Heliyon 10 (2024) e34410

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identification of novel NLRP3 inhibitors as therapeutic options for epilepsy by machine learning-based virtual screening, molecular docking and biomolecular simulation studies

Maryam Zulfat [a], Mohammed Ageeli Hakami [b], Ali Hazazi [c,d], Arif Mahmood [e], Asaad Khalid [f,**], Roaya S. Alqurashi [g], Ashraf N. Abdalla [g], Junjian Hu [h], Abdul Wadood [a,*], Xiaoyun Huang [i,***]

[a] *Department of Biochemistry, Computational Medicinal Chemistry Laboratory, Abdul Wali Khan University, Mardan, Pakistan*
[b] *Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Shaqra University, Al-Quwayiyah-19257, Riyadh, Saudi Arabia*
[c] *Department of Pathology and Laboratory Medicine, Security Forces Hospital Program, Riyadh, Saudi Arabia*
[d] *College of Medicine, Alfaisal University, Riyadh, Saudi Arabia*
[e] *Department of Biochemistry, Quaid-i-Azam University Islamabad, Pakistan*
[f] *Substance Abuse and Toxicology Research Center, Jazan University, P.O. Box: 114, Jazan 45142, Saudi Arabia*
[g] *Department of Pharmacology and Toxicology, College of Pharmacy, Umm Al-Qura University, Makkah 21955, Saudi Arabia*
[h] *Department of Central Laboratory, SSL, Central Hospital of Dongguan City, Affiliated Dongguan Shilong People's Hospital of Guangdong Medical University, Dongguan, China*
[i] *Department of Neurology, Houjie Hospital and Clinical College of Guangdong Medical University, Dongguan, China*

ARTICLE INFO

Keywords:
NLRP3
Epilepsy
Machine learning
Molecular docking
MD simulation

ABSTRACT

The NOD-Like Receptor Protein-3 (NLRP3) inflammasome is a key therapeutic target for the treatment of epilepsy and has been reported to regulate inflammation in several neurological diseases. In this study, a machine learning-based virtual screening strategy has investigated candidate active compounds that inhibit the NLRP3 inflammasome. As machine learning-based virtual screening has the potential to accurately predict protein-ligand binding and reduce false positives outcomes compared to traditional virtual screening. Briefly, classification models were created using Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (KNN) machine learning methods. To determine the most crucial features of a molecule's activity, feature selection was carried out. By utilizing 10-fold cross-validation, the created models were analyzed. Among the generated models, the RF model obtained the best results as compared to others. Therefore, the RF model was used as a screening tool against the large chemical databases. Molecular operating environment (MOE) and PyRx software's were applied for molecular docking. Also, using the Amber Tools program, molecular dynamics (MD) simulation of potent inhibitors was carried out. The results showed that the KNN, SVM, and RF accuracy was 0.911 %, 0.906 %, and 0.946 %, respectively. Moreover, the model has shown sensitivity of 0.82 %, 0.78 %, and 0.86 % and specificity of 0.95 %, 0.96 %, and 0.98 % respectively. By applying the model to the ZINC and South African databases, we identified 98 and 39 compounds, respectively, potentially possessing *anti*-NLRP3 activity. Also, a molecular docking analysis produced ten ZINC

* Corresponding author.
** Corresponding author.
*** Corresponding author. Department of Neurology, Houjie Hospital and Clinical College of Guangdong Medical University, Dongguan, China
E-mail addresses: akahmed@jazanu.edu.sa (A. Khalid), awadood@awkum.edu.pk (A. Wadood), hxydg21@163.com (X. Huang).

https://doi.org/10.1016/j.heliyon.2024.e34410
Received 6 August 2023; Received in revised form 6 July 2024; Accepted 9 July 2024
Available online 9 July 2024
2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

and seven South African compounds that has comparable binding affinities to the reference drug. Moreover, MD analysis of the two complexes revealed that the two compounds (ZINC000009601348 and SANC00225) form stable complexes with varying amounts of binding energy. The in-silico studies indicate that both compounds most likely display their inhibitory effect by inhibiting the NLRP3 protein.

