

Identification of potential antagonists of CRF1R for possible treatment of stress and anxiety neuro-disorders using structure-based virtual screening and molecular dynamics simulation

Abdullahi Ibrahim Uba¹ and Chun Wu*

College of Science and Mathematics, Rowan University, Glassboro, NJ, 08028 USA

* To whom correspondence should be addressed: Chun Wu: wuc@rowan.edu:

Abstract

G protein-coupled-receptors (GPCRs) are the largest family of cell surface receptors with tremendous therapeutic potential. They mediate signal transduction activities via G protein-dependent signaling pathways, G protein-independent signaling pathways, and other complicated regulatory processes. The corticotropin-releasing factor receptor type 1 (CRF1R) is a member of class B GPCRs that is predominantly found in the central nervous system, where it plays a key role in stress-related neuro-disorders. To date, no drug targeting this receptor has been approved, partly due to inadequate understanding of the activation mechanism of class B GPCRs.

Previously, using MD simulation, we demonstrated that the CRF1R complexed with a small-molecule antagonist CP-376395 maintains a conformation of its transmembrane domain (TMD). Here, using the most abundant structures derived from those simulations, we carried out a structure-based virtual screening of ZINC15 “Druglike” library containing approximately 17 million compounds. The docking complexes of the CRF1R with the top 30 hits were submitted to MD simulation to examine the stability of ligand binding mode. Furthermore, MM-GBSA binding energy calculations were performed on all the complexes to rank them with improving accuracy. Hit **1** (ZINC000046079839) and hit **20** (ZINC000032907937) span the allosteric site of the CRF1R, persistently forming interactions with transmembrane helices 3 and 6. These interactions are likely to keep the receptor in an inactive state since both transmembrane helices play a critical role in the activation of the receptor.

Keywords: Class B GPCRs, CRF1R, small-molecule antagonist, structure-based virtual screening, MD simulation.

Introduction

G protein-coupled-receptors (GPCRs) are the largest family of cell surface receptors with tremendous therapeutic potential (over 40% of the approved drugs target GPCRs)(Liu & Kokubo, 2017). They are used by cells to convert extracellular signals into intracellular responses, mediating signal transduction activities via G protein-dependent signaling pathways, G protein-independent signaling pathways, and other complicated regulatory processes (Azzi et al., 2003; Rajagopal, Rajagopal, & Lefkowitz, 2010). They share a conserved structural feature comprising seven-transmembrane (7TM) helices connected by alternating intracellular and extracellular loops (ICLs and ECL), an extracellular N-terminus, and an intracellular C-terminus. Five distinct classes of GPCRs are identified based on sequence homology and functional similarity: Class A (rhodopsin), Class B (secretin), Class C (glutamate), Class D (adhesion), and Class E (frizzled) (Pal, Melcher, & Xu, 2012). Class A is the largest GPCR class that is best studied due to the wealth of structural information available, which greatly increases molecular understanding of functions and activation mechanisms. On the other hand, Class B is poorly studied due to the limited structural availability. Members of this class are distinguished by their large cysteine-rich extracellular domain (ECD) which plays an important role during activation. 15 known receptors in this family are implicated in various disease conditions (Kaspar Hollenstein et al., 2014), including stress, anxiety, and related neuro-disorders (Garelja et al., 2020; Harmar, 2001).

The corticotropin-releasing factor receptor type 1 (CRF1R) is a representative member of class B, predominantly found in the central nervous system, where it is involved in the regulation of adrenocorticotrophic hormone (ACTH)—a key modulator in stress response (Kean, Bortolato, Hollenstein, Marshall, & Jazayeri, 2015). Thus, CRF1R is a good drug target for anxiety,

depression, inflammation, and other stress-related neuro-disorders (Teleb, Kuppast, Spyridaki, Liapakis, & Fahmy, 2017).

Molecular dynamics (MD) simulation has been proven to be effective in studying the natural motion of proteins and other biomolecules at the atomic level and time resolution (Hollingsworth & Dror, 2018). Using MD simulation, the activation mechanism of the transmembrane domain (TMD) of CRF1R, which involves large conformational changes of the TM helices, was demonstrated (Seidel, Zarzycka, Zaidi, Katritch, & Coin, 2017; Singh, Ahalawat, & Murarka, 2015). Since the activation mechanism of full-length CRF1R was unclear, we built a full-length CRF1R model using available crystal structures of the N-terminal ECD (PDB ID: 3EHU) and transmembrane domain (TMD) (PDB ID: 4KBY) (**Figure 1**). Using molecular dynamics simulation, we demonstrated that in the presence of a peptide agonist urocortin 1, the CRF1R undergoes large-scale conformational changes involving breakage of networks of inter-helical/regional H-bonds and salt bridges and observed movement of transmembrane helix 6 (TM6). On the other hand, the small molecule antagonist CP-376395-bound CRF1R maintains the initial inactive conformation of the transmembrane domain (TMD)(Uba, Scorese, Dean, Liu, & Wu, 2021).

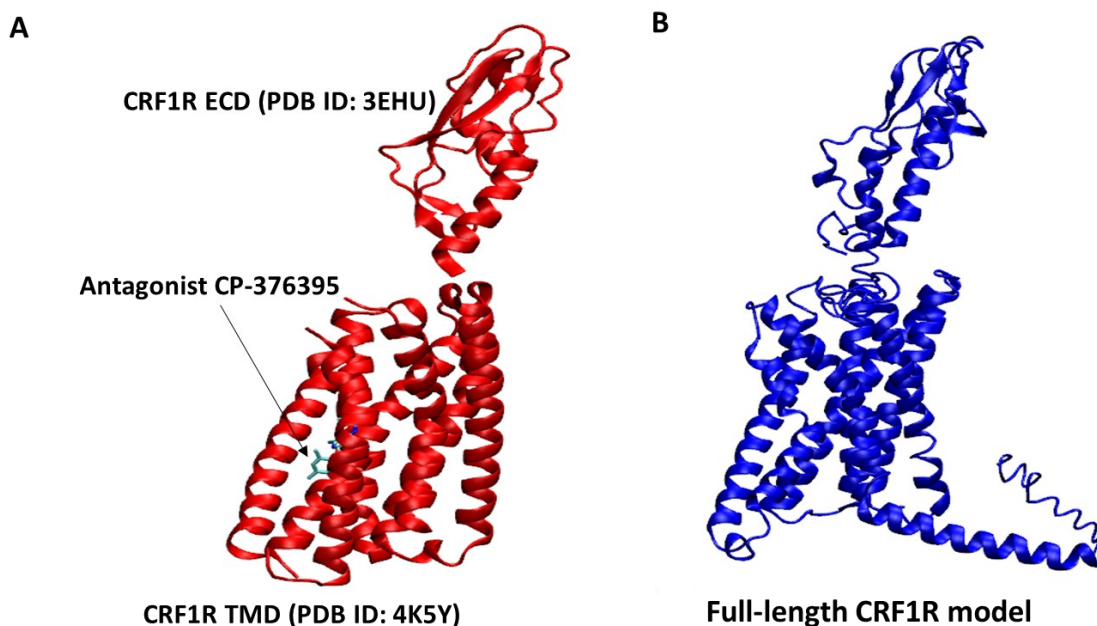


Figure 1. (A) Crystal structure of the N-terminal extracellular domain (ECD) (PDB ID: 4KBY) and transmembrane domain (TMD) of CRF1R. (B) A full-length CRF1R model built by homology modeling.

Here, using multiple structures derived from the afore-mentioned MD simulations, structure-based virtual screening of ZINC “Druglike” library containing 17 million compounds (T. Sterling & J. J. Irwin, 2015) was conducted. The top 30 compounds complexed with the CRF1R were subjected to MD simulation and MM-GBSA binding energy calculation. Hits with MM-GBSA free energy of binding scores higher than that of the antagonist CP-376395 and better physicochemical properties are considered potential CRF1R antagonists. These compounds span the allosteric site of the CRF1R, forming strong interactions with the transmembrane helices. The interactions are likely to keep the receptor in an inactive state.

