Geometric Deep learning Prioritization and validation of cannabis phytochemicals as anti-HCV non- nucleoside direct-acting inhibitors

Ssemuyiga Charles
charles.ssemuyiga@kiu.ac.ug

Kampala International University

Mulumba Pius Edgar
Kampala International University

Research Article

Keywords: Graph Neural network, Deep learning, Hepatitis C virus, Cannabis sativa, and Molecular simulation dynamics

Posted Date: February 21st, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3961716/v1

Additional Declarations: No competing interests reported.

Abstract

Introduction:

The rate of acute hepatitis C increased by 7% between 2020 and 2021, after the number of cases doubled between 2014 and 2020. With the current adoption of pan-genotypic HCV therapy, there is a need for improved availability and accessibility of this therapy. However, double and triple DAA-resistant variants have been identified in genotypes 1 and 5 with resistance-associated amino acid substitutions (RAASs) in NS3/4A, NS5A, and NS5B [1]. The role of this research was to screen for novel potential NS5B inhibitors from the cannabis compound database (CBD) using Deep Learning.

Methods

Virtual screening of the CBD compounds was performed using a trained Graph Neural Network (GNN) deep learning model. Re-docking and conventional docking were used to validate the results for these ligands since some had rotatable bonds > 10. 31 of the top 67 hits from virtual screening and docking were selected after ADMET screening. To verify their candidacy, six random hits were obtained for FEP/MD and Molecular Simulation Dynamics.

Results

The top 200 compounds from the deep learning virtual screening were selected, and the virtual screening results were validated by re-docking and conventional docking. The ADMET profiles were optimal for 31 hits. Simulated complexes indicate that these hits are likely inhibitors with suitable binding affinities and FEP energies. Phytil Diphosphate and glucaric acid were suggested as possible ligands against NS5B.

Introduction

The primary cause of chronic hepatitis, cirrhosis, and hepatocellular carcinoma (HCC) is the hepatitis C virus (HCV). Approximately 3% of the global population have a chronic HCV infection, and for 10–30 years, 30% of carriers are predicted to experience significant liver-related illnesses, such as HCC [2]. The number of cases of acute hepatitis C that have been reported has increased by 129% since 2014 [3].

HCV is an enveloped + ssRNA hepacivirus of the Flaviviridae family with 7 known genotypes in addition to unknown genotypes having several subtypes that mostly affect the liver. Its 9.6Kb genome is flanked by 5' and 3' UTRs with a single ORF that codes for a polyprotein with 3,000 amino acids[4] and is post-translationally processed by cellular and viral proteases to yield 11 viral proteins composed of seven nonstructural proteins (NS) which are the two proteins necessary for the formation of the virion (p7 and NS2), as well as five proteins that make up the cytoplasmic viral replication complex (NS3, NS4A, NS4B, NS5A, and NS5B), and three structural proteins (one nucleocapsid protein and two envelope proteins). Also encoded by the core region is an alternative open reading frame protein (ARFP) or F protein, whose function is still unknown [5].

NS5B is a 66 kDa heart-shaped catalytic component of the HCV replication complex, and owing to selectivity and non-toxicity, it makes a good therapeutic target due to the lack of mammalian counterparts. Like other polymerases, it has the palm, thumb, and finger domains surrounding the enzyme active site in the palm domain [6]. Many anti-HCV Direct Acting Antivirals (DAAs) that are either nucleotide inhibitors (NIs) or non-nucleotide inhibitors (NNIs) targeting NS5B are in development [7,8]. At least 4 allosteric inhibitory sites have been reported (thumb site I or II, palm site I or II) [9] with their mechanisms of action detailed elsewhere [10–15]. However, these allosteric inhibitors have several drawbacks including low potency in enzymatic assays [16], lack of cellular potency during an HCV sub-genomic replicon assay [17], high lipophilic character, and low genetic barrier to resistance [17]. It is possible for distinct forms of HCV with various amino acid changes that confer treatment resistance to coexist inside the same host in the context of quasispecies. While some alterations are not linked to drug resistance, others can result in a phenotypic decrease in susceptibility to one or more antiviral drugs. [18] documented several substitutions with no discernible antiviral resistance to Dasabuvir and Sofosbuvir. Multi-drug-resistant RAASs variants of NS3/4A in GP1 and GP5 along with DAA-specific NS3/4A, NS5A, and NS5B were identified pan-genotypically [1]. Resistance to daclatasvir and sofosbuvir was conferred by L2003M and S2702T of NS5B, respectively. D1194A NS3/4A was triple DAA (simeprevir, faldaprevir, and asunaprevir) resistant. Double-drug resistant variants included R1181K (faldaprevir and asunaprevir), A1182V and Q1106K/R (faldaprevir and simeprevir), T1080S (faldaprevir and telaprevir), while single drug-resistant variants were V1062L (telaprevir), D1194E/T (simeprevir), D1194G (asunaprevir), S1148A/G (simeprevir), and Q1106L (Boceprevir) of NS3/4A were determined [1]. Many NS5B RAASs in genotypes 1, 3, 4, and 5 have been reported [1]. Other mutations in NS5B that were related to DAAs resistance include E237G, S282R, L320F, V321A, and V321I [19]. S282T induces high sofosbuvir-resistance, Q309R is a ribavirin-associated resistance, E237G was identified in the successfully amplified non-responder sample [20].

With a multiverse of biochemical compounds (cannabinoids and non-cannabinoids (Phenolics, Terpenes, and Alkaloids)), *cannabis spp* has been reported by several communities to have medicinal activities against several disease-causing pathogens[21] including HCV. HCV Population had lower rates of diabetes and obesity when they consumed cannabis, however, whether this translates into lower mortality should be investigated [22]. It has been observed to address key challenges like nausea, and depression in HCV Patients. Some individuals receiving therapy for HCV may benefit

virologically and symptomatically from modest cannabis use by helping them stick to the difficult medication regimen [23,24]. Faster decay of HIV RNA among cannabis users who were HCV co-infected and reduced risk of steatosis was observed [25,26]. The advantages of using cannabis for treating HCV from a biological and clinical standpoint, as well as the efficacy of this treatment, should be investigated in bigger study populations. Since there is no known vaccine or treatment for this virus, there is an urgent need for effective ways to manage and eventually eradicate this illness. This work aimed to investigate, using in-silico methodologies, if cannabis chemicals could block NS5B from HCV.

Materials and Methods

The protein structure was downloaded from pdb (PDB ID: 3FQL) while cannabis small molecules were downloaded from the Cannabis Compound Database (CBD).

Geometric deep learning Virtual screening.

PDBbind database [27] protein-ligand complexes were used as inputs for training the model. Complexes that were also part of CASF-2016 [28] and those that failed pre-processing were excluded. The remaining complexes were randomly divided into training (14,000) and test (2,367) sets. The detailed representation of ligand molecules is as in [29] and the protein targets were processed using a pipeline described by [30]. Two separate residual graph convolutional neural networks with the same architecture, one for the ligand and the other for the target, were used to extract features to build the model, the extracted features were concatenated and used to build a mixed-density network (MND).

Initially, a linear layer is used to project the node and edge features to a 128-dimensional embedding. Each node and edge was updated using a series of three GNNs depending on the nodes that were next to them and the kinds of edges that connected them. The GNN initially updates each edge in the graph by using a multi-layer perceptron (MLP) on the concatenation of the edge features and the features of the two connecting nodes and the updated edge features are used to update the node feature [29]. The modified edge and node features can be utilized as input for a subsequent convolution round because they contain information about not only the core atom but also its surrounding neighbors. 3 convolutions were used. The node and edge features were then processed by the remaining GNN blocks [29].

After being pairwise concatenated, the node features recovered by the GNNs and residual GNNs for the ligand and target are fed into an MND [31]. Concatenated target and ligand node information are combined to construct a hidden representation by the MND using an MLP. The MND's outputs are computed using the hidden representation. Furthermore, by connecting neighboring nodes, the retrieved ligand node properties were utilized to forecast the type of bond and atom, aiding in the learning of molecular structures and speeding up the training. Every MLP that is utilized consists of a linear layer, batch normalization, and an ELU activation function. A dropout rate of 0.1 was employed [29].

Training

The Adam optimizer was utilized to update the model weights at a learning rate of 0.002. The loss function was minimized during model training. 16 protein-ligand complexes were used as the batch size for 150 epochs of training the model. A potential specific to a given target-ligand complex was defined using the loss function. This potential was then used to score the target-ligand complex's three-dimensional structure by summing over all possible pairs, calculating the negative log-likelihood for each target-ligand node pair, and calculating the distances separating each target node from each ligand node in that particular conformation. The likelihood of finding the target-ligand combination in that particular conformation increases with a decreasing value [28,29].

Benchmarking

The CASF-2016 benchmark [28], which includes 285 carefully chosen protein-ligand complexes, was used to evaluate this. The preprocessing of the structures from this benchmark was identical to that of the training set. The power of screening and scoring was assessed [29].

Prediction of binding conformations

The relative location of the ligand in Euclidean space, the dihedral angles of all rotatable bonds in the molecule, and the Euler angles were used to represent the ligand conformation as a vector. Differential evolution [32] was used to determine which ligand conformation would interact with the target binding site most likely following the model, i.e., minimize the potential learned by the model for that particular complex. Using a population size of 150, the global optimization was performed up to 500 iterations with a recombination constant of 0.8 and a mutation constant randomly modified from 0.5 to 1.0 in every generation. Euler angles and the dihedrals of rotatable bonds were limited to values between $-\pi$ and π without seeding [29]. The detailed geometric learning protocol can be accessed in the original publication [29]. The deep neural network learns the parameters of a mixture model that is employed as a probability density function. This probability density function is used to determine the most likely distance separating a ligand atom from a specific point in the molecular surface of the binding site. The potential is determined as the combination of the negative log-likelihood of all pairwise combinations of ligand atoms and points in the molecular surface. The optimal conformation is the one that minimizes the potential, that is, the ligand conformation in which every atom is separated from the target surface by the most likely distance.

