	0	0	0	0	0
Brenk	2	2	1	2	3	4
Lead likeness	Yes	No	No	Yes	No	No
SA score	4.61	3.81	2.30	4.51	3.33	3.84
Lipinski	0	0	0	0	0	0
XLogP3	1.91	4.26	3.49	1.93	4.78	3.40
Solubility (LogS)	− 3.15	− 5.79	− 4.47	− 3.02	− 5.81	− 4.12
GI absorption	Low	High	High	Low	High	High

Table 6 Residues of CpACS1 interacting with the ligand (R134) over a period of 100 ns. The bond distance and angle thresholds were kept unchanged in PLIP

Time (ns)	Interactions		
	Hydrogen bonds	Hydrophobic interactions	Salt bridges
0	Ser257, Thr259, 4xThr260, Gly261, Tyr262	Ile665	NA
10	2xSer257, Thr523	Val572, Lys664, Ile665	NA
20	2xSer257	2xLys664, Ile665	NA
30	3xAsp525, Arg540	Lys664	Asp525
40	2xAsp525, Lys664	NA	Asp525
50	Asp525, Arg540	NA	Asp525
60	Asp525, Arg540	Thr260, Glu556	Asp525
70	Asp525, Arg540	NA	Asp525
80	Thr523, 2xAsp525, Arg540	Thr259, Glu559	Asp525
90	Thr260, Asp525, 2xArg540	2xPro421	Asp525
100	Gly443, Ser445, Glu446, Thr523, 2xAsp525, 2xArg540	NA	Asp525

Discussion

With nitazoxanide being the only FDA-approved drug currently available for treatment of cryptosporidiosis, we consider the identification of new, high affinity leads compounds against *C. parvum* LC-FACS isoforms to be an important step towards tackling future outbreaks. By basing our work off Triacsin C, an experimentally verified inhibitor of the same, we prepared a set of 514 potential lead compounds using a combination of literature search, database screening, and bioisostere generation by scaffold hopping and pharmacophore alignment. The resulting screening library had reasonable diversity in terms of chemotypes and molecular properties (Tanimoto coefficient ≈ 0.5, see Supplementary Fig. S3 for details). We propose the molecule R134 as a putative adaptive inhibitor of both *C. parvum* LC-FACS isoforms, because of its consistent presence among the top three most potent ligands as determined by virtual screening. A subsequent 100-ns molecular dynamics simulation suggested that R134 was reasonably stable when bound to either of the two CpACS isoforms. ADMET predictions are suggestive of its favorable PK/PD properties, although the relatively low hydrophobicity and Brenk structural alerts might indicate the requirement for further optimizations to the scaffold. CDB2 displayed a higher affinity for CpACS1, but we have chosen to ignore the same because it did not display any corresponding potency against CpACS2. The BEI-SEI plots (Fig. 5) of the screening library showed an almost linear and horizontal spread of molecules with average BEI scores of 18.04 and 19.60 for CpACS1 and CpACS2 respectively. The molecules tended to aggregate above the BEI-SEI diagonal, each cluster having a SEI value below 15. The tendency of the molecules to cluster above the diagonal is indicative of their high potency per unit MW, but slightly under-optimized polar surface area.

Table 7 Residues of CpACS2 interacting with the ligand (R134) over a period of 100 ns. The bond distance and angle thresholds were kept unchanged in PLIP

Time (ns)	Interactions		
	Hydrogen bonds	Hydrophobic interactions	Salt bridges
0	Tyr250, Thr251, Ser252, 2xAla410, Gly434, Thr436	Tyr250, Thr436, Tyr495, Val505	NA
10	Ala410, Gly434, Met435	Thr251, Thr436, Glu501, Ala504, Val505	NA
20	Ala410, Tyr433, Gly434	Thr436, 2xVal505	Glu437
30	Tyr433, Gly434	Tyr250, Thr251, Thr436	Glu437
40	Tyr433, Gly434	Tyr250, Thr251, Glu501, 2xVal505	Glu437
50	Tyr433, Gly434	Thr251, Val505	Glu437
60	Tyr433, Gly434	Thr251, Thr436, Tyr495, 2xVal505	Glu437
70	Tyr433, Gly434	Thr251	Glu437
80	Gly434, 2xLys655	Tyr250, Lys259, Leu438	Glu437
90	Gly434, 2xLys655	Tyr250, Lys259, Tyr495	Glu437
100	2xThr251, Gly434, Thr436	Tyr250, Thr251, 2xLys259, Thr436, Val505	Glu437