Methods

Protein preparation

The crystal structures of the N-terminal extracellular domain (ECD) of CRF1R (PDB ID: 3EHU) (Pioszak, Parker, Suino-Powell, & Xu, 2008) and that of the CRF1R transmembrane domain (TMD) bound to small molecule antagonist CP-376395 (PDB ID: 4K5Y) (K. Hollenstein et al., 2013) were retrieved from the protein data bank (<https://www.rcsb.org>) and prepared using Maestro's Protein Preparation Wizard (Sastry, Adzhigirey, Day, Annabhimoju, & Sherman, 2013). During this preparation, the protein was assigned correct bond orders, missing side chain atoms and hydrogen atoms were added, disulfide bonds were created, and water molecules beyond 5 Å were deleted. The charge state of the titratable residues was optimized using PROPKA at a pH of 7. A restrained minimization was done to relax the protein using an OPLS3 force field². These structures were used as a template to build a full-length model of CRF1R. The details of the model building and microsecond MD simulations are available in our previous work (Uba et al., 2021). Three conformations of the CRF1R derived from the MD simulations of antagonist CP-376395 bound CRF1R (Figure S1)(Uba et al., 2021) were used to generate a receptor grid box of 15 Å cube for virtual screening. The ligand-binding site was identified from the crystal structure of the CRF1R TMD bound to the antagonist CP-376395 (PDB ID: 4K5Y) (K. Hollenstein et al., 2013). The grid map encloses key binding site residues (F232^{3.44}, N312^{5.50}, and Y356^{6.53}) that are conserved in class B GPCRs.

Compound Library preparation

A prepared ZINC15 “Druglike” library was downloaded from the ZINC database (Teague Sterling & John J. Irwin, 2015), in which ChemAxon's JChem was used to protonate and prepare biologically relevant tautomers at Physiological pH of 7 (Csizmadia, 2000).

High-throughput virtual screening (HTVS)

Virtual screening was performed using the Glide docking program (Friesner et al., 2004; Halgren et al., 2004). ZINC15 drug-like library containing 17,900,742 ligand entries was screened against the CRF1R allosteric site (Teague, Davis, Leeson, & Oprea, 1999). The druglike library contains compounds that have been filtered based on Lipinski's "rule-of-five" parameters: molecular weight, lipophilicity, and hydrogen bonding potential. **Figure 2** shows the virtual screening workflow comprising multiple scoring functions: high throughput virtual screening (HTVS), standard precision (SP), and extra precision (XP). HTVS serves as the first-pass filter, so only the top 10% of hits were selected. SP further reduces the thoroughness of the final torsional refinement and sampling, retaining 10% of hits. XP performs more extensive sampling, considering ligand-receptor shape complementarity. QikProp module in Schrödinger rapidly screens the resulting hits for physicochemical properties. Canvas task identifies diverse structures based on molecular property descriptors.

Further ligand preparation and induced-fit docking

The ionization/tautomeric state of the ligand was generated at a pH of 7 using an empirical pKa (Epik) prediction module in Maestro (Sastry et al., 2013). To ensure plausible ligand binding mode by incorporating flexibility into the protein side-chain atoms, induced-fit docking (IFD) of these top 30 prepared ligands into the CRF1R allosteric pocket was performed using the Glide IFD module (Friesner et al., 2004). IFD generates a more accurate protein-ligand complex structure, so the ligands bind in a similar mode to that of the cocrystal ligand, the antagonist CP-376395.

Molecular dynamics simulation

Molecular dynamics simulation systems of CRF1R complexes with each of the top 30 hits were constructed. All systems were solvated using the SPC water model and neutralized by the

addition of Na^+ ions at a concentration of 0.15 M NaCl, and modeled using the OPLS3 force field (Harder et al., 2016b) in the Desmond simulation package (Bowers et al., 2006). As applied to another GPCR in our recent MD simulation study (Uba, Aluwala, Liu, & Wu, 2022), the default protocol of relaxation for membrane protein was employed here. This protocol comprises eight steps, viz: (i) Minimization with restraints on solute heavy atoms; (ii) Minimization without any restraints; (iii) Simulation with heating from 0 K to 310 K, with H_2O barrier and gradual restraining; (iv) Simulation in NPT (constant number of particles, constant pressure of 1 bar and constant temperature of 310 K) ensemble with H_2O barrier and with heavy atoms restrained; (v) Simulation in NPT ensembles with equilibration of solvent and lipids; (vi). Simulation in NPT ensemble with protein-heavy atoms annealing from 10.0 kcal/mol to 2.0 kcal/mol; (vii) Simulation in NPT ensemble with $\text{C}\alpha$ atoms restrained at 2 kcal/mol; and (viii). Simulation for 1.5 ns in NPT ensemble with no restraints. Finally, a 200 ns-production run was carried out under the NPT ensemble using the default protocol. During the simulation, the temperature was controlled using the Nosé-Hoover chain coupling scheme (Ikeguchi, 2004) with a coupling constant of 1.0 ps, and pressure was controlled using the Martyna-Tuckerman-Klein chain coupling scheme (Ikeguchi, 2004) with a coupling constant of 2.0 ps. M-SHAKE (Bailey & Lowe, 2009) was used to constrain all bonds connecting hydrogen atoms to enable a 2.0 fs time step in the simulations. The long-range electrostatic interactions under periodic boundary conditions (charge grid spacing of $\sim 1.0 \text{ \AA}$, and direct sum tolerance of 10^{-9}) were treated using the k-space Gaussian split Ewald method (Shan, Klepeis, Eastwood, Dror, & Shaw, 2005). The cutoff distance for short-range non-bonded interactions was set to $10 \text{ \AA}$, and the long-range van der Waals interactions were based on a uniform density approximation. Non-bonded forces were calculated using an r-RESPA integrator (Stuart, Zhou, & Berne, 1996), where the short-range

forces were updated every step and the long-range forces every three steps. Snapshots were collected every 50.0 ps for analysis.

Simulation results analysis

To check the convergence of the MD simulations, C α protein and ligand RMSDs were calculated. To determine the most abundant/dominant structure, trajectory clustering analysis was performed using the Desmond trajectory clustering tool (Bowers et al., 2006). By employing hierarchical clustering with average linkage, backbone RMSD was used as the structural similarity, with merging distance cutoff set at 2.5 Å. The Representative structure (centroid) of each cluster is the structure with the greatest number of neighbors in the structural family.

MM-GBSA binding energy calculations and decomposition

Molecular mechanics generalized Born surface area (MM-GBSA) method with an implicit membrane (a slab-shaped region with a low dielectric constant (~2)) predicts the binding affinity of the ligand with improved prediction accuracy than the docking method (Ghosh, Rapp, & Friesner, 1998; Jianing Li et al., 2011). The MM-GBSA calculations adopted an OPLS3 force field (Harder et al., 2016a), a VSGB 2.0 solvation model (J. Li et al., 2011), and the default Prime protocol. The default procedure consists of three steps: computation of energies of receptor alone, ligand alone, and finally receptor-ligand complex. The interaction terms are Coulombic, H-bond, GB solvation, van der Waals, pi-pi packing, self-contact, and lipophilic interactions. The total binding free energy equation is given as:

$$\Delta E_{\text{(bind)}} = E_{\text{complex}} - (E_{\text{ligand}} + E_{\text{receptor}})$$

The interaction terms were then merged into three components, Electrostatics, E_{vdW} , and $E_{\text{lipophilic}}$, where

$$E_{\text{electrostatics}} = H_{\text{bond}} + E_{\text{coulomb}} + E_{\text{GB_solvation}}$$

$$E_{vdW} = E_{vdW} + E_{\pi-\pi} + E_{\text{self-contact}} \text{ and } E_{\text{lipophilic}}$$

The MM-GBSA scoring function lacks the solute conformational entropy which results in higher negative values when compared to the actual values. Nevertheless, it has proven to be extremely useful for ranking different drugs targeting receptors with comparable binding entropy values (Harder et al., 2016b). Therefore, MM-GBSA binding energy was calculated for the top 30 diverse hits against the CRF1R on the snapshots collected during the last 10 ns of the simulation.