Conventional docking

Structure preparation

The protein structure was cleaned and preprocessed by assigning bond orders using the CCd database, adding Hydrogens, creating zero bond orders, and creating disulfide bonds as well as generation of het states using Epik [33] (PH 7 ± 2 units), water, and Ligands were removed and H-bond assignment was done using PROPKA in Maestro 12.8 in the protein preparation wizard. The ligands were prepared by Ligprep [34]. Briefly, the ligands were imported into the maestro workspace, and only those with a maximum of 300 atoms were subjected to OPLS4 Force force field [35], and ionization was done using Epik [33] to generate possible states at PH = 7 ± 2 units.

Glide docking

The prepared structures were subjected to XP Glide [36] with only ligands of atoms and rotatable bonds equal to or below 300 and 100 respectively were selected with a Vander Waals scaling factor of 0.8 and a partial charge cutoff of 0.15. The receptor was rigid with flexible ligand sampling and to sample nitrogen inversions and ring conformations. Bias sampling torsion was set for all predefined functional groups and Epik state penalties were added to docking scores. Post-docking minimization was performed with 10 poses per ligand with a threshold of rejecting minimized pose set to 0.5KCal/mol.

Binding Affinity determination

SeeSAR 12.1.0 [37] was used to perform binding affinity calculations. A binding site was defined by the co-crystallized ligand in the receptor PDB file and copied to the docking mode after the generated protein and docking library were loaded into the Biosolveit workspace [37]. Docking calculations were done for each compound in a standard docking mode with defaulted settings and parameters. The affinity of the generated poses was then assessed and the best poses were selected based on these affinities.

Binding energy is calculated from the HYDE score function (Eq. 1), which relies on intrinsically balanced terms of atom-specific desolvation, hydration, and hydrogen bonds based on the logP atomic increment system [38], while also taking into account the Torsion angle values to the binding conformation of the protein-ligand. The quotient of G and the number of non-hydrogen atoms in the molecule is used to define a ligand's binding affinity (Eq. (2)) [39] and expressed as.

$\Delta G_{Hyde} = \Sigma_{atom\ i} [\Delta G_i\ Dehydn + \Delta G_i\ h\text{-}bonds]$.....1

$$LE = \Delta G / N \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

where $\Delta G = -RT \ln K_i$ and N = number of non-hydrogen atoms.

The selection of the best poses was based on their visual HYDE scores while also considering a statistics-based torsional analysis.

ADMET Screening

All compounds were subjected to SeeSAR [37], AdmetLab2.0 [40], and QikProp [41,42]. The protocol used for SeeSar is as in the Binding Energy determination above. The concatenated sdf file was uploaded to the AdmetLab2.0 website (https://admetmesh.scbdd.com/) in the ADMET screening mode. It employs a Multi-task Graph Attention (MGA) framework made up of input, Relation graph convolution network (RGCN) layers, attention layer, and fully connected (FC) layers.

Molecular Simulation Dynamics

All simulations were done in GROMACS 2021.4 [43]. The protein topology was generated using AMBER99SB [44] force field from *pdb2gmx* module while ligands were parametrized by using Generalized Amber Force Field (GAFF2) [45] at antechamber website. The complexes were placed in the hexahedral box using *gmx editconf*, solvation was done using *gmx solvate*, TIP3P water molecules were added to the systems, and Na^+ and Cl^- ions were added using *genion* to a concentration of 0.15M. Equilibration was done with steps set to 20,000 in 10ns at 298K and 1atm with position restraints Force Constant set to 700KJmol^{-1}. This allowed water molecules and ions to move freely. Production MD simulations lasted for 100ns and they were monitored by checking system energies during the simulations. Pymol, vmd, and Gromacs binaries [43] were used for the analysis of the results.

MM/G(P)BSA calculations

Three energetic terms are taken into account in the computation of binding energy when determining the free energy of complex formation in conjunction with MD simulations: 1) variations in the system's potential energy in a vacuum; 2) polar and apolar solvation of the various species; and 3) the entropy related to complex development during the gaseous phase.

The Uni-GBSA-based tool *unigbsa-traj* was utilized to do MM/PBSA (Molecular Mechanics/ Poisson-Boltzmann Surface Area) calculations automatically for every simulated system [41,46]. The formula for estimation of free energy is explained in detail in [41,47,48].

Free Energy Perturbation (FEP) Calculations

The CHARMM-GUI Free Energy Calculator and NAMD were used to calculate absolute free energy [42,49,50]. The simulated ligands (docked postures) were uploaded as a single concatenated SDF file, and CHARMM General Force Field (CGenFF v1.x) [51] was used for ligand parametrization and topology construction to build the NAMD inputs and post-processing scripts. Counter ions (KCl) were used to neutralize all of the systems that were collected to produce input files and post-process scripts. Applying restraining potentials to limit the ligand's location in a receptor during FEP/MD allowed for the calculation of binding free energy using the double decoupling approach. With the resulting inputs, the TIP3P water model and Langevin piston pressure were applied to the system's NPT ensemble at 300K and 1atm. The FE values were captured in history files obtained via FEP lambda replica exchange MD (λ-REMD) using simple overlap sampling (SOS). By measuring the FE values throughout the last 6ns, the final FE values from 10ns FEP/λ-REMD simulations were computed [42]. The sequence of the used Methods is shown in Fig. 1.

Results and Discussion

Geometric deep learning Virtual Screening.

The results show that all optimal binding conformations had negative values (Fig. 2 and Table 1) and only the best 100 molecules were considered for further studies based on their potential scores, with malic acid having the best score of -286.78, however, the complexes of those ligands with many rotatable bonds (10≤) could not achieve successful optimizations and terminated after 500 iterations (Fig. 1), this is because an increase in rotatable bonds is associated with the inefficiency of the optimization algorithms when working with a large number of degrees of freedom. These ligands justified the need for conventional docking methods integration, all compounds having rotatable bonds above 10 underwent the same re-docking process [52] and average potential scores were reported.

Conventional docking and Binding Affinity determination.

The results of docking and binding affinity determination are shown in Table 1 and they agree with the results of geometric learning, validating the compounds under investigation. Phytyl diphosphate had the best score of -12.275Kcalmol^{-1}, but iso-citric acid had the best ligand efficiency of -3.100. The compounds had acceptable values of Binding affinities [37] with Isovitexin having the highest value (Table 1).

ADMET screening

The results for some important selected properties are presented in Table 1, and the summary for all properties is shown in Fig. 3 and Supplementary File 2, because the presence of PAINS alert is not enough to justify the elimination of a hit candidate [41], Compounds with pain alerts were only eliminated if at least other 4 ADMET properties were out of range (except the number of bonds, which was not considered as an ADMET property). The Synthetic Accessibility Score (SA Score) quantifies how simple it is to synthesize drug-like compounds. A score of less than 6 indicates that the compound is simple to synthesize [40,42]. An estimate of the hazardous dosage threshold of substances in humans can be obtained from the maximum recommended daily dose (FDAMDD). The likelihood of being toxic is the output value, and it ranges from 0-0.3 for excellent (less likely to be harmful); 0.3–0.7 for medium; and 0.7-1.0 for poor (more likely to be toxic) [40]. The mutagenicity test is the Ames test. The most used test for determining a compound's mutagenicity is the mutagenic effect, which closely correlates with carcinogenicity. The numbers indicate the probability of being harmful; 0-0.3 indicates excellent (less likely to be mutagenic), 0.3–0.7 indicates medium, and 0.7-1.0 indicates poor (more likely mutagenic) [40]. A fathead minnow LC50FM is defined as the concentration of the test chemical in water which results in 50% of the minnows dying after 96 hours. For LC50FM, the unit is − log10[(mg/L)/(1000*MW)] [40]. One reliable sign of DNA damage and other cellular stressors is the activation of p53. The output values of SR-p53 are the probabilities of being active, with 0-0.3 representing excellent (likely inactive), 0.3–0.7 representing medium, and 0.7-1.0 representing poor (most likely active) [40]. The rest of the computed ADMET properties are presented in Supplementary File 2[37,40] and their tSNE distribution is in Figure S1.

Molecular Dynamics Simulation Studies

The potential of six protein-ligand complexes as NS5B inhibitors was demonstrated by the formation of stable complexes by each of the simulated complexes.

The analysis of Root Mean Square Deviation (RMSD) evaluates the long-term structural stability of biomolecular simulations.

Table I: The summary of the results of virtual screening, molecular docking, and ADMET studies. Fun: A potential specific to a given target-ligand complex, SA Score: Synthetic Accessibility Score, LC50FM: 96-hour fathead minnow LC50, FDAMDD: maximum recommended daily dose, Ames: Mutagenicity test, SR-p53: Probability of being p53 actives.