1. Introduction

Epilepsy is commonly prevailed diseases which is reported in people of different ethnicities, genders, and ages and has been known since the beginning of medical history [1]. Over 70 million individuals worldwide suffer from epilepsy which makes it the third most common neurological condition [2]. The identified risk factors for epilepsy include gene mutations, brain damage (such as trauma, stroke, brain tumors or status epilepticus), infections of the CNS (central nervous system), metabolic disorders and autoimmune diseases [3]. Yet, the etiology of epilepsy is unclear for many people. Moreover, epilepsy is usually accompanied by neurological comorbidities, such as cognitive impairment, depression, and anxiety, as well as mental morbidities, such as autism spectrum disorders. This has a significant impact on patients' quality of life. These coexisting conditions vary in severity and frequency depending on the cause of epilepsy and the age at which the disease manifests [4]. The development of epilepsy is thought to be mediated by a variety of underlying processes, including localized or systemic uncontrolled inflammatory responses, incorrect neural connections, and excessive stimulation of neuronal networks [5]. Moreover, research indicates that inflammation may both induce and result in epilepsy. Pro-inflammatory cytokines are released as a result of infection or hyperthermia-induced inflammation, which then raises the risk of seizures. On the other hand, protracted seizures and BBB (blood brain barrier) collapse cause an inflow of inflammatory chemicals that ultimately results in neuroinflammation [2].

Recent data from epilepsy patients and experimental models suggest that the inflammasome pathway may be a novel inflammatory signaling system implicated in epileptogenesis [6]. A potential strategy for treating epilepsy and altering its course is the inhibition of pro-inflammatory signaling molecules. It has been reported that the NLRP3 inflammasome, is an example of a significant regulator responsible for the release of pro-inflammatory cytokines, that promotes Caspase-1 activation, which in turn cleaves pro-Interleukin-1β (IL-1β) to generate the mature Interleukin-1β, is a crucial mediator of Interleukin-1β function [7,8]. In 2014, a study was conducted in rats with the goal of examining the role of NLRP3 inflammasome in neuroinflammation, hippocampal neuronal loss, and spontaneous recurrent seizures in an amygdala kindling-induced status epilepticus (SE) model. The results of this study provided the first indication of a correlation between NLRP3 and epilepsy [9]. In 2022, Zhang et al. discovered that NLRP3 inflammasome activation could speed up epilepsy by regulating the CREB/REST/SP1 signaling pathway and increasing adenosine kinase expression in mouse and cell epilepsy models [10]. This is consistent with the findings of Gong et al. who observed increased expression of NLRP3 in epileptic neurons and rats in addition to downregulated expression of miR-138-5p, upregulated expression of LncRNA ZNF883 and ubiquitin-specific peptidase 47 (USP47), and concurrent with aggravated inflammation and apoptosis [11]. These findings indicate that inhibiting the NLRP3 inflammasome may be a therapeutic target for epilepsy. Consequently, several companies have shown interest in developing NLRP3-targeting compounds in recent years, given the promising potential of targeting NLRP3 for anti-epileptic therapy. Due to the complex signaling cascade of the NLRP3 inflammasome, several targets can be identified to directly or indirectly inhibit NLRP3 activity. But it is important to emphasize that the best way to treat NLRP3-related diseases such as epilepsy is to use specific NLRP3 inhibitors [12].

The NLRP3 inflammasome [13], consists of NLR family, pyrin domain containing 3 (NLRP3), apoptosis-associated speck-like protein containing a caspase recruitment domain (ASC) and caspase-1, regulates IL-1β transcription and perform other tasks by interacting with the NF-κB inflammatory pathway [13,14]. It has been shown that innate immunity and inflammatory regulation of the central nervous system are associated with the NLRP3 inflammasome [15–19]. Noteworthy factors that can activate NLRP3 include the production of ROS (reactive oxygen species), high levels of extracellular ATP and K+ ions, acidosis, cell swelling, hypoxia and an increase in intracellular Ca2+ concentration [20–22]. Yet, there is evidence linking the abnormal stimulation of the NLRP3 inflammasome to a number of inflammatory diseases, like inflammatory bowel disease, Alzheimer's disease, type 2 diabetes and epilepsy. Hence, it is imperative to carefully control NLRP3 inflammasome overstimulation to provide enough immune protection without affecting the host tissues. It is evident from prior research that blocking the NACHT domain is a useful strategy for preventing NLRP3 inflammasome overactivation [23].