Further ADMET prediction

Prediction of ADMET properties for the top 30 compounds was performed on the SwissADME web server (<http://www.swissadme.ch/>) developed by the Swiss Institute of Bioinformatics to enable computational estimation of physiochemical descriptors and pharmacokinetic properties, and drug-like small molecule inhibitors. The SMILE code for each compound was uploaded to the webserver and their ADMET properties were computed. These include gastrointestinal (GI) absorption, blood-brain barrier permeability, Lipinski's "rule-of-5" parameters, liver metabolic (CYP450) enzymes inhibition potential, and PAINS filtering.

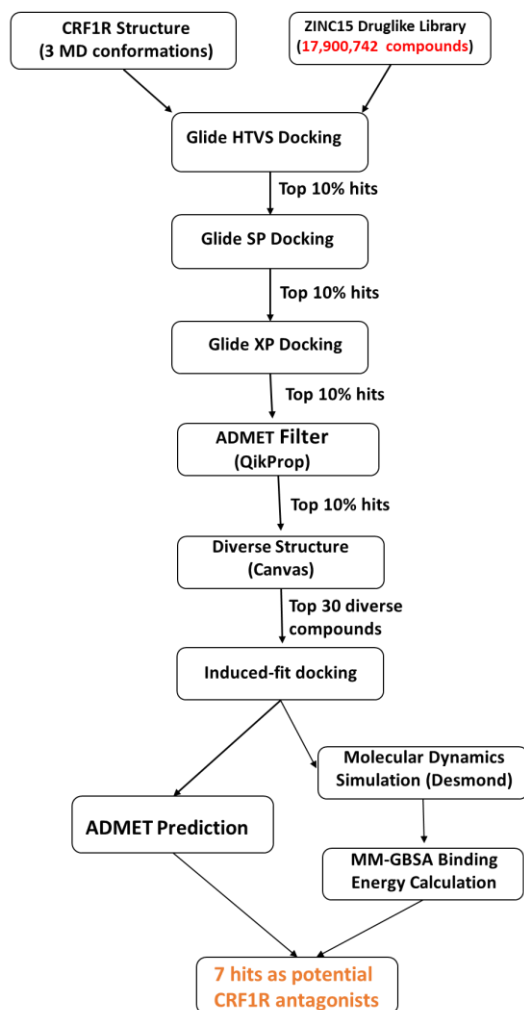


Figure 2. Virtual screening workflow for the identification of CRF1R antagonists.

Results

The initial hits show strong binding potential and good ADMET profiles

A virtual screening workflow comprising various filters is depicted in **Figure 2** above. The final compounds were selected based on an MM-GBSA binding energy score higher than that of the antagonist CP-376395 (**Table 1**) and good ADMET properties prioritizing blood-brain barrier permeability, high GI absorption, and 0 PAINS alert (**Table 2**). PAINS are chemicals that non-specifically target several biological targets due to their disruptive functional groups (Baell

& Walters, 2014). A complete list of the ADMET properties for the individual hits is given in Figure S2. These compounds are diverse and bear heterocyclic groups as surface recognition moiety (**Figure 3**).

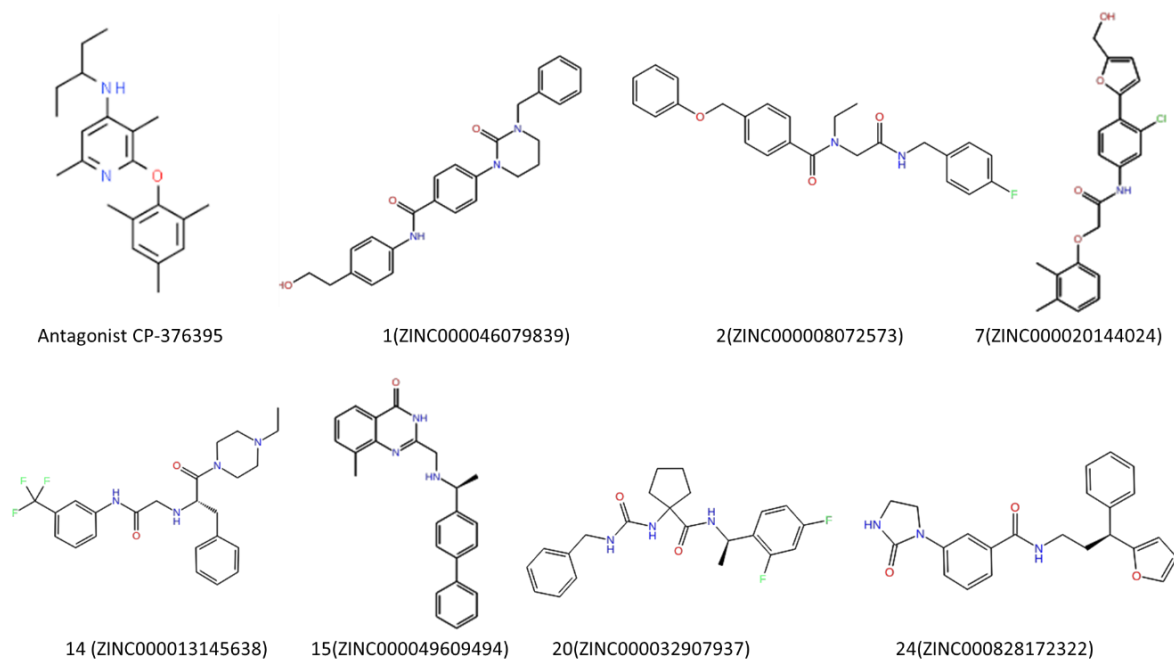


Figure 3. Structures of the top 7 hits identified by virtual screening of ZINC druglike library as potential antagonists of CRF1R.

Table 1. Binding energy and RMSD of the top 30 hits compared to the reference compound, antagonist CP-376395. Hits 1 to 10 against conformation 1(green); 11 to 20 against conformation 2 (blue); 21 to 30 against conformation 3 (orange).