Generic Name	Fun	Docking score ($KCalmol^{-1}$)	Glide ligand efficiency ln	Glide XP Hbond	QikProp Stars	Binding Affinity Range (mM)	SA Score	LC50FM	FDAMDD	Ames	SR-p53
[r]Phytyl diphosphate	-197.87	-12.275	-2.811	-2.000	3	0.198–19.699	4.851	5.102	0.905	0.004	0.004
***Apigenin-7-o-p-coumarylglucoside	-256.98	-12.038	-2.541	-4.427	5	1.2E6–1.2E9	4.081	5.818	0.122	0.686	0.969
Isocitric acid	-204.56	-11.050	-3.100	-2.480	3	1E3–1E5	3.356	3.261	0.006	0.021	0.004
*[r]LPA (18:0/0:0)	-187.45	-11.050	-2.530	-2.384	3	139–13835	3.463	3.18	0.765	0.015	0.009
Glucaric acid	-263.768	-10.667	-2.931	-2.664	2	57–5628	4.092	2.556	0.004	0.02	0.004
*[r]LPA (18:1(11Z)/0:0)	-242.98	-10.621	-2.432	-2.102	3	70986-7052884	3.664	2.805	0.887	0.024	0.008
*[r]LPA (18:2(9Z,12Z)/0:0)	-214.19	-10.210	-2.338	-1.310	3	0.768-76	3.846	3.58	0.947	0.004	0.01
[r]Sativic acid	-194.50	-10.062	-2.408	-1.777	3	19-1933	3.524	3.041	0.019	0.028	0.017
[r]LPA (18:1(9Z)/0:0)	-216.31	-9.978	-2.285	-2.568	3	1.515-150	3.664	3.257	0.889	0.006	0.009
**[p]Cynaroside	-196.45	-9.935	-2.225	-3.695	2	29-2868	3.924	4.845	0.031	0.757	0.835
[p]Cannabisin A	-222.58	-9.873	-2.064	-2.558	5	34041-3382134	2.817	4.116	0.493	0.494	0.809
**[p]Quercetin-o-glucoside	-277.72	-9.621	-2.140	-5.511	4	6179-613939	4.008	4.785	0.02	0.809	0.781
Gluconic acid	-268.92	-9.602	-2.694	-3.600	2	3393-337072	4.06	-0.392	0.002	0.053	0.003
Cannabitriol	-264.72	-8.908	-2.112	-1.184	0	2728-2710000	3.799	5.268	0.82	0.113	0.568
Malic acid	-286.78	-8.808	-2.755	-1.628	4	596-59228	2.893	2.579	0.011	0.02	0.004
Quinic acid	-222.48	-8.736	-2.450	-3.300	2	15396-1529705	3.598	0.807	0.016	0.031	0.003
[r]cannabisin d	-222.08	-8.722	-1.806	-2.754	4	7398-735004	3.412	5.418	0.946	0.791	0.811
[p]Orientin	-256.86	-8.627	-1.932	-4.613	5	947-94045	4.072	5.014	0.012	0.822	0.714
Isocannflavin B	-204.54	-8.540	-1.988	-1.567	0	3103-308292	2.724	5.884	0.303	0.521	0.893
**Apigenin-7-o-glucoside	-256.76	-8.522	-1.922	-1.790	1	83708-8316933	3.804	4.804	0.019	0.693	0.871
[p]N-trans-Caffeoyltyramine	-266.49	-8.499	-2.077	-1.912	0	8.240–819	2.061	4.611	0.64	0.386	0.82
3-[2-(3-Isoprenyl-4-hydroxy-5-methoxy-phenyl)ethyl]-5-methoxyphenol	-209.68	-8.456	-2.004	-1.296	1	269-26722	2.454	5.3	0.827	0.017	0.845
Cannabistilbene I	-196.04	-8.378	-2.026	-0.136	0	5.135-510	2.361	5.545	0.917	0.023	0.465
[p]Quercetin	-201.67	-8.360	-2.043	-2.400	0	9.593-953	2.545	5.222	0.31	0.657	0.888
Cannabidiolic acid	-228.47	-8.282	-1.945	-1.180	0	185-18368	3.626	5.699	0.787	0.023	0.513
[r]Arachidic acid	-148.87	-8.237	-2.013	-1.158	5	10.4–1037	1.665	4.743	0.014	0.005	0.034

Generic Name	Fun	Docking score ($KCalmol^{-1}$)	Glide ligand efficiency ln	Glide XP Hbond	QikProp Stars	Binding Affinity Range (mM)	SA Score	LC50FM	FDAMDD	Ames	SR-p53
Etofenoprox	-176.93	-8.220	-1.897	0.000	2	0.440-44	2.025	6.485	0.782	0.186	0.005
Delta-9-tetrahydro cannabinolic acid A	-279.84	-8.202	-1.926	-1.380	1	1028-102145	3.617	5.55	0.689	0.012	0.413
*Resmethrin	-232.34	-8.053	-1.909	0.000	1	11-1133	3.317	7.826	0.631	0.022	0.002
**Astragalin	-222.78	-8.016	-1.795	-2.564	2	3225-320374	3.884	4.732	0.011	0.775	0.853
Isovitexin	-197.23	-8.002	-1.805	-2.446	3	619407-61541783	4.081	5.818	0.122	0.686	0.969
ʳSofosbuvir	-201.60	-6.887	-1.503	-0.494	0	64-6368	4.375	4.036	0.9	0.371	0.02

* This compound has an ester and may undergo hydrolysis at high or low pH.

** This compound contains an acetal/aminal-like group (X-CH(R)-Y where X, Y are N, S, or O) that may be acid/base labile, releasing an aldehyde.

*** Both ester and acetal/ aminal-like groups.

ʳ Rotatable bonds above 10.

ᵖ PAINS alert (1)

It estimates the typical difference between an atom's location in a molecular structure and its reference structure [42,53]. The RMSDs of each system show that during the experiment, every simulated complex reached stability (Figs. 4A and S2A). Comparatively, the glucaric acid complex was more stable than the other compounds. The observed RMSD differences between the complexes and their corresponding proteins exhibit a variation that is driven by the ligand. When utilizing 1D RMSD, it is easy to believe that two structures that have the same RMSD from a reference frame are similar, but in practice, they can differ significantly. Alternatively, significantly more information can be obtained by computing the RMSD of each frame in the trajectory to all other frames in the other trajectory to give 2D RMSD [41]. Pairwise RMSDs of each trajectory were calculated to itself and the results are shown in Figure S5, the diagonal represents 0 (RMSD of a structure to its self), and all the structures had RMSDs $\leq 1.8A^{o}$, over simulation showing that the complexes were stable over simulation and had nearly same states but not identical except Phytyl diphosphate complex had more re-visited states than other complexes (Figure S5). These results agree with 1D RMSD results and show that the glucaric acid complex was more stable evidenced by lower RMSD values.

Root Mean Square Fluctuation (RMSF) analysis is another way to understand the flexibility and dynamic behavior of individual atoms or residues within a biomolecular system as well as their contribution to the flexibility of the whole molecule [54]. The main interacting residues have minimum RMSF values during simulation time compared to the ligand-free protein, supporting their stability and interactions with the simulated compounds, while the RMSF of non-interacting residues shows somewhat greater oscillations (Fig. 4C). The distribution of the rmsf further supports the stability of all complexes (Figure S4D).

The Radius of Gyration (Rg) is a measure of the compactness or spread of a biomolecular structure in three-dimensional space valuable for analyzing the overall shape and structural fluctuations of biomolecules [41,54]. A relatively constant Rg value fundamentally signifies a stably folded structure and a reduction in Rg signifies an increase in stability [41,42,54]. All of the simulated systems had Rg values between 23.7 and 24.2 A^{o}, indicating their stability. The simulation revealed a fairly progressive reduction in the gyration radius over time, indicating a gradual increase in the compactness and stability of all systems (Fig. 4D). The distribution of Rg also shows that the Glucaric acid complex is more stable relative to other complexes (Figure S3C).

Solvent Accessible Surface Area (SASA) is a measure of the surface area of a biomolecule that is accessible to solvent molecules. It plays a crucial role in analyzing the interactions between biomolecules and their surrounding solvent environment [41]. A higher value suggests an increase in the protein's volume, indicating lower stability, whereas stable proteins typically exhibit minimal fluctuation throughout the simulation. The binding of a small molecule can alter the solvent-accessible surface area (SASA) and significantly impact the protein's structure [41]. The SASA consistently decreased across all systems during the 100ns simulation (Figure S2B). This decrease in SASA indicates a rise in compactness, consequently indicating enhanced stability across all systems. The parallel patterns observed in both SASA and Rg affirm the validity of the molecular dynamics simulation outcomes [41]. Their distribution also shows that all systems were stable over simulation with the glucaric acid complex being more stable than others (Figure S3C and S3D).

During the simulation, the minimum distances between active site residues and ligands were calculated. Since most of the time, these minimum distances were less than 3A^{o}, a conventional hydrogen bond can still form as long as the acceptor and donor are orientated correctly (Figure S3A).

Non-bonded Molecular Mechanics (MM) interaction energy between the Ligands and their receptor was calculated to evaluate the magnitude of the interaction between the ligand and the protein [55,56]. The total interaction energy for all the systems was negative over 100ns simulation with the complex of phytyl-diphosphate having the lowest energy values (Figs. 4B and S3B). Vander Waals forces are short-range interactions that include London dispersion forces and dipole-dipole interactions, they enable the close interaction between the nonpolar regions of the ligand and the protein, ensure complex shape complementarity, predict the strength of the protein-ligand interaction with higher energy indicating stronger binding [57]. Apigenin coumaryl glucoside had the best values while that of Isocitric acid was the worst, a fact justified by their size and proximity to the protein groups (Figure S9). The stability of the complex and total binding energy can be enhanced by the long-range electrostatic interactions, which can direct protein molecules toward their pre-binding orientations [58]. Phytyl Diphosphate had the best values of electrostatic energy while apigenin coumaryl glucoside had the worst values since the former has phosphate groups which results in strong electrostatic interaction energy. These energies are good predictions of binding affinity since they are considered while computing those affinities. The simulated complexes remained stable throughout the simulation according to this data (Figure S2C and S2D). But it's crucial to remember that this quantity is neither a binding energy nor a free energy [55]. Figure 5 displays the specifics of the molecular interactions that occur in the middle of the simulation. The summary and distribution of all these energy terms are shown in Figures S3B, S4A, S4B, and S4C.