The process of discovering novel drugs is expensive and time-consuming; it might take 10–15 years to introduce a single medicine to the market. Machine learning algorithms are widely employed in fields like drug discovery since they are cost-effective and efficient [24]. These methods are perfect for categorizing compounds as active or inactive through virtual screening of vast compound libraries [25]. In this study, we describe a machine learning (ML) based virtual screening (VS) method that was used to find new candidate NLRP3 inhibitors. Our method involved docking assessments, molecular dynamics (MD) simulations, and finally, the prediction of novel and potential drug-like compounds to inhibit the NLRP3-mediated epileptic symptoms.

2. Methodology

2.1. Collection and preparation input dataset

The ChEMBL database, Binding DB (http://www.bindingdb.org), and literature were surveyed to collect the dataset of NLRP3 inhibitors. The decoys inhibitors of NLRP3 were generated while using LIDeB (https://lideb.biol.unlp.edu.ar/). The SMILE notation was applied to filter out duplicate molecules from the database [26]. Furthermore, the dataset was split out into active and inactive ligands in order to develop ML-based binary classification models.

2.2. Descriptors computation and features selection

The 2D descriptors of the dataset were calculated through MOE [27], and features selection was achieved by using RFE (recursive feature elimination) which eliminates features one at a time until only a certain number of features are left, hence reducing the model's complexity.

2.3. Generation of features-based ML models

In order to develop various classification and regression modules for machine learning models, we implemented the Scikit-learn Python version 3.9 [28], which includes a large number of built-in modules. For the binary classification of our data, we employed numerous algorithm (listed below). To test the accuracy of all the models without bias, the test data in the compound library was then split into 918 (70 %) training and 394 (30 %) test set compounds.

2.3.1. Support vector machine (SVM) model

A well-liked ML technique for regression and classification is the support vector machine model. The Scikit learn library of Python 3.9 [28,29] utilizes the training data set of ligands to develop a SVM model to classify and distinguish between actives and inactive molecules of the input chemical library. In this approach, each data value is represented as a point in n-dimensional space. The best hyperplane for separating the two classes is then determined, followed by classification.

2.3.2. Random Forest (RF) model

RF is also a standard ML technique for regression and classification [30]. It is based on the concept of ensemble learning. Several classifiers are integrated into ensemble learning to tackle a complicated issue and enhance model performance. It examines the forecast from multiple trees and produces the final result based on the majority of projected outcomes. The high numerical value of trees in the RF contributes to higher accuracy and minimizes data overfitting.

2.3.3. K-nearest neighbors (KNN) model

The k-NN is one of the simplest and earliest classification [31] method that was first employed in statistical applications in the early 1970s [32]. The basic idea behind KNN is that it ranks the query data points based on the votes received from the nearest k data point, (e.g., based on distance functions), utilizing different types of distances including Euclidean distances, hamming distances, and cosine distances. The regression prediction value for the first sample was chosen to be the average value of the K ($K = 4$) samples that were closest to the target to be predicted.

2.4. Machine learning (ML) model validation

The K-fold cross-validation (CV) approach is often exploited to assess the performance of classifier since it eliminates the problem of overfitting. The k-fold cross-validation approach does not train the model using the complete dataset, but rather randomly divides it into k equalized smaller subsets. A selected subset of the k subsets is kept as validation data for testing the model, while the remaining $k-1$ subsets are utilized as training data. The CV procedure is then repeated k times (the folds), with each of the k subsets utilized once as test data to evaluate the model's performance. The average of the k-fold findings may then be utilized to assess the performance of the created model.

One of the main tools for diagnostic test assessment is a receiver operating characteristic (ROC) plot, which is made by comparing the true positive (TP) rate against the false positive (FP) rate at various threshold levels. This approach is frequently used to determine the area under the ROC curve (AUC) to assess the effectiveness of a classification model. The greater the AUC (a perfect prediction will make AUC equal to 1), the higher will be the model's predictive accuracy.