#	Compound	Docking Score (kcal/mol)	VDW (kcal/mol)	ELE (kcal/mol)	Hydrophobic (kcal/mol)	MM-GBSA (kcal/mol)	Receptor RMSD(Å) ^a	Ligand RMSD(Å) ^a
Ref.	Antagonist CP-376395	NA	-66.4±1.0	3.9±1.2	-30.5±1.8	-93.0±4.4	8.6±1.4	0.4±0.1
1	ZINC000046079839	-12.2	-68.1±3.6	-14.5±9.4	-52.1±3.3	-134.8±12.5	4.8±0.3	1.63±0.1
2	ZINC000008072573	-12.1	-61.3±3.9	-6.1±4.9	-51.8±3.7	-119.2±8.4	6.4±0.1	2.2±0.1
3	ZINC000001154395	-11.9	-60.8±5.2	-10.0±11.6	-54.8±4.3	-125.6±13.9	3.9±0.1	2.6±0.1
4	ZINC000065062688	-11.8	-61.4±3.9	-10.2±4.0	-52.6±3.2	-124.3±8.5	4.8±0.3	1.7±0.2
5	ZINC000409176962	-11.8	-58.8±2.9	-19.0±2.9	-41.4±2.8	-119.2±6.1	4.6±0.4	1.3±0.3
6	ZINC000214746700	-11.8	-52.3±1.2	-3.3±2.4	-39.0±2.9	-94.6±4.3	5.3±0.3	0.3±0.2
7	ZINC000020144024	-11.8	-57.4±5.6	5.3±6.1	5.3±4.9	-109.6±13.7	6.0±0.2	1.8±0.3
8	ZINC000009730882	-11.7	-51.2±4.1	-28.1±12.2	-42.0±2.5	-121.5±12.7	5.2±0.2	2.8±0.1
9	ZINC000077119068	-11.6	-61.5±3.9	-21.1±4.9	-56.1±2.8	-138.7±6.9	7.3±0.2	1.7±0.0
10	ZINC000001926343	-11.6	-46.6±4.8	-13.6±6.9	-35.1±3.2	-95.3±9.7	4.6±0.1	2.1±0.1
11	ZINC000224809098	-13.7	-62.4±4.4	-17.4±4.6	-58.8±4.1	-138.5±10.7	5.4±0.1	1.9±0.1
12	ZINC000224898313	-13.2	-55.4±4.8	-9.9±3.4	-59.3±5.6	-124.6±11.1	8.4±0.4	2.7±0.2
13	ZINC000224117283	-13.1	-66.3±2.7	-7.9±7.1	-53.3±3.7	-127.6±7.9	4.4±0.2	2.1±0.4
14	ZINC000013145638	-13.1	-48.6±4.0	-13.6±3.4	-44.5±3.8	-106.7±7.7	6.1±0.2	2.6±0.1
15	ZINC000049609494	-12.9	-52.1±4.2	-12.3±4.9	-44.7±3.3	-109.1±8.3	6.3±0.2	1.5±0.2
16	ZINC000224728551	-12.9	-55.3±3.9	-10.6±5.1	-51.8±3.9	-117.7±6.6	7.1±0.2	2.1±0.1
17	ZINC000224631654	-12.9	-61.3±2.9	-0.3±8.4	-55.6±3.4	-117.1±10.1	4.6±0.2	2.5±0.1
18	ZINC000224761269	-12.8	-63.6±4.0	-2.3±5.2	-58.6±3.4	-124.4±9.3	7.8±0.1	1.7±0.1
19	ZINC000224669931	-12.8	-56.6±3.6	-5.4±4.9	-55.3±3.2	-117.3±7.5	6.2±0.3	2.4±0.0
20	ZINC000032907937	-12.8	-61.9±3.8	-12.4±3.1	-56.5±3.6	-130.8±3.8	4.9±0.3	1.9±0.2
21	ZINC000004521247	-12.2	-50.5±4.1	-2.6±11.7	-37.0±3.2	-90.1±10.9	4.4±0.5	1.7±0.1
22	ZINC000000959167	-12.0	-50.4±3.4	-2.4±9.3	-37.1±4.3	-89.0±8.4	4.3±0.4	1.7±0.1
23	ZINC000000057555	-12.0	-51.6±4.0	-18.2±7.8	-41.8±2.9	-111.6±10.5	4.3±0.4	1.7±0.1
24	ZINC000828172322	-11.9	-50.0±6.6	-21.7±14.5	-48.7±3.7	-120.4±12.1	6.3±0.1	2.2±0.2
25	ZINC000031167739	-11.5	-54.8±3.5	-17.5±5.4	-51.4±3.4	-123.7±7.3	6.4±0.3	1.6±1.1
26	ZINC000012374475	-11.5	-54.8±3.5	-17.5±5.4	-51.3±3.4	-123.7±7.2	6.3±2.1	2.1±0.1
27	ZINC000020761418	-11.5	-56.0±4.3	-17.5±11.0	-42.2±4.1	-115.7±13.4	8.8±0.9	2.1±0.1
28	ZINC000031159228	-11.5	-58.2±3.5	-14.2±6.5	-55.3±3.9	-127.7±9.5	5.0±0.1	2.5±0.1
29	ZINC000143132475	-11.3	-57.4±3.7	-16.4±7.6	-59.8±3.4	-133.7±8.5	5.5±0.2	0.5±0.1
30	ZINC000067673743	-11.3	-62.0±7.3	-23.6±3.2	-41.8±3.7	-127.4±12.2	6.4±0.2	1.4±0.1

^a Based on the snapshots from the last 10 ns simulation.

Compounds selected based on MM-GBSA scores <-93.0 Kcal/mol are represented in **bold** font.

Table 2. Druglike and ADMET properties of the top 30 compounds predicted using the SwissADME server. Compounds shown in **bold** have higher MM-GBSA energy scores than that of the cocrystal ligand Antagonist CP-376395 (PDB ID: 4KBY). Compounds highlighted in green show good ADMET properties, including high gastrointestinal (GI) absorption and blood-brain barrier permeability.

S/N	Compound	GI absorption	BBB permeant	CYP1A2 inhibitor	CYP2C19 inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	CYP3A4 inhibitor	Lipinski's "Rule of 5"	PAINS
Ref.	Antagonist CP-376395	High	Yes	No	Yes	No	Yes	No	Yes; 1 violation: MLOGP>4.15	0 alert
1	ZINC000046079839	High	Yes	Yes	Yes	Yes	Yes	Yes	0 violation	0 alert
2	ZINC000008072573	High	Yes	No	Yes	Yes	Yes	Yes	0 violation	0 alert
3	ZINC000001154395	Low	No	Yes	Yes	Yes	No	Yes	0 violation	0 alert
4	ZINC000065062688	High	No	No	Yes	Yes	No	Yes	0 violation	0 alert
5	ZINC000409176962	High	No	Yes	Yes	Yes	No	Yes	0 violation	0 alert
6	ZINC000214746700	Low	No	No	Yes	Yes	No	No	0 violation	1 alert: ene_rhod
7	ZINC000020144024	High	Yes	Yes	Yes	Yes	Yes	Yes	0 violation	0 alert
8	ZINC000009730882	High	No	No	No	Yes	No	No	0 violation	0 alert
9	ZINC000077119068	High	No	Yes	Yes	Yes	Yes	Yes	0 violation	0 alert
10	ZINC000001926343	Low	No	No	No	Yes	No	No	0 violation	1 alert: sulfonamide
11	ZINC000224809098	High	No	Yes	Yes	Yes	Yes	Yes	0 violation	0 alert
12	ZINC000224898313	High	No	No	Yes	Yes	Yes	Yes	0 violation	0 alert
13	ZINC000224117283	High	No	Yes	Yes	Yes	Yes	Yes	0 violation	0 alert
14	ZINC000013145638	High	Yes	No	No	No	Yes	Yes	0 violation	0 alert
15	ZINC000049609494	High	Yes	No	Yes	Yes	Yes	Yes	0 violation	0 alert
16	ZINC000224728551	High	No	Yes	Yes	Yes	Yes	Yes	0 violation	0 alert
17	ZINC000224631654	High	No	Yes	Yes	Yes	Yes	Yes	0 violation	0 alert
18	ZINC000224761269	High	No	Yes	Yes	Yes	Yes	Yes	0 violation	0 alert
19	ZINC000224669931	High	No	Yes	Yes	Yes	Yes	Yes	0 violation	0 alert
20	ZINC000032907937	High	Yes	No	Yes	No	Yes	Yes	0 violation	0 alert
21	ZINC000004521247	High	No	No	No	No	Yes	No	0 violation	0 alert
22	ZINC000000959167	High	No	No	Yes	Yes	No	No	0 violation	0 alert
23	ZINC000000057555	High	No	No	No	No	Yes	No	0 violation	0 alert
24	ZINC000828172322	High	Yes	No	Yes	Yes	Yes	Yes	0 violation	0 alert
25	ZINC000031167739	High	No	No	No	Yes	No	No	0 violation	1 alert: catechol A.
26	ZINC000012374475	High	No	No	No	No	Yes	No	0 violation	0 alert
27	ZINC000020761418	Low	No	No	Yes	Yes	No	No	0 violation	0 alert
28	ZINC000031159228	High	No	No	No	Yes	No	No	0 violation	1 alert: catechol A.
29	ZINC000143132475	High	No	No	No	No	Yes	No	0 violation	0 alert
30	ZINC000067673743	High	No	No	Yes	No	No	No	0 violation	0 alert

Note: hits 1-10 against conformation 1; 11-20 against conformation 2; 21-30 against conformation 3.