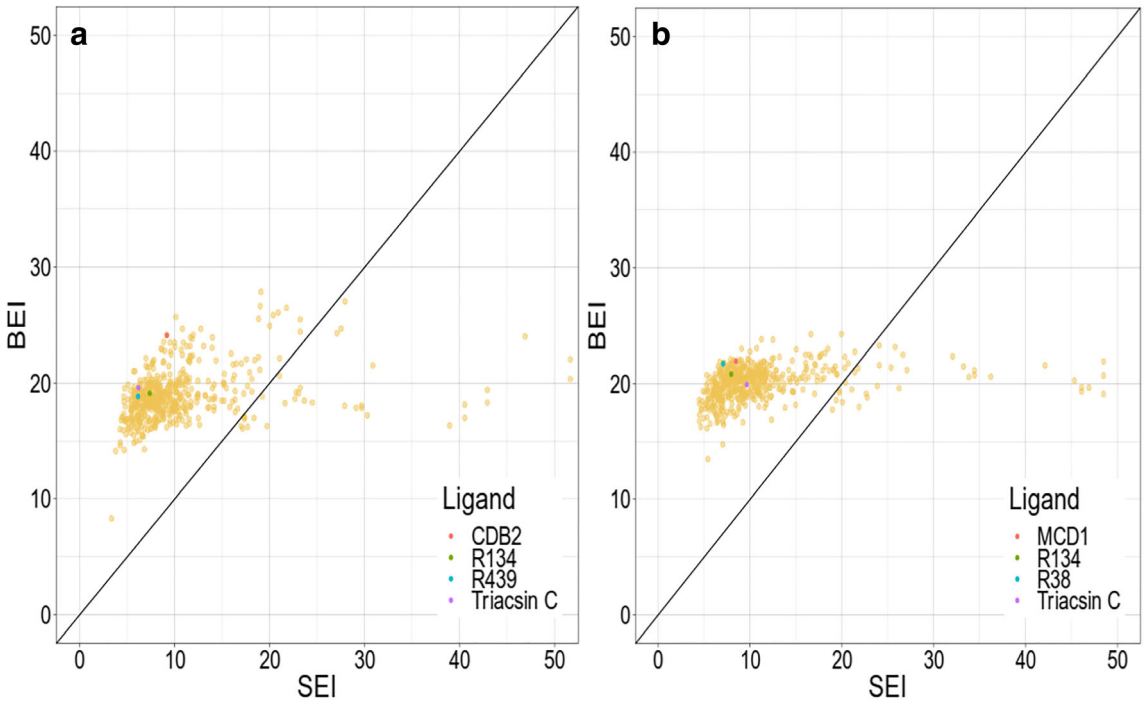

Fig. 5 Plot of BEI (binding efficiency index: potency per unit kDa of molecular weight) vs. SEI (surface binding efficiency index: potency per angstrom of polar surface area) of the screening library molecules (n = 515) against **a** CpACS1 and **b** CpACS2. Highlighted points represent the top three inhibitors (in terms of binding affinity), along with Triacsin C. The most efficient molecules occupy the BEI-SEI diagonal plane (see Supplementary Tables S4 and S5)

This points towards the viability of the screening library as a source of putative lead compounds. Nevertheless, the aim of this work was to generate viable models for CpACS1 and CpACS2, whose crystal structures do not yet exist, and predict putative potent inhibitors against the same. The authors maintain that although the protein models used are of reasonable quality, the results obtained here are predictions and remain hopeful of future in vitro experimental undertakings which might be able to validate the same.

Acknowledgments The authors would like to acknowledge the Bioinformatics Lab Facility of School of Biotechnology, KIIT deemed to be university during the course of the study. The authors would like to thank Dr. Nivedita Jena of KIIT TBI (Technology Business Incubator) for her proposal of the structure of the MCD1 molecule.

References

Abad-Zapatero C, Metz JT (2005) Ligand efficiency indices as guideposts for drug discovery. Drug Discov Today 10(7):464–469

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1:19–25

Azam A, Peerzada MN, Ahmad K (2015) Parasitic diarrheal disease: drug development and targets. Front Microbiol 6:1183

Beisken S, Meinl T, Wiswedel B, de Figueiredo LF, Berthold M, Steinbeck C (2013) KNIME-CDK: workflow-driven cheminformatics. BMC bioinformatics 14(1):257

Borowski H, Thompson RCA, Armstrong T, Clode PL (2010) Morphological characterization of *Cryptosporidium parvum* life-cycle stages in an in vitro model system. Parasitology 137(1):13–26

Chellan P, Sadler PJ, Land KM (2017) Recent developments in drug discovery against the protozoal parasites Cryptosporidium and Toxoplasma. Bioorg Med Chem Lett 27(7):1491–1501

Corso PS, Kramer MH, Blair KA, Addiss DG, Davis JP, Haddix AC (2003) Costs of illness in the 1993 waterborne Cryptosporidium outbreak, Milwaukee, Wisconsin. Emerg Infect Dis 9(4):426–431

da Silva AWS, Vranken WF (2012) ACPYPE-antechamber python parser interface. BMC Res Notes 5(1):367

Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7:42717

Ertl P, Schuffenhauer A (2009) Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. J Cheminformatics 1(1):8

Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY et al (2006) Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics 15(1):5–6

Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V (2014) SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res 42(W1):W32–W38

Guerrant RL (1997) Cryptosporidiosis: an emerging, highly infectious threat. Emerg Infect Dis 3(1):51–57

Guo, F., 2014. Long chain fatty acyl-coenzyme A synthetase as novel drug targets in *Cryptosporidium parvum* (doctoral dissertation)

Guo F, Zhang H, Fritzler JM, Rider SD Jr, Xiang L, McNair NN et al (2013) Amelioration of *Cryptosporidium parvum* infection in vitro and in vivo by targeting parasite fatty acyl-coenzyme A synthetases. J Infect Dis 209(8):1279–1287

Hayes EB, Matte TD, O'Brien TR, McKinley TW, Logsdon GS, Rose JB et al (1989) Large community outbreak of cryptosporidiosis due to

contamination of a filtered public water supply. N Engl J Med 320(21):1372–1376

Heo L, Park H, Seok C (2013) GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic Acids Res 41(W1):W384–W388

Hewitt RG, Yiannoutsos CT, Higgs ES, Carey JT, Geiseler PJ, Soave R et al (2000) Paromomycin: no more effective than placebo for treatment of cryptosporidiosis in patients with advanced human immunodeficiency virus infection. Clin Infect Dis 31(4):1084–1092

Hiroshi T, Kazuaki I, Satoshi O (1987) Inhibition of acyl-CoA synthetase by triacsins. Biochim Biophys Acta (BBA)-Lipids Lipid Metabol 921(3):595–598

Hunter PR, Nichols G (2002) Epidemiology and clinical features of Cryptosporidium infection in immunocompromised patients. Clin Microbiol Rev 15(1):145–154

Kim Y, George D, Prior AM, Prasain K, Hao S, Le DD et al (2012) Novel triacsin C analogs as potential antivirals against rotavirus infections. Eur J Med Chem 50:311–318

Konc J, Janežič D (2010) ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. Bioinformatics 26(9):1160–1168

Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. Proteins 47(3):393–402

Mac Kenzie WR, Hoxie NJ, Proctor ME, Gradus MS, Blair KA, Peterson DE, Kazmierczak JJ, Addiss DG, Fox KR, Rose JB, Davis JP (1994) A massive outbreak in Milwaukee of Cryptosporidium infection transmitted through the public water supply. N Engl J Med 331(3):161–167

O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. J Cheminformatics 3(1):33

O'Donoghue PJ (1995) Cryptosporidium and cryptosporidiosis in man and animals. Int J Parasitol 25(2):139–195

Puleston RL, Mallaghan CM, Modha DE, Hunter PR, Nguyen-Van-Tam JS, Regan CM et al (2014) The first recorded outbreak of cryptosporidiosis due to *Cryptosporidium cuniculus* (formerly rabbit genotype), following a water quality incident. J Water Health 12(1): 41–50

Rotz LD, Khan AS, Lillibridge SR, Ostroff SM, Hughes JM (2002) Public health assessment of potential biological terrorism agents. Emerg Infect Dis 8(2):225–230

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5(4):725–738

Ryan U, Hijjawi N (2015) New developments in Cryptosporidium research. Int J Parasitol 45(6):367–373

Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M (2015) PLIP: fully automated protein–ligand interaction profiler. Nucleic Acids Res 43(W1):W443–W447

Samie A, Al-Qahtani A, El Bakri A, Ehdaie B (2015) Challenges and innovative strategies to interrupt cryptosporidium transmission in resource-limited settings. Curr Trop Med Rep 2(3):161–170

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455–461

Tzipori S, Ward H (2002) Cryptosporidiosis: biology, pathogenesis and disease. Microbes Infect 4(10):1047–1058

Walters WP, Murcko MA (2002) Prediction of 'drug-likeness'. Adv Drug Deliv Rev 54(3):255–271

Wanyiri J, Ward H (2006) Molecular basis of Cryptosporidium–host cell interactions: recent advances and future prospects. Future Med 1: 201–208

Wass MN, Kelley LA, Sternberg MJ (2010) 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Res *38*-(suppl_2):W469–W473

Yang H, Lou C, Sun L, Li J, Cai Y, Wang Z et al (2019) admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. Bioinformatics 35(6):1067–1069

Yoshida K, Okamoto M, Umehara K, Iwami M, Kohsaka M, Aoki H, Imanaka H (1982) Studies on new vasodilators, WS-1228 A and B. J Antibiotics 35(2):151–156