PCA (Principal Component Analysis) was performed particularly on alpha carbon data from the last 25ns to acquire insights into the dynamic behavior of both the complexes, incorporating structural and energy data. This analysis aimed to explore the conformational range of the complexes, distinguishing various regions within the energy landscape explored during the MD simulation. The complexes defined discrete conformational clusters and suggested stability by occupying compact subspaces. A graph illustrating the motion and displacement of atomic fluctuations within the complexes was created by utilizing eigenvectors 1 and 2. The first few eigenvalues had greater values, and the remaining eigenvalues were in a declining order [41]. All complexes showed PC ranging from ~ 15 to ~-15 (Figure S6). The simulation results of the apo-protein are shown in Figure S10. Interacting residues have increased positive correlation, as indicated by the Dynamic Cross-Correlation (DCC) data, indicating that they are interacting with the target. Residues located in the active site showed slightly elevated positive cross-correlation peaks upon ligand binding, suggesting a high occupant binding affinity (Figure S8) [41].

The average binding free energy of simulated hits was estimated using MM/P/GBSA calculations, and the outcomes are in good agreement with other studies. Since all energies were negative, proteins and ligands were strongly bound together (Table 2). Additionally, FEP/MD simulations were performed; Table 2's results show that these ligands have sufficient binding affinities, making them eligible for in-vitro research. To evaluate the convergence and dependability of the findings, the FE values were averaged using the standard error of the mean (SOS). The SEM of all data is less than 0.5kCal mol^{-1}, indicating that all systems have converged in less than 10ns [42,49,59].

Because of many RAASs in HCV proteins, reported DAAs resistance-enhancing RAASs were introduced in this protein with several combinations as shown in Table S1 and these mutants were re-docked with the identified compounds. 7 ligands still had significant docking scores with Glucaric acid having the highest values (Fig. 6 and Table S2). Paired 2-tailed t-tests to compare the wild type and mutants are also shown in tables S4, S5, S6, and S7. These results show that the docking scores of many ligands are significantly reduced except for the 7 ligands. This may suggest that some of these compounds, with glucaric acid leading, may still bind the resistant phenotypes. Studies have demonstrated the hepatoprotective, anti-inflammatory, cholesterol-lowering, anti-oxidant, and anti-carcinogenic properties of glucaric acid and its derivative D-saccharic acid-1,4-lactone [60].

Table 2
The Table shows the results of MM_G/PBSA and FEP/MD energy calculations.

Ligand	MM_GBSA (KCalmol^{-1})	MM_PBSA (KCalmol^{-1})	ΔG FE (k cal mol^{-1})
Phytyl Diphosphate	-57.331	-31.784	-9.45
Apigenin Coumarylglucoside	-46.342	-20.304	-8.24
Isocitric Acid	-43.111	-29.160	-9.06
LPA (18:0/0:0)	-60.753	-32.981	-8.64
Glucaric Acid	-66.876	-35.834	-7.32
LPA (18:1(11Z)/0:0)	-53.496	-28.507	-7.86

With the pan genomic approach to HCV therapy, the candidates were also docked with NS3/4 protease (PDB ID: 3P8N), and the results are presented in Table S3, 4 ligands including Glucaric acid had docking scores better than − 7.00KCalmol^{-1}. Having inhibitors inhibiting both proteins will be a bonus for drug development. To further discover the global therapeutic world of glucaric acid, insight into the compound's potential pharmacological targets and associated biological pathways was determined. This was done using Swiss Target Prediction to predict other possible targets of Glucaric Acid. Results (Fig. 7) show that this compound may be active against several targets including the Neuronal acetylcholine receptor protein

alpha-7 subunit (Probability = 0.765) and Squalene synthetase (Probability = 0.103). Inhibiting the former means that this compound can also be a competitive antagonist at a neuromuscular junction and hence can competitively compete with bungarotoxin minimizing its blockade activities at the junction [61] while inhibiting the latter is a potential therapeutic strategy for lowering cholesterol levels in individuals with hypercholesterolemia, as well as Anti-Cancer Potential Since cholesterol, is essential for the formation of lipid rafts and cell signaling pathways involved in cancer cell proliferation and survival, inhibiting squalene synthase can potentially inhibit tumor growth and metastasis [60,62]. The identified hits are shown in Figure S11.

Conclusion

We have used geometric deep learning to screen for possible HCV inhibitors of NS5B from cannabis sativa natural compounds. These compounds' data show that they may be potential DAAs against HCV. It is however very important to keep the RAASs in check, given that they represent the biggest challenge to HCV treatments, and screening them and recommending treatment post their identification will be a step ahead. The best hits identified were Glucaric acid and phytyl diphosphate, with the former retaining the ability to interact with mutants studied as well NS3/4 hence a possibility of polypharmacology.

Declarations

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The corresponding author has the data supporting this work, which can be shared upon inquiry.

References

1. Izhari, M. A. (2023). Molecular Mechanisms of Resistance to Direct-Acting Antiviral (DAA) Drugs for the Treatment of Hepatitis C Virus Infections. *Diagnostics 2023, Vol. 13, Page 3102*, *13*(19), 3102. https://doi.org/10.3390/DIAGNOSTICS13193102
2. Shepard, C. W., Finelli, L., & Alter, M. J. (2005). Global epidemiology of hepatitis C virus infection. *The Lancet. Infectious Diseases*, *5*(9), 558–567. https://doi.org/10.1016/S1473-3099(05)70216-4
3. *2021 Hepatitis C | Viral Hepatitis Surveillance Report | CDC*. (n.d.). Retrieved January 14, 2024, from https://www.cdc.gov/hepatitis/statistics/2021surveillance/hepatitis-c.htm
4. Venkatesan, A., & Prabhu Dass J., F. (2019). Review on chemogenomic approaches towards hepatitis C viral targets. *Journal of Cellular Biochemistry*, *120*(8), 12167–12181. https://doi.org/10.1002/JCB.28581
5. Xu, Z., Choi, J., Yen, T. S. B., Lu, W., Strohecker, A., Govindarajan, S., Chien, D., Selby, M. J., & Ou, J. H. (2001). Synthesis of a novel hepatitis C virus protein by ribosomal frameshift. *The EMBO Journal*, *20*(14), 3840–3848. https://doi.org/10.1093/EMBOJ/20.14.3840
6. Hang, J. Q., Yang, Y., Harris, S. F., Leveque, V., Whittington, H. J., Rajyaguru, S., Ao-Leong, G., McCown, M. F., Wong, A., Giannetti, A. M., Le Pogam, S., Talamás, F., Cammack, N., Nájera, I., & Klumpp, K. (2009). Slow binding inhibition and mechanism of resistance of non-nucleoside polymerase inhibitors of Hepatitis C virus. *Journal of Biological Chemistry*, *284*(23), 15517–15529. https://doi.org/10.1074/jbc.M808889200
7. Moradpour, D., & Penin, F. (2013). Hepatitis C virus proteins: from structure to function. *Current Topics in Microbiology and Immunology*, *369*, 113–142. https://doi.org/10.1007/978-3-642-27340-7_5
8. Irekeola, A. A., Ear, E. N. S., Amin, N. A. Z. M., Mustaffa, N., & Shueb, R. H. (2022). Antivirals against HCV infection: the story thus far. In *Journal of Infection in Developing Countries* (Vol. 16, Issue 2, pp. 231–243). Journal of Infection in Developing Countries. https://doi.org/10.3855/jidc.14485
9. Sorbo, M. C., Cento, V., Di Maio, V. C., Howe, A. Y. M., Garcia, F., Perno, C. F., & Ceccherini-Silberstein, F. (2018). Hepatitis C virus drug resistance associated substitutions and their clinical relevance: Update 2018. *Drug Resistance Updates*, *37*, 17–39.