The top hits show stable binding mode

To examine the stability of ligand binding mode, both protein and ligand RMSDs were computed over time. The protein-C α RMSDs increase due to the high mobility of the N-terminal

ECD and C-terminal regions, and later converge through the simulation. On the other hand, decreased ligand RMSD trends are observed (**Figure 4**), suggesting stability of binding mode. While the RMSF values of the N-terminal ECD, C-terminal, and loop regions increase, the TMD show decreased fluctuation (Figure S3). Also, the overall secondary structure is maintained with some loss of helicity (Figure S4).

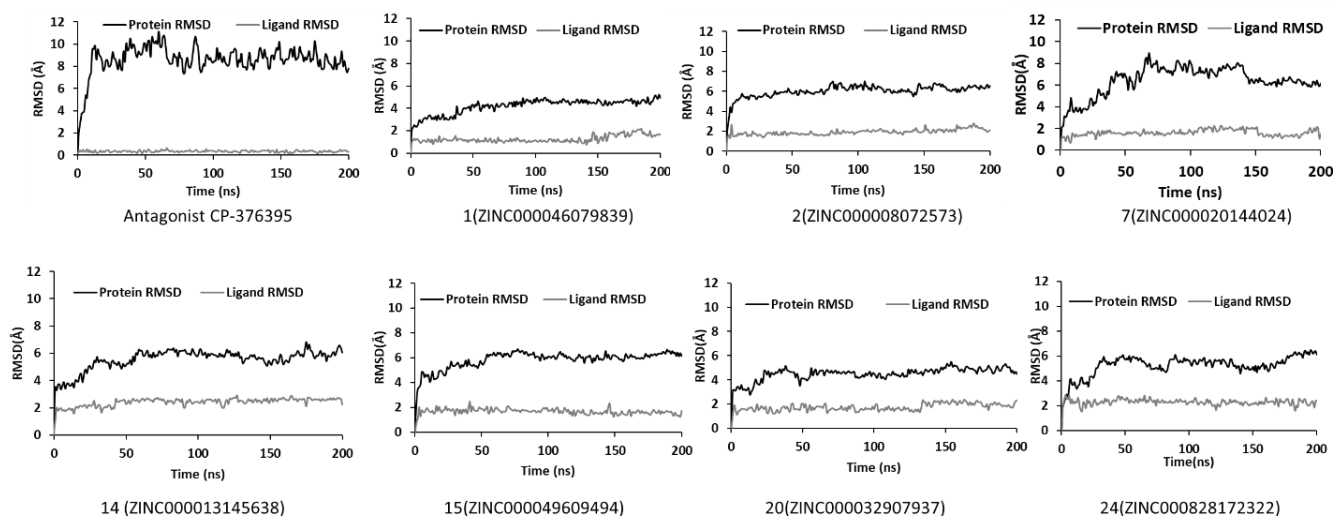


Figure 4. Protein C α and ligand RMSDs of the CRF1R complexes with the top 7 hits identified by structure-based virtual screening.

Plausible protein-ligand interactions

A summary of protein-ligand interactions is presented in given in **Figure 5**. Like the antagonist CP-376395, the key interactions formed by the top compounds are mostly hydrophobic, with a couple of polar and charged contacts. The antagonist CP-376395 forms persistent interactions with N312^{5.50}, and Y356^{6.53} (The Ballesteros–Weinstein numbering scheme(Ballesteros & Weinstein, 1995)). These residues enclose the allosteric antagonist binding site of the CRF1R. Other less persistent interactions are formed with L209, F232^{3.44}, and M235^{3.47}. In the case of the CRF1R MD conformation 1, hit **1** forms similar interaction as the

antagonist CP-376395, with additional polar interactions with R194 and N231. Similarly, hits **2** and **7** engage both Y356^{6,53} and F232^{3,44} in hydrophobic interactions. For conformation 2, hit **14** bearing a benzyl group along its length, does not share any interactions with antagonist CP-376395; it rather forms a persistent hydrophobic interaction with L352^{6,49}, and a charged interaction with E238^{3,50} via a water molecule. Hit **15** forms a persistent interaction with L352^{6,49} like hit **14**, with only other non-persistent interactions with F232^{3,44} like formed by the antagonist CP-376395. Despite having a different structure from the antagonist CP-376395, hit **20** forms similar interacting residues enclosing the CRF1R allosteric site. In the case of conformation 3, hit **24** forms persistent polar interactions with residues along TM3, TM6, and TM7.