The performance evaluation metrics we utilize are sensitivity, specificity, precision, overall accuracy and F1 score:

$$\text{Sensitivity } (Sn) = \frac{TP}{(TP + FN)}$$

$$\text{Specificity } (Sp) = \frac{TN}{(TN + FP)}$$

$$\text{Precision} = \frac{TP}{(TP + FP)}$$

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FN + FP)}$$

$$\text{F1 score} = \frac{2 \times TP}{(2 \times TP + FN + FP)}$$

Where TP indicates the total number of properly predicted actives, TN represents the total number of correctly predicted inactive, FP represents the number of inactive miss-predicted as active, and FN represents the number of active recognized as inactive.

Many statistical metrics were used to assess each ML model. To analyze several models, we computed parameters such as sensitivity, specificity, accuracy, F1 score, and precision. The percentage of inhibitors properly predicted by the model is referred to as sensitivity. Although specificity is defined as the percentage of non-inhibitors accurately predicted by the model and accuracy is defined as the proportion of data points successfully predicted. A measure of the model's ability to make accurate positive predictions is called precision. In other words, it includes false positives because it gives the ratio of the number of correctly predicted positive outcomes to all positive outcomes. For virtual screening, the model with highest evaluation values for the training and test sets were employed [33].

2.5. Prediction on an external dataset

We used commercially available ZINC and Southern African databases for virtual screening purpose. Particularly, some areas of those databases that are intended for the identification of new NLRP3 inhibitors were taken into consideration and put through the same data curation procedure as training and test sets, and descriptors were calculated. In order to find probable novel active compounds, the best generated model was applied for screening the in-house database (comprised of ZINC and South African Databases). During the screening, only compounds that had the best probability score for each database were retrieved.

2.6. Molecular docking using MOE

To identify the probable interaction between ligands and the NLRP3 protein, docking was carried out using the MOE program [34]. The 3D structure of NLRP3 was retrieved from PDB databank, where it's PDB ID is 7ALV, and docking was performed using this structure [35]. Prior to docking, the structure of NLRP3 was cleaned by deleting the previously inbound ligand to make the active site available for the new compounds. Moreover, the co-crystal water molecules, salts and other unwanted molecules were also removed so that the interaction site would not be obscured during the docking. The *MOE structure preparation* module was then employed to fix errors, correct structure and add missing atoms to the structure of NLRP3 if any.

2.7. Re-docking of compounds using PyRx

It is crucial to carry out docking with another software package in order to further validate the findings. In order to achieve this goal, we choose the open-source virtual screening tool PyRx, with two built-in features: Vina Wizard (Auto-dock vina) for docking and Open Babel for file conversion [36]. Using a graphical user interface, PyRx operates. The target protein was initially obtained from the PDB and then further processed using the Discovery Studio visualizer to get rid of any repeating chains, water molecules, heteroatoms and ligands that were already attached. After eliminating all redundant chains, only the chain of interest at which the active site was located remained. This file was then stored in pdb format. By loading already prepared receptor into PyRx software, the redocking process was carried out. By utilizing the import option, a 2D conformer of the ligands in structure data files (SDF) format was imported. Prior to docking, the ligands' energy was minimized and then changed to pdbqt format. Finally, the Grid Box was set up around the protein's active site using the Vina Wizard option, allowing the programme to only dock and search for the maximum score in a specific region within the box. The start option was used to get the docking result.

2.8. Molecular dynamics (MD) simulation protocol

Molecular dynamics simulation using the Amber20 algorithm was used to evaluate the stability of the top-scoring compounds [37]. The topology of NLRP3 protein was prepared using the ff14SB force field. An antechamber was utilized to study ligand's topology. The TIP3P water model with a box dimension of 10.0 Å was used to adequately solvate each system. The neutrality of the system was maintained by counter-ions (Na^+ and Cl^-). Each system was relaxed using the 1000 steps steepest descent energy minimization followed by 500 steps of conjugate gradient energy minimization in order to eliminate undesirable atoms collisions. The temperature was then raised to 300 K for each system. Each system was heated and then subjected to a two-steps equilibration process at constant pressure (1 atm) and temperature (300 K). The temperature of each system was regulated by a Langevin thermostat [38]. The PMEMD (Particle Mesh Ewald algorithm) was utilized to handle long-range electrostatic interactions [39]. The SHAKE algorithm was used to address the covalent bonds [40]. For each system, the molecular dynamics simulation production steps were carried out by the GPU-supported pmemd code [41] and the trajectories were examined through the CPPTRAJ package. The output files from the

CPPTRAJ module [42] were graphically visualized and interpreted using the Origin Pro program [43], and the protein visualization was done using MOE2016 and Pymol v2.0.