https://doi.org/10.1016/J.DRUP.2018.01.004

10. Winquist, J., Abdurakhmanov, E., Baraznenok, V., Henderson, I., Vrang, L., & Danielson, U. H. (2013). Resolution of the interaction mechanisms and characteristics of non-nucleoside inhibitors of hepatitis C virus polymerase. *Antiviral Research*, *97*(3), 356–368. https://doi.org/10.1016/J.ANTIVIRAL.2012.12.027
11. Yi, G., Deval, J., Fan, B., Cai, H., Soulard, C., Ranjith-Kumar, C. T., Smith, D. B., Blatt, L., Beigelman, L., & Kao, C. C. (2012). Biochemical Study of the Comparative Inhibition of Hepatitis C Virus RNA Polymerase by VX-222 and Filibuvir. *Antimicrobial Agents and Chemotherapy*, *56*(2), 830. https://doi.org/10.1128/AAC.05438-11
12. Caillet-Saguy, C., Simister, P. C., & Bressanelli, S. (2011). An objective assessment of conformational variability in complexes of hepatitis C virus polymerase with non-nucleoside inhibitors. *Journal of Molecular Biology*, *414*(3), 370–384. https://doi.org/10.1016/J.JMB.2011.10.001
13. Boyce, S. E., Tirunagari, N., Niedziela-Majka, A., Perry, J., Wong, M., Kan, E., Lagpacan, L., Barauskas, O., Hung, M., Fenaux, M., Appleby, T., Watkins, W. J., Schmitz, U., & Sakowicz, R. (2014). Structural and regulatory elements of HCV NS5B polymerase - β-loop and C-terminal tail - Are required for activity of allosteric thumb site II inhibitors. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0084808
14. Han, D., Wang, H., Wujieti, B., Zhang, B., Cui, W., & Chen, B. Z. (2021). Insight into the drug resistance mechanisms of GS-9669 caused by mutations of HCV NS5B polymerase via molecular simulation. *Computational and Structural Biotechnology Journal*, *19*, 2761–2774. https://doi.org/10.1016/j.csbj.2021.04.026
15. Hang, J. Q., Yang, Y., Harris, S. F., Leveque, V., Whittington, H. J., Rajyaguru, S., Ao-Leong, G., McCown, M. F., Wong, A., Giannetti, A. M., Le Pogam, S., Talamás, F., Cammack, N., Nájera, I., & Klumpp, K. (2009). Slow binding inhibition and mechanism of resistance of non-nucleoside polymerase inhibitors of Hepatitis C virus. *Journal of Biological Chemistry*, *284*(23), 15517–15529. https://doi.org/10.1074/jbc.M808889200
16. Stammers, T. A., Coulombe, R., Rancourt, J., Thavonekham, B., Fazal, G., Goulet, S., Jakalian, A., Wernic, D., Tsantrizos, Y., Poupart, M. A., Bös, M., McKercher, G., Thauvette, L., Kukolj, G., & Beaulieu, P. L. (2013). Discovery of a novel series of non-nucleoside thumb pocket 2 HCV NS5B polymerase inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *23*(9), 2585–2589. https://doi.org/10.1016/J.BMCL.2013.02.110
17. Eltahla, A. A., Luciani, F., White, P. A., Lloyd, A. R., & Bull, R. A. (2015). Inhibitors of the Hepatitis C Virus Polymerase; Mode of Action and Resistance. *Viruses*, *7*(10), 5206–5224. https://doi.org/10.3390/V7102868
18. Hasanshahi, Z., Hashempour, A., Ghasabi, F., Moayedi, J., Musavi, Z., Dehghani, B., Sharafı, H., & Joulaei, H. (2021). First report on molecular docking analysis and drug resistance substitutions to approved HCV NS5A and NS5B inhibitors amongst Iranian patients. *BMC Gastroenterology*, *21*(1). https://doi.org/10.1186/s12876-021-01988-y
19. Nguyen, T. K., & Van Le, D. (2022). Identification of NS5B Resistance-Associated Mutations in Hepatitis C Virus Circulating in Treatment Naïve Vietnamese Patients. *Infection and Drug Resistance*, *15*, 1547–1554. https://doi.org/10.2147/IDR.S353605
20. Ahmed, H. R., Waly, N. G. F. M., Abd El-Baky, R. M., Yahia, R., Hetta, H. F., Elsayed, A. M., & Ibrahem, R. A. (2021). Distribution of naturally -occurring NS5B resistance-associated substitutions in Egyptian patients with chronic Hepatitis C. *PLoS ONE*, *16*(4). https://doi.org/10.1371/JOURNAL.PONE.0249770
21. Ssemuyiga, C. (2019). *In-vitro determination of antibacterial activity of Cannabis Sativa against staphylococcus aureusIn-vitro determination of antibacterial activity of Cannabis Sativa against staphylococcus aureus*. http://dissertations.mak.ac.ug/handle/20.500.12281/7027
22. Barré, T., Bourlière, M., Ramier, C., Carrat, F., Di Beo, V., Protopopescu, C., Marcellin, F., Bureau, M., Cagnot, C., Dorival, C., Zoulim, F., Zucman-Rossi, J., Duclos-Vallée, J. C., Fontaine, H., & Carrieri, P. (2022). Cannabis Use Is Inversely Associated with Metabolic Disorders in Hepatitis C-Infected Patients (ANRS CO22 Hepather Cohort). *Journal of Clinical Medicine*, *11*(20), 6135. https://doi.org/10.3390/JCM11206135/S1
23. Fischer, B., Reimer, J., Firestone, M., Kalousek, K., Rehm, J., & Heathcote, J. (2006). Treatment for hepatitis C virus and cannabis use in illicit drug user patients: implications and questions. *European Journal of Gastroenterology & Hepatology*, *18*(10), 1039–1042. https://doi.org/10.1097/01.MEG.0000236869.93527.B9
24. Sylvestre, D. L., Clements, B. J., & Malibu, Y. (2006). Cannabis use improves retention and virological outcomes in patients treated for hepatitis C. *European Journal of Gastroenterology and Hepatology*, *18*(10), 1057–1063. https://doi.org/10.1097/01.meg.0000216934.22114.51
25. Marcellin, F., Miailhes, P., Santos, M., Mercié, P., Di Beo, V., Salmon-Céron, D., Barré, T., Wittkop, L., Protopopescu, C., Zucman, D., Sogni, P., Carrieri, P., Salmon, D., Wittkop, L., Sogni, P., Esterle, L., Trimoulet, P., Izopet, J., Serfaty, L., … Wittkop, L. (2020). Cannabis Use and Plasma Human Immunodeficiency Virus (HIV) RNA Levels in Patients Coinfected With HIV and Hepatitis C Virus Receiving Antiretroviral Therapy: Data From the ANRS CO13 HEPAVIH Cohort. *Clinical Infectious Diseases*, *71*(9), 2536–2538. https://doi.org/10.1093/CID/CIAA117
26. Nordmann, S., Vilotitch, A., Roux, P., Esterle, L., Spire, B., Marcellin, F., Salmon-Ceron, D., Dabis, F., Chas, J., Rey, D., Wittkop, L., Sogni, P., Carrieri, P., Salmon, D., Trimoulet, P., Izopet, J., Serfaty, L., Paradis, V., Valantin, M. A., … Rosellini, S. (2018). Daily cannabis and reduced risk of steatosis in human immunodeficiency virus and hepatitis C virus-co-infected patients (ANRS CO13-HEPAVIH). *Journal of Viral Hepatitis*, *25*(2), 171–179. https://doi.org/10.1111/JVH.12797
27. Liu, Z., Li, Y., Han, L., Li, J., Liu, J., Zhao, Z., Nie, W., Liu, Y., & Wang, R. (2015). PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics*, *31*(3), 405–412. https://doi.org/10.1093/BIOINFORMATICS/BTU626
28. Su, M., Yang, Q., Du, Y., Feng, G., Liu, Z., Li, Y., & Wang, R. (2019). Comparative Assessment of Scoring Functions: The CASF-2016 Update. *Journal of Chemical Information and Modeling*, *59*(2), 895–913. https://doi.org/10.1021/ACS.JCIM.8B00545/SUPPL_FILE/CI8B00545_SI_001.PDF

29. Méndez-Lucio, O., Ahmad, M., del Rio-Chanona, E. A., & Wegner, J. K. (2021). A geometric deep learning approach to predict binding conformations of bioactive molecules. *Nature Machine Intelligence 2021 3:12*, *3*(12), 1033–1039. https://doi.org/10.1038/s42256-021-00409-9
30. Gainza, P., Sverrisson, F., Monti, F., Rodolà, E., Boscaini, D., Bronstein, M. M., & Correia, B. E. (2020). Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, *17*(2), 184–192. https://doi.org/10.1038/S41592-019-0666-6
31. Bishop, C. M. (1994). *Mixture Density Networks*. http://www.ncrg.aston.ac.uk/
32. Storn, R., & Price, K. (1997). Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, *11*(4), 341–359. https://doi.org/10.1023/A:1008202821328
33. Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., & Uchimaya, M. (2007). Epik: A software program for pKa prediction and protonation state generation for drug-like molecules. *Journal of Computer-Aided Molecular Design*, *21*(12), 681–691. https://doi.org/10.1007/s10822-007-9133-z
34. User Manual, L. (2015). *Schrödinger Press LigPrep 3.4 User Manual*.
35. Lu, C., Wu, C., Ghoreishi, D., Chen, W., Wang, L., Damm, W., Ross, G. A., Dahlgren, M. K., Russell, E., von Bargen, C. D., Abel, R., Friesner, R. A., & Harder, E. D. (2021). OPLS4: Improving force field accuracy on challenging regimes of chemical space. *Journal of Chemical Theory and Computation*, *17*(7), 4291–4300. https://doi.org/10.1021/ACS.JCTC.1C00302
36. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., & Mainz, D. T. (2006). Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of Medicinal Chemistry*, *49*(21), 6177–6196. https://doi.org/10.1021/jm051256o
37. *SeeSAR Beginner's Guide Version 13-Midas*. (2023).
38. Brethon, A., Chantalat, L., Christin, O., Clary, L., Fournier, J. F., Gastreich, M., Harris, C. S., Isabet, T., Pascau, J., Thoreau, E., Roche, D., & Rodeschini, V. (2017). New Caspase-1 inhibitor by scaffold hopping into bio-inspired 3D-fragment space. *Bioorganic and Medicinal Chemistry Letters*, *27*(24), 5373–5377. https://doi.org/10.1016/j.bmcl.2017.11.015
39. Schärfer, C., Schulz-Gasch, T., Ehrlich, H. C., Guba, W., Rarey, M., & Stahl, M. (2013). Torsion angle preferences in druglike chemical space: A comprehensive guide. *Journal of Medicinal Chemistry*, *56*(5), 2016–2028. https://doi.org/10.1021/jm3016816
40. Xiong, G., Wu, Z., Yi, J., Fu, L., Yang, Z., Hsieh, C., Yin, M., Zeng, X., Wu, C., Lu, A., Chen, X., Hou, T., & Cao, D. (2021). ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Research*, *49*(W1), W5–W14. https://doi.org/10.1093/NAR/GKAB255
41. Charles, S., Edgar, M. P., & Kasoma, N. A. (2023). The Hunt for Antipox Compounds against Monkeypox Virus Thymidylate Kinase and Scaffolding Protein Leveraging Pharmacophore Modeling, Molecular Docking, ADMET Studies and Molecular Dynamics Simulation Studies. *Virology & Mycology*, *12*(4), 1–14. https://doi.org/10.35248/2161-0517.23.12.280
42. Charles, S., & Mahapatra, R. K. (2023). Artificial intelligence based de-novo design for novel Plasmodium falciparum plasmepsin (PM) X inhibitors. *Journal of Biomolecular Structure and Dynamics*, 1–16. https://doi.org/10.1080/07391102.2023.2279700
43. Bauer, P., Hess, B., & Lindahl, E. (2022). *GROMACS 2022.1 Manual*. https://doi.org/10.5281/ZENODO.6451567
44. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, *78*(8), 1950–1958. https://doi.org/10.1002/PROT.22711
45. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, *25*(9), 1157–1174. https://doi.org/10.1002/JCC.20035
46. Yang, M., Wang, D., & Zheng, H. (n.d.). *UNI-GBSA: AN AUTOMATIC WORKFLOW TO PERFORM MM/GB(PB)SA CALCULATIONS FOR VIRTUAL SCREENING*. https://hermite.dp.tech/
47. Wang, J., Hou, T., & Xu, X. (2006). Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Current Computer Aided-Drug Design*, *2*(3), 287–306. https://doi.org/10.2174/157340906778226454
48. Gilson, M. K., & Honig, B. (1988). Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins*, *4*(1), 7–18. https://doi.org/10.1002/PROT.340040104
49. Zhang, H., Kim, S., Giese, T. J., Lee, T. S., Lee, J., York, D. M., & Im, W. (2021). CHARMM-GUI Free Energy Calculator for Practical Ligand Binding Free Energy Simulations with AMBER. *Journal of Chemical Information and Modeling*, *61*(9), 4145–4151. https://doi.org/10.1021/acs.jcim.1c00747
50. Gumbart, J., Hénin, J., & Chipot, C. (2023). *In silico alchemy: A tutorial for alchemical free-energy perturbation calculations with NAMD*. www.ks.uiuc.edu/Training/Tutorials/
51. Soteras Gutiérrez, I., Lin, F. Y., Vanommeslaeghe, K., Lemkul, J. A., Armacost, K. A., Brooks, C. L., & MacKerell, A. D. (2016). Parametrization of halogen bonds in the CHARMM general force field: Improved treatment of ligand-protein interactions. *Bioorganic & Medicinal Chemistry*, *24*(20), 4812–4825. https://doi.org/10.1016/J.BMC.2016.06.034
52. Li, H., Leung, K. S., Ballester, P. J., & Wong, M. H. (2014). istar: a web platform for large-scale protein-ligand docking. *PloS One*, *9*(1). https://doi.org/10.1371/JOURNAL.PONE.0085678