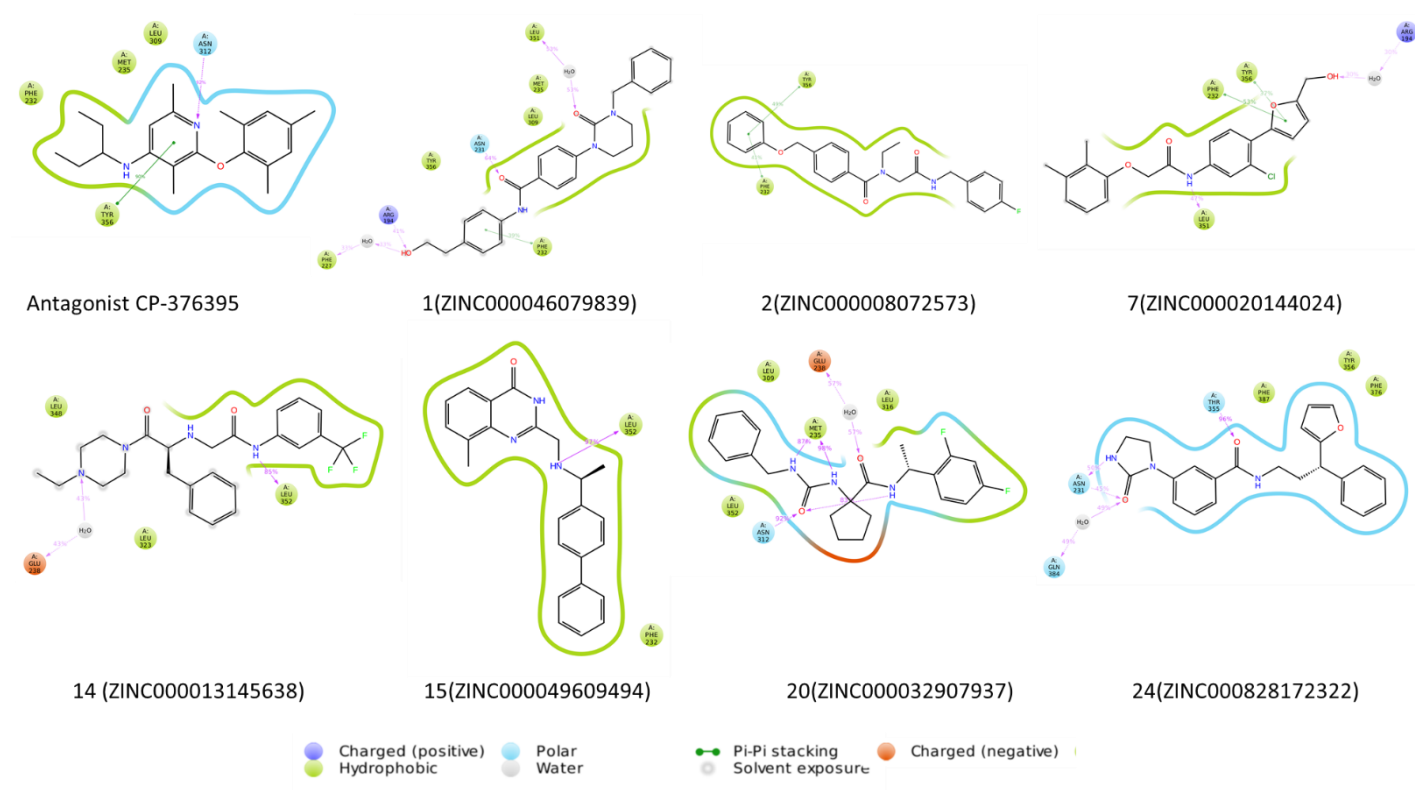


Figure 5. Protein-ligand interaction in the CRF1R complexes with the antagonist CP-376395 and the top 7 hits identified by structure-based virtual screening.

Ligand conformational adjustment

Superimposition of the induced-fit docking pose with the MD simulation poses for the most abundant structure of each ligand reveals ligand conformational adjustment (**Figure 6**). Like the antagonist CP-376395, hits **1** and **2** do not show significant movement, but rather a slight shift of aromatic recognition group. Hit **7** shifts horizontally towards TM3 and TM6 but remains bound deeper into the allosteric pocket. The conformations of hits **14**, and **20** become more extended, while hit **15** slightly moves to the edge of the pocket. However, hit **24**—the only hit against conformation 3, moves away from the pocket, resulting in a completely different set of persistent interactions. Hence, of all the 7 hits, **24** should be considered with caution.

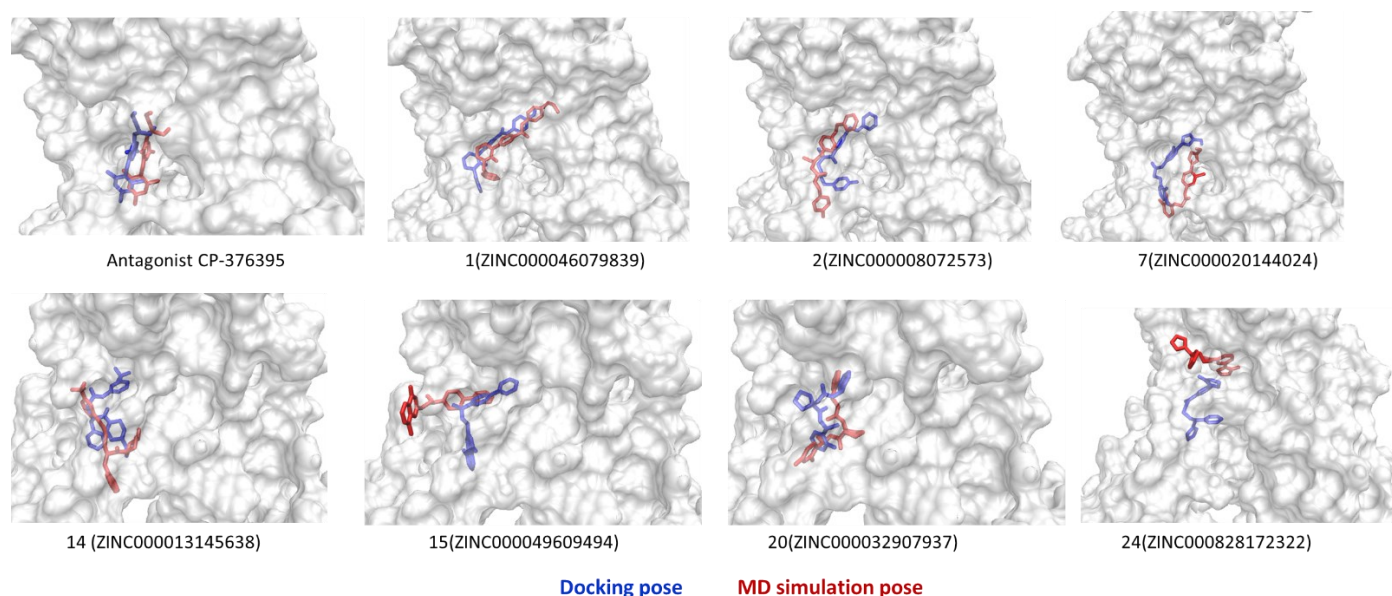


Figure 6. Superimposition of the docking pose with MD simulation pose of the antagonist CP-376395 and the top 7 hits identified by structure-based virtual screening of ZINC druglike library against CRF1R.

Discussion

Despite the implication of CRF1R in stress-related neuro-disorders, no drug targeting the receptor has been approved yet. To identify novel small non-peptide antagonists of CRF1R, we carried out an elaborate virtual screening campaign that comprises various filters from a

combination of multiple computational approaches. Using the 3 most abundant conformations derived from our previous MD simulation of the antagonist-bound CRF1R(Uba et al., 2021), we identified a total of 30 compounds as the initial hits from virtual screening. These were subjected to MD simulations coupled with MM-GBSA binding energy calculation and ADMET prediction. Seven compounds having MM-GBSA binding energy lower than that of the antagonist CP-376395, high gastrointestinal absorption, and blood-brain barrier permeability were selected as the final hits. The selection of the MM-GBSA free energy of binding scores higher than that of the antagonist CP-376395 increases the reliability of the identified hits since the MM-BGSA method is more accurate in predicting ligand binding affinity than molecular docking method (Hou, Wang, Li, & Wang, 2011a), even though the former ignores entropic contribution to the overall free energy of binding (Sun et al., 2018). Another method for calculating ligand binding affinity is molecular mechanics-Poisson Boltzmann surface area (MM-PBSA)(Fogolari, Brigo, & Molinari, 2003). However, the MM-GBSA method has been demonstrated to perform better than MM-PBSA in predicting both correct binding poses and binding free energy for the examined protein-ligand systems (Hou, Wang, Li, & Wang, 2011b). Also, the further ADMET filtering performed was meant to increase the chance of obtaining compounds with better physicochemical properties (van de Waterbeemd & Gifford, 2003).

Structural stability was measured using protein and ligand RMDs. Increased protein RMSD in the antagonist CP-376395 is due to the large movement of the N-terminal ECD and C-terminal region (Uba et al., 2021). On the other hand, all the 7 hits show lower protein RMSDs which converge over time. Hits **1**, **2**, and **7** stabilize at <2.0 Å against conformation 1, while hits **14**, **15**, **20** stabilize at about 2.0 Å against conformation 2, and so does hit **24** against conformation 3. Thus, hits **1**, **2**, and **7** demonstrate similar dynamic behavior with the antagonist CP-376395.

These decreased ligand RMSD trends suggest binding mode stability (Liu & Kokubo, 2017; Uba, Weako, Keskin, Gürsoy, & Yelekçi, 2019). Consistent with the protein RMSD data, the N-terminal ECD, C-terminal region, and loops show higher fluctuation compared to the TMD region within which the antagonist ligands sit. Increased movement of these regions of CRF1R has been reported previously (Seidel et al., 2017).