3. Results and discussion

3.1. Input dataset collection and machine learning-based model generation

Numerous literature has been reported the use of machine learning (ML) in computer-aided drug designing (CADD) [44–47]. In this study, we have also exploited the ML-based virtual screening of commercially available ZINC and Southern African databases to identify novel and more potent inhibitors against the NLRP3 protein which is implicated in epilepsy. Inhibitors against the NLRP3 were retrieved from literature, binding DB and CHEMBL databases and decoys were generated based on these inhibitors using LIDeB, an online tool. All of the duplicate compounds were eliminated once the dataset's redundancy was examined. The compounds obtained from literature, binding DB and CHEMBL were labeled as active compounds, while the decoy compounds and some inactive compounds from CHEMBL database were presumed as inactive compounds. Our results confirmed 425 active and 887 inactive compounds against the NLRP3 during the initial curation of the already existed data of the NLRP3 protein. The dataset's active compounds were given the label "1," whilst its inactive compounds got the label "0." The molecules' weight and logP were used to define the chemical space. This training set contains enough chemical space to be employed for the creation of ML models due to the wide variety of molecular weights (which vary from around 165 to 600) and the h-logP (which ranges from 2.5 to 10.0) (Fig. 1). The dataset containing active and inactive compounds were randomly divided into 70 % training and 30 % test sets. In total, there were 918 compounds in the training set and 394 compounds in the test set, as shown in Fig. 1.

3.2. Descriptor calculation

The chemical features of the ligands are represented by molecular descriptors utilizing numerical values. By using MOE program [27], we generated 208 descriptors for our data. The descriptor data set was used to create machine learning models utilizing various algorithms.

3.3. Feature selection

One of the core ideas of machine learning, the feature selection has a valuable effect on model's performance. Feature selection is very beneficial before modeling data because it improves accuracy, reduces overfitting and training time. In this study, feature selection was performed with the help of Recursive Feature Elimination (RFE) and reduced the 2D features from 208 to 42. The selected 2D features included a total of 42 optimum features were further used for ML based model generation. All the irrelevant features were removed from the dataset.

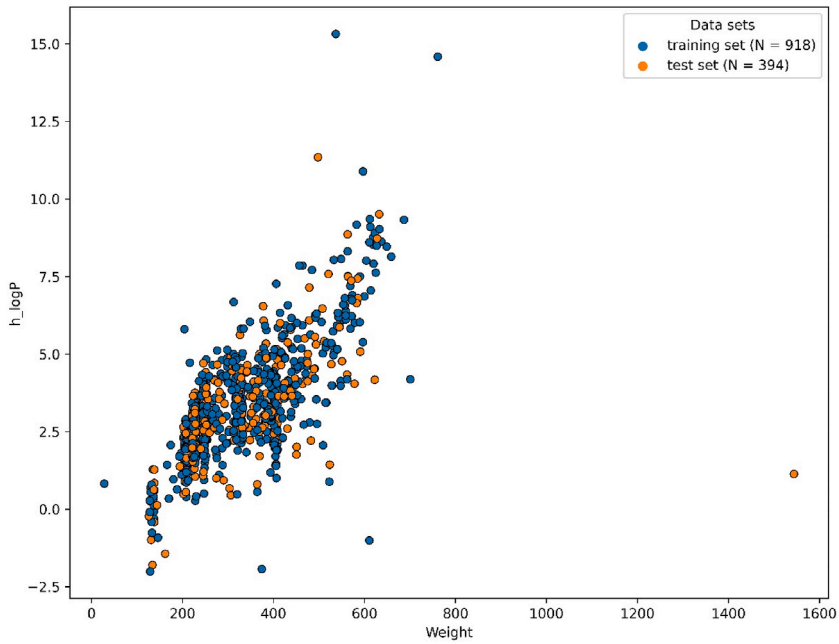

Fig. 1. Chemical space analysis of the molecules in the training set, with molecular weight on the X-axis and logP on the y-axis.