53. Martínez, L. (2015). Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. *PLOS ONE*, *10*(3), e0119264. https://doi.org/10.1371/JOURNAL.PONE.0119264
54. Charles, S., Edgar, M. P., & Mahapatra, R. K. (2023). Artificial intelligence based virtual screening study for competitive and allosteric inhibitors of the SARS-CoV-2 main protease. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2023.2188419
55. Arantes, P. R., Polêto, M. D., Pedebos, C., & Ligabue-Braun, R. (n.d.). *Making it rain: cloud-based molecular simulations for everyone Dynamics of bioactive compounds and impacts on molecular recognition View project Design and development of thrombin inhibitors View project*. https://doi.org/10.33774/chemrxiv-2021-9f2m5
56. *Host-Directed anti-fusion aptamers and small molecules as Respiratory Syncytial Virus (RSV) inhibitors: An Insilco-based study*. (2023). https://doi.org/10.21203/RS.3.RS-3318240/V1
57. Bitencourt-Ferreira, G., Veit-Acosta, M., & de Azevedo, W. F. (2019). Van der waals potential in protein complexes. *Methods in Molecular Biology*, *2053*, 79–91. https://doi.org/10.1007/978-1-4939-9752-7_6
58. Zhang, Z., Witham, S., & Alexov, E. (2011). On the role of electrostatics on protein-protein interactions. *Physical Biology*, *8*(3), 035001. https://doi.org/10.1088/1478-3975/8/3/035001
59. Jiang, W., & Roux, B. (2010). Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *Journal of Chemical Theory and Computation*, *6*(9), 2559–2565. https://doi.org/10.1021/ct1001768
60. Ayyadurai, V. A. S., Deonikar, P., & Fields, C. (2023). Mechanistic Understanding of D-Glucaric Acid to Support Liver Detoxification Essential to Muscle Health Using a Computational Systems Biology Approach. *Nutrients*, *15*(3). https://doi.org/10.3390/NU15030733/S1
61. Chang, C. C., Chen, T. F., & Chuang, S. -T. (1973). Influence of chronic neostigmine treatment on the number of acetylcholine receptors and the release of acetylcholine from the rat diaphragm. *The Journal of Physiology*, *230*(3), 613–618. https://doi.org/10.1113/JPHYSIOL.1973.SP010207
62. Seiki, S., & Frishman, W. H. (2009). Pharmacologic inhibition of squalene synthase and other downstream enzymes of the cholesterol synthesis pathway: a new therapeutic approach to treatment of hypercholesterolemia. *Cardiology in Review*, *17*(2), 70–76. https://doi.org/10.1097/CRD.0B013E3181885905

Figures

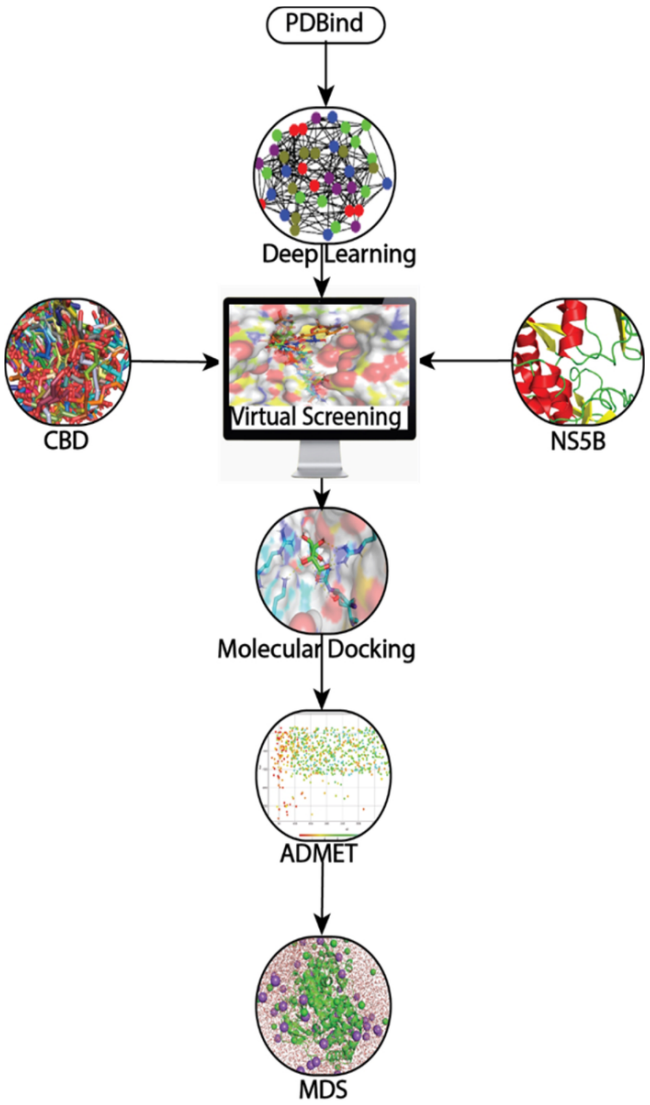

Figure 1

The sequence of the steps followed in this study.

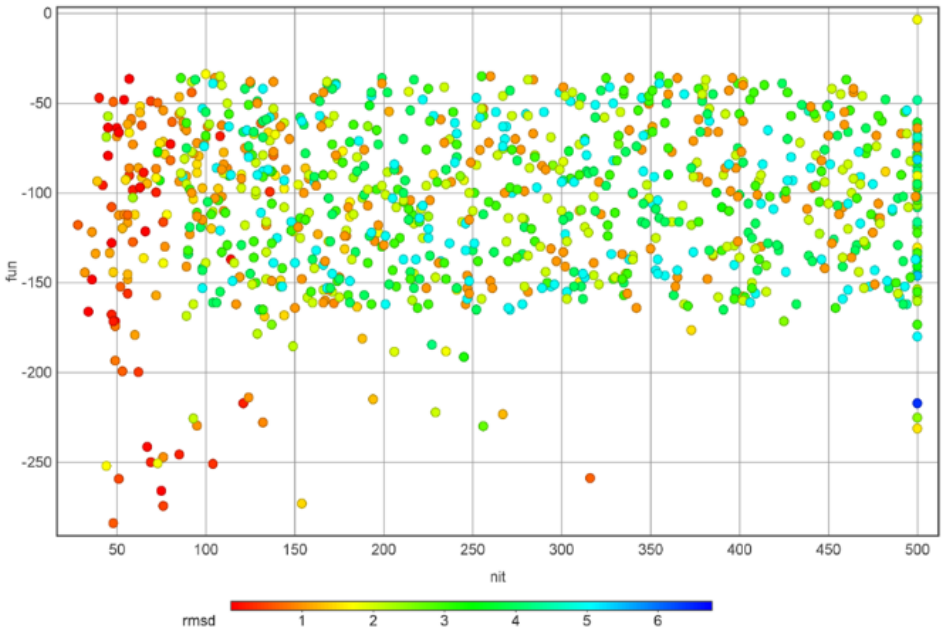

Figure 2

The Results of geometric Deep learning showing the distribution of some key parameters. Fun: Statistical potential, nit: Number of iterations, and rmsd: Root Mean Square Deviation.