The histogram of the protein-ligand interaction fractions reveals more interactions with higher fractions formed by the hits than those formed by the antagonist CP-376395 (**Figure 7**). A timeline representation of the interactions and contacts (H-bonds, hydrophobic, ionic, water bridges) showing the total number of specific contacts the protein makes with the ligand over the course of the trajectory is given in Figure S5. In particular, hits **1** and **20** demonstrate similar interaction patterns to the antagonist CP-376395. In addition, being larger than the reference compounds, they span the whole CRF1R allosteric pocket, forming persistent polar and hydrophobic interactions with mostly TM3 and TM6 via amide groups. Both TM3 and TM6 play a critical role in class B activation (Garelja et al., 2020). Thus, these interactions are likely to keep the receptor in an inactive state.

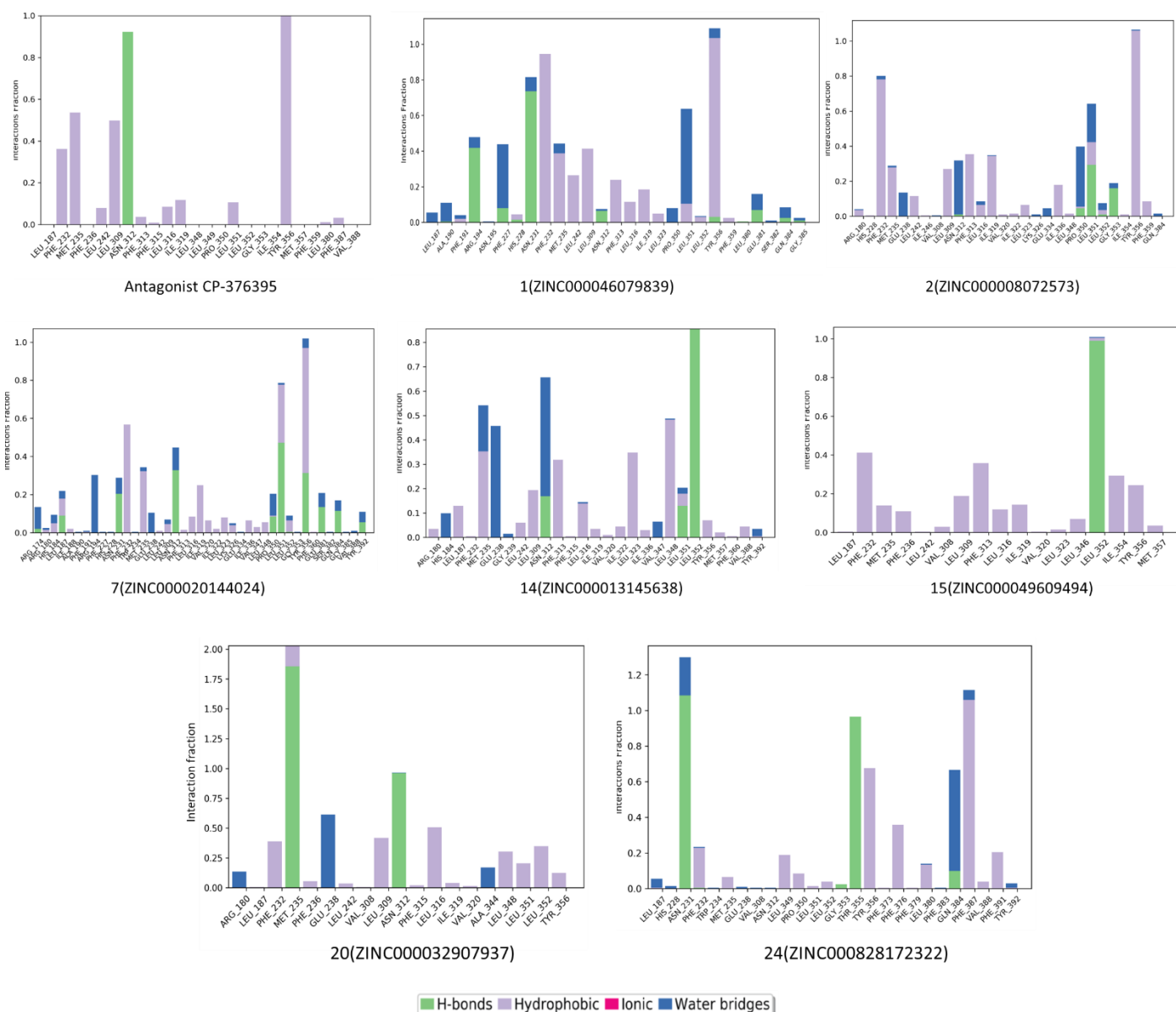


Figure 7. Histogram of protein-ligand interaction showing interaction fractions in the complexes of CRF1R with the top 7 hits identified from structure-based virtual screening of ZINC15 druglike library.

Conclusion

Despite the therapeutic potential of the CRF1R, no drugs targeting the receptor have been approved to date, partly due to inadequate understanding of the activation mechanism.

Previously, we showed that the conformation of CRF1R TMD is maintained when bound to the

antagonist CP-376395. Here, we used the conformations derived from those simulations to identify potential antagonists of CRF1R by structure-based virtual screening and MD simulations. Of the top 7 hits identified, **1** (ZINC000046079839) and hit **20** (ZINC000032907937) demonstrate plausible binding to residues at the TM3 and TM7. Since both TMs play a critical role in the activation of the receptor, hits **1** and **20** are likely to maintain the inactive conformation of the receptor. Therefore, these are proposed as potential CRF1R antagonists for the possible treatment of stress and anxiety-related neuro-disorders.

Acknowledgments

We acknowledge the New Jersey Health Foundation (PC 76-21), the US National Science Foundation under Grants NSF ACI-1429467/RUI-1904797, and XSEDE MCB 170088, and the Anton2 machine time at the Pittsburgh Supercomputing Center (MCB170090P) which is generously made available by D. E. Shaw Research.

Reference

- Azzi, M., Charest, P. G., Angers, S., Rousseau, G., Kohout, T., Bouvier, M., & Piñeyro, G. (2003). β -Arrestin-mediated activation of MAPK by inverse agonists reveals distinct active conformations for G protein-coupled receptors. *Proceedings of the National Academy of Sciences*, 100(20), 11406-11411. doi:10.1073/pnas.1936664100
- Baell, J., & Walters, M. A. (2014). Chemistry: Chemical con artists foil drug discovery. *Nature*, 513(7519), 481-483. doi:10.1038/513481a
- Bailey, A. G., & Lowe, C. P. (2009). MILCH SHAKE: An Efficient Method for Constraint Dynamics Applied to Alkanes. *J. Comput. Chem.*, 30(15), 2485-2493. doi:10.1002/jcc.21237
- Ballesteros, J. A., & Weinstein, H. (1995). [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In *Receptor Molecular Biology* (pp. 366-428).
- Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., . . . Shaw, D. E. (2006). *Scalable algorithms for molecular dynamics simulations on commodity clusters*. Paper presented at the Proceedings of the 2006 ACM/IEEE conference on Supercomputing, Tampa, Florida. http://delivery.acm.org/10.1145/1190000/1188544/a84-bowers.pdf?ip=150.250.5.19&id=1188544&acc=ACTIVE%20SERVICE&key=7777116298C9657D%2E414C84AC7BC68407%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=943096243&CFTOKEN=88922320&_acm_ =1496278456_ afaa38972ff106d0072c2b84f83af9e6

- Csizmadia, F. (2000). JChem: Java Applets and Modules Supporting Chemical Database Handling from Web Browsers. *Journal of Chemical Information and Computer Sciences*, 40(2), 323-324. doi:10.1021/ci9902696
- Fogolari, F., Brigo, A., & Molinari, H. (2003). Protocol for MM/PBSA molecular dynamics simulations of proteins. *Biophys J*, 85(1), 159-166. doi:10.1016/S0006-3495(03)74462-2
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., . . . Shenkin, P. S. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, 47(7), 1739-1749. doi:10.1021/jm0306430
- Garelja, M. L., Au, M., Brimble, M. A., Gingell, J. J., Hendrikse, E. R., Lovell, A., . . . Hay, D. L. (2020). Molecular Mechanisms of Class B GPCR Activation: Insights from Adrenomedullin Receptors. *ACS Pharmacology & Translational Science*, 3(2), 246-262. doi:10.1021/acspsci.9b00083
- Ghosh, A., Rapp, C. S., & Friesner, R. A. (1998). Generalized Born Model Based on a Surface Integral Formulation. *J. Phys. Chem. B*, 102(52), 10983-10990. doi:10.1021/jp982533o
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem*, 47(7), 1750-1759. doi:10.1021/jm030644s
- Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., . . . Friesner, R. A. (2016a). OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J Chem Theory Comput*, 12(1), 281-296. doi:10.1021/acs.jctc.5b00864
- Harder, E., Damm, W., Maple, J., Wu, C. J., Reboul, M., Xiang, J. Y., . . . Friesner, R. A. (2016b). OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.*, 12(1), 281-296. doi:10.1021/acs.jctc.5b00864
- Harmar, A. J. (2001). Family-B G-protein-coupled receptors. *Genome Biol*, 2(12), REVIEWS3013. doi:10.1186/gb-2001-2-12-reviews3013
- Hollenstein, K., de Graaf, C., Bortolato, A., Wang, M.-W., Marshall, F. H., & Stevens, R. C. (2014). Insights into the structure of class B GPCRs. *Trends in Pharmacological Sciences*, 35(1), 12-22. doi:10.1016/j.tips.2013.11.001
- Hollenstein, K., Kean, J., Bortolato, A., Cheng, R. K., Dore, A. S., Jazayeri, A., . . . Marshall, F. H. (2013). Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature*, 499(7459), 438-443. doi:10.1038/nature12357
- Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. *Neuron*, 99(6), 1129-1143. doi:10.1016/j.neuron.2018.08.011
- Hou, T., Wang, J., Li, Y., & Wang, W. (2011a). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model*, 51(1), 69-82. doi:10.1021/ci100275a
- Hou, T., Wang, J., Li, Y., & Wang, W. (2011b). Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J Comput Chem*, 32(5), 866-877. doi:10.1002/jcc.21666
- Ikeguchi, M. (2004). Partial Rigid-Body Dynamics in NPT, NPAT and NPgammaT Ensembles for Proteins and Membranes. *J. Comput. Chem.*, 25(4), 529-541. doi:10.1002/jcc.10402
- Kean, J., Bortolato, A., Hollenstein, K., Marshall, F. H., & Jazayeri, A. (2015). Conformational thermostabilisation of corticotropin releasing factor receptor 1. *Scientific Reports*, 5, 11. doi:10.1038/srep11954
- Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., & Friesner, R. A. (2011). The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins*, 79(10), 2794-2812. doi:10.1002/prot.23106

- Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., & Friesner, R. A. (2011). The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins: Structure, Function, and Bioinformatics*, 79(10), 2794-2812. doi:10.1002/prot.23106
- Liu, K., & Kokubo, H. (2017). Exploring the Stability of Ligand Binding Modes to Proteins by Molecular Dynamics Simulations: A Cross-docking Study. *Journal of Chemical Information and Modeling*, 57(10), 2514-2522. doi:10.1021/acs.jcim.7b00412
- Pal, K., Melcher, K., & Xu, H. E. (2012). Structure and mechanism for recognition of peptide hormones by Class B G-protein-coupled receptors. *Acta Pharmacologica Sinica*, 33(3), 300-311. doi:10.1038/aps.2011.170
- Pioszak, A. A., Parker, N. R., Suino-Powell, K., & Xu, H. E. (2008). Molecular Recognition of Corticotropin-releasing Factor by Its G-protein-coupled Receptor CRFR1. *Journal of Biological Chemistry*, 283(47), 32900-32912. doi:10.1074/jbc.M805749200
- Rajagopal, S., Rajagopal, K., & Lefkowitz, R. J. (2010). Teaching old receptors new tricks: biasing seven-transmembrane receptors. *Nature Reviews Drug Discovery*, 9(5), 373-386. doi:10.1038/nrd3024
- Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des*, 27(3), 221-234. doi:10.1007/s10822-013-9644-8
- Seidel, L., Zarzycka, B., Zaidi, S. A., Katritch, V., & Coin, I. (2017). Structural insight into the activation of a class B G-protein-coupled receptor by peptide hormones in live human cells. *eLife*, 6. doi:10.7554/eLife.27711
- Shan, Y., Klepeis, J. L., Eastwood, M. P., Dror, R. O., & Shaw, D. E. (2005). Gaussian Split Ewald: A Fast Ewald Mesh Method for Molecular Simulation. *J. Chem. Phys.*, 122(5), 54101-54101. doi:10.1063/1.1839571
- Singh, R., Ahlawat, N., & Murarka, R. K. (2015). Activation of corticotropin-releasing factor 1 receptor: insights from molecular dynamics simulations. *J Phys Chem B*, 119(7), 2806-2817. doi:10.1021/jp509814n
- Sterling, T., & Irwin, J. J. (2015). ZINC 15--Ligand Discovery for Everyone. *J Chem Inf Model*, 55(11), 2324-2337. doi:10.1021/acs.jcim.5b00559
- Sterling, T., & Irwin, J. J. (2015). ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, 55(11), 2324-2337. doi:10.1021/acs.jcim.5b00559
- Stuart, S. J., Zhou, R., & Berne, B. J. (1996). Molecular Dynamics with Multiple Time Scales: The Selection of Efficient Reference System Propagators. *J. Chem. Phys.*, 105(4), 1426-1436. doi:10.1063/1.472005
- Sun, H., Duan, L., Chen, F., Liu, H., Wang, Z., Pan, P., . . . Hou, T. (2018). Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. *Phys Chem Chem Phys*, 20(21), 14450-14460. doi:10.1039/c7cp07623a
- Teague, S. J., Davis, A. M., Leeson, P. D., & Oprea, T. (1999). The Design of Leadlike Combinatorial Libraries. *Angew Chem Int Ed Engl*, 38(24), 3743-3748. doi:10.1002/(sici)1521-3773(19991216)38:24<3743::Aid-anie3743>3.0.Co;2-u
- Teleb, M., Kuppast, B., Spyridaki, K., Liapakis, G., & Fahmy, H. (2017). Synthesis of 2-imino and 2-hydrazono thiazolo[4,5-d]pyrimidines as corticotropin releasing factor (CRF) antagonists. *European Journal of Medicinal Chemistry*, 138(Supplement C), 900-908. doi:<https://doi.org/10.1016/j.ejmech.2017.07.016>
- Uba, A. I., Aluwala, H., Liu, H., & Wu, C. (2022). Elucidation of Partial Activation of Cannabinoid Receptor Type 2 and Identification of Potential Partial Agonists: Molecular Dynamics Simulation and Structure-Based Virtual Screening. *Computational Biology and Chemistry*. doi:10.1016/j.compbiolchem.2022.107723

- Uba, A. I., Scorese, N., Dean, E., Liu, H., & Wu, C. (2021). Activation Mechanism of Corticotrophin Releasing Factor Receptor Type 1 Elucidated Using Molecular Dynamics Simulations. *ACS Chemical Neuroscience*, 12(9), 1674-1687. doi:10.1021/acscchemneuro.1c00126
- Uba, A. I., Weako, J., Keskin, Ö., Gürsoy, A., & Yelekçi, K. (2019). Examining the stability of binding modes of the co-crystallized inhibitors of human HDAC8 by molecular dynamics simulation. *Journal of Biomolecular Structure and Dynamics*, 1-10. doi:10.1080/07391102.2019.1615989
- van de Waterbeemd, H., & Gifford, E. (2003). ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov*, 2(3), 192-204. doi:10.1038/nrd1032