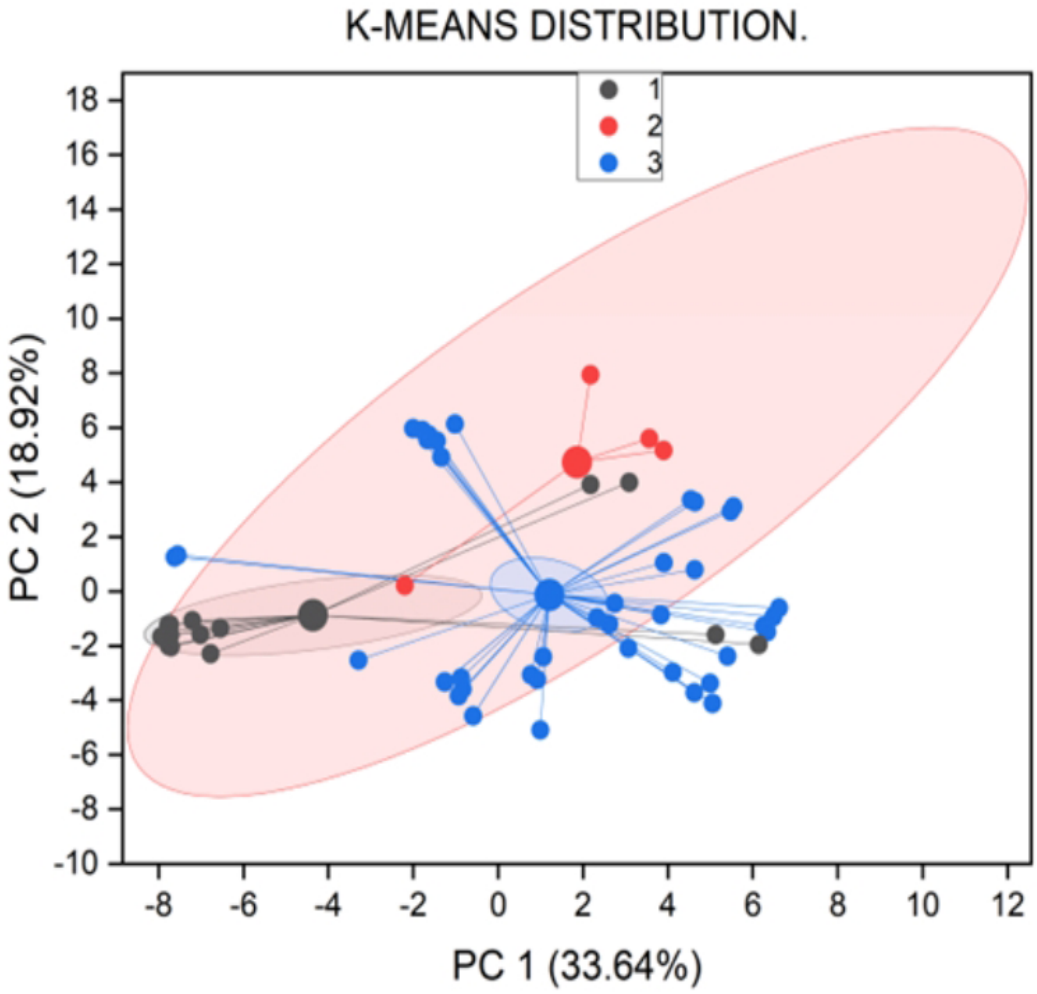

Figure 3

The K-Means Cluster analysis of ADMET properties

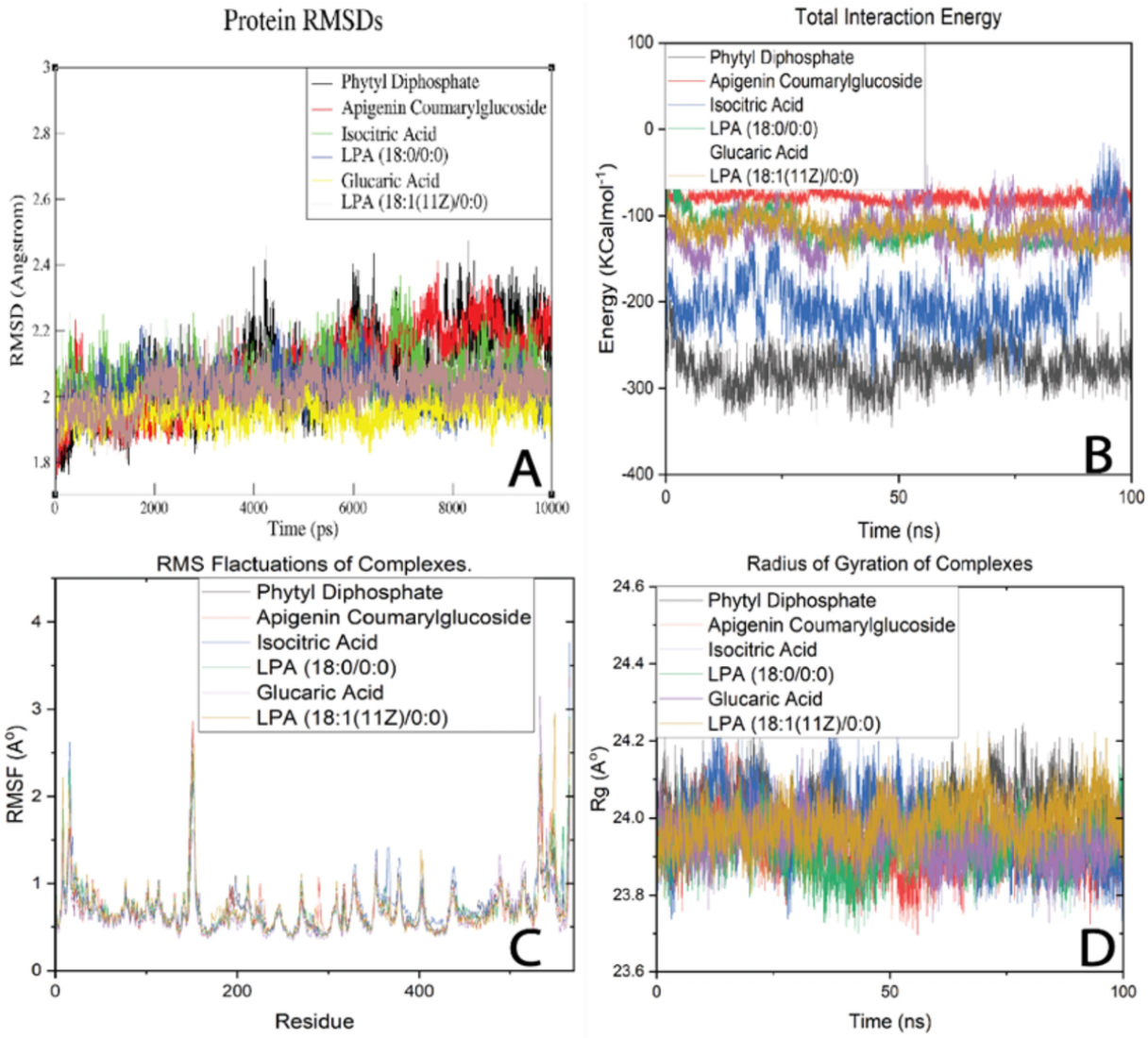

Figure 4

The results of molecular dynamics simulation. (A) Shows the variation of protein RMSDs from various complexes; (B) shows the total interaction energy for each system during the simulation; (C) depicts the variation of RMSF during the simulation; and (D) displays the variation of Rg during the simulation.

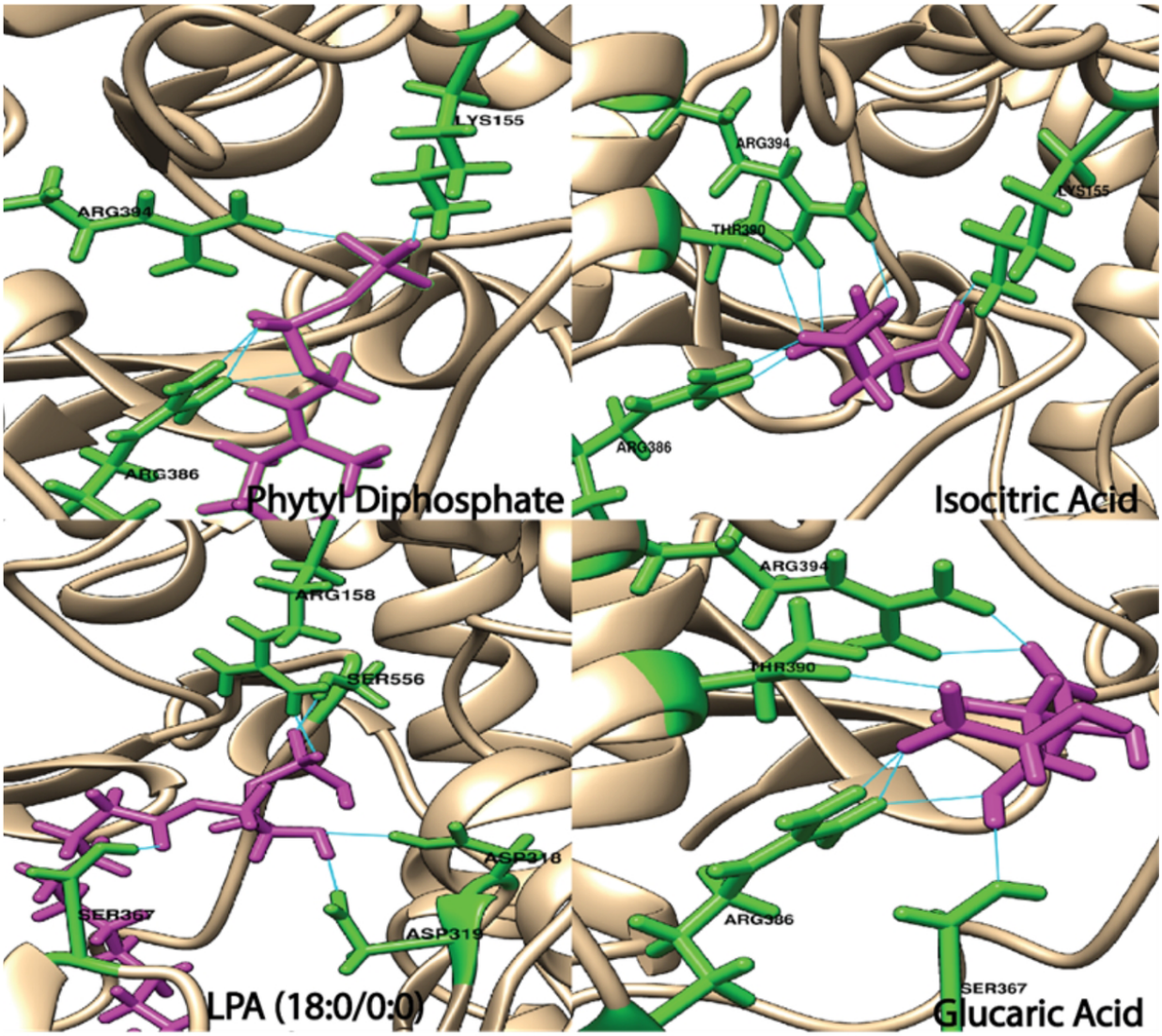

Figure 5

A glance of the interactions between the selected compounds and the target. The participating residues and compounds are depicted in licorice, with residues in green and compounds in pink, while the remaining protein is represented in a cartoon format.

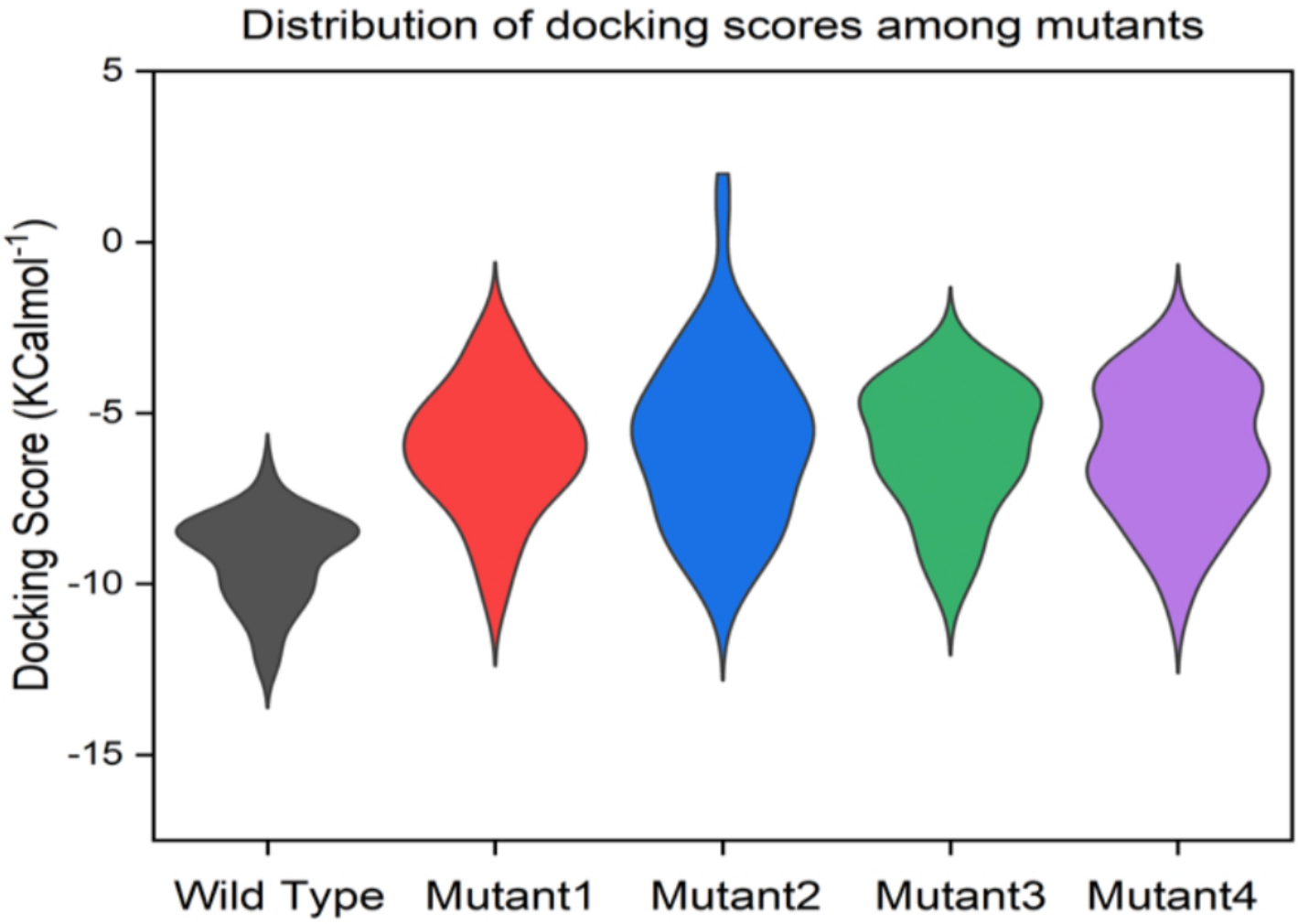

Figure 6

The results of docking the ligands into the active sites of mutants.

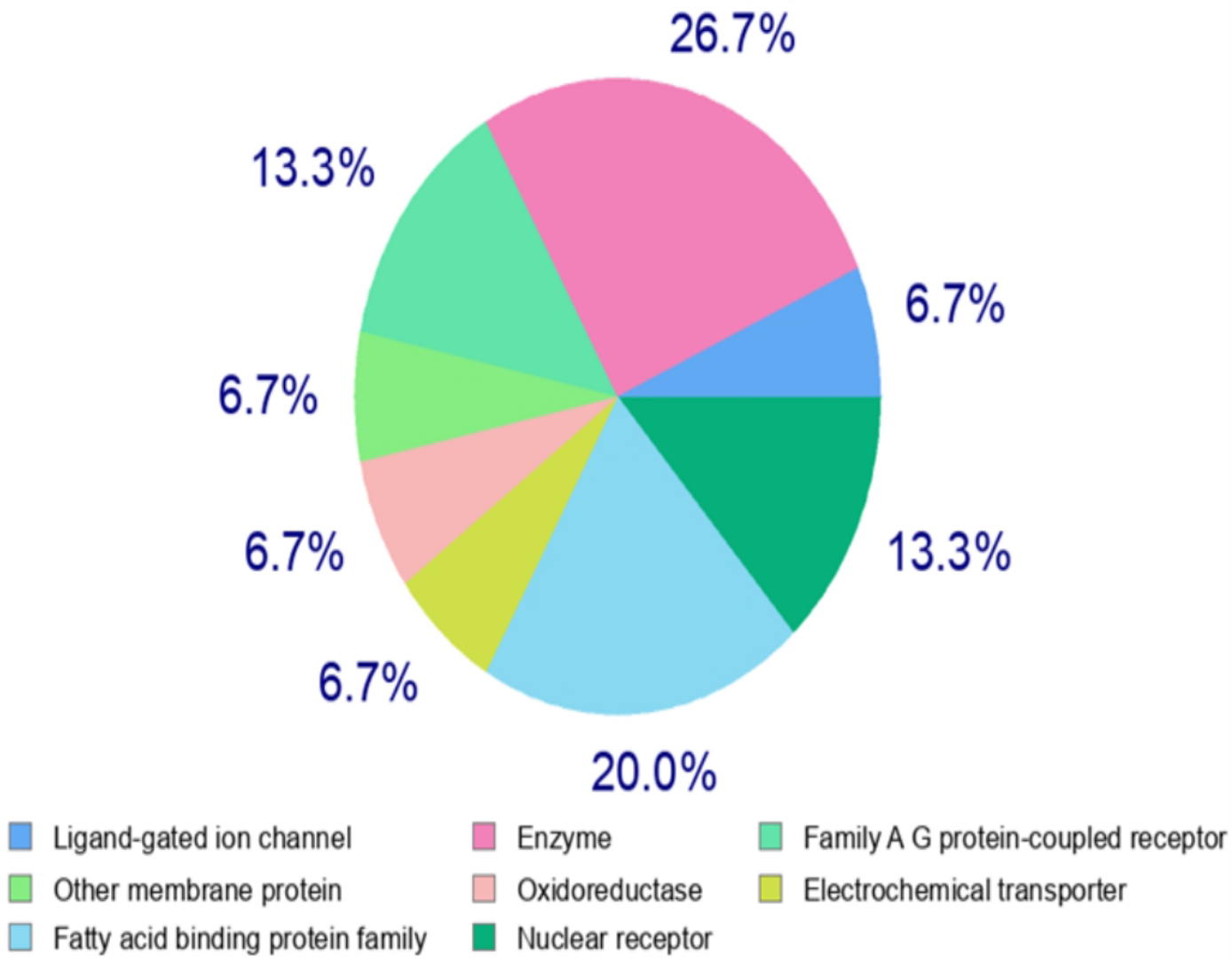

Figure 7

The results show other possible targets for Glucaric acid.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile1..docx