3.4. ML-based model generation and evaluation

To create ML models, Python version 3.9's Scikit-learn package [28] was used. Using the descriptor data, we designed three distinct models that are frequently employed to address binary classification problems. The accuracy for the KNN classifier was 0.911, while the precision and roc_auc scores were 0.88 and 0.886, respectively. The SVM model accuracy was 0.906, and roc_auc score was 0.871. The SVM model precision value was 0.90. The accuracy of the RF model was 0.946, while the roc_auc score was 0.925. The accuracy of the RF model was 0.946, while the roc_auc score was 0.925. The RF model has precision value of 0.95. The RF model, followed by the KNN, had the highest value in terms of statistical parameters when all of the statistical values of the other models were compared. The RF model has precision value of 0.95. The RF model, followed by the KNN, had the highest value in terms of statistical parameters when all of the statistical values of the other models were compared. All the statistical values of several models in this study have been displayed in Table 1. In order to screen the ZINC and Southern African databases using the ML approach, the RF model was utilized based on its best statistical results as compared to other models. Since the ML model reported in our study has obtained significantly higher parameters, therefore we argue that the prediction of active hits by this model will have better probability. Fig. 2 shows the ROC curves of the ML models. ROC plot is the representation of false positive rate (X-axis) and true positive rate (Y-axis) for all the samples thresholds between 0 and 1.

3.5. ML-based screening of ZINC and Southern African databases

The Southern African and ZINC databases, which include 1012 and 136,564 compounds, respectively, were utilized to find possible hits against the NLRP3 protein. Initially, 2D descriptors were computed with the MOE program and save in the CSV (comma-separated value) format. The descriptor file was then loaded as a data package into Scikit Learn using pandas. Using the dataset descriptor file, essential features were selected and utilized to create the ML model for the NLRP3. A new data frame was created after the descriptors were extracted, and the RF model was used to screen the results. We were able to screen 39 entries from the South African database and 98 hits from the ZINC database (Table 2).

3.6. Molecular docking calculations with MOE

We docked (98 and 39 compounds from ZINC and Southern African databases, respectively) in the active site of the NLRP3 NACHT domain. The molecular docking results showed that the reference compound "MCC950" interacted with the NLRP3 NACHT domain with a docking score (S-score) of −7.6432 kcal/mol. In order to narrow down the list of possible inhibitors from the docking simulation, we set a threshold of docking score −7.5 kcal/mol. Application of this threshold value yielded 10 molecules from ZINC database [ZINC000009601348, ZINC000003103298, ZINC000020938182, ZINC000017135518, ZINC000024985355, ZINC000019810420, ZINC000001376250, ZINC000009795937, ZINC000033063933 and ZINC000044168026] and 7 molecules from South African database [SANC00225_minRM1.pdb, SANC00529_minRM1.pdb, SANC00639_minRM1.pdb, SANC01077_minRM1.pdb, SANC01125_minRM1.pdb, SANC00101_minRM1.pdb and SANC00457_minRM1.pdb]. Molecular docking study demonstrated that most of our identified inhibitors showed higher docking score than the reference compound MCC950. By analyzing the docking results of ZINC database, it was observed that all the inhibitors showed promising interaction with the NLRP3 protein, but the most potent compound was ZINC000009601348 (hereafter hit1) with a docking score of −10.422. Hit1 fitted well in the active site of NLRP3 NACHT domain and established three H-bond acceptor interactions with ALA 228 and ARG 351 residues, two pi-cation interactions with ARG 178 and one Pi-H interaction with ALA 228. The second most potent compound was ZINC000019810420 (hereafter hit2) followed by ZINC000017135518 (hit3) with a docking score of −9.9841 and −9.194, respectively. Hit2 formed interactions with ALA 228, ARG 351, LEU 187 and ALA 227 active site residues. Similarly, Hit3 made interactions with ALA 228, THR 524, VAL 414 and VAL 353 active site residues of the NLRP3 NATCH domain. The docking score, interacting residues of NLRP3 NATCH domain, and types, distance and energy of each interaction established by the newly identified inhibitors from the ZINC database are shown in Table 3. By Docking the active hits from the South African database with the active site of the NLRP3 NACHT domain compound SANC00225_minRM1.pdb (hereafter hit4) revealed a best docking score of −8.9764. Hit4 formed strong H-bond interactions with GLU 527, ALA 228, ARG 351, and ALA 227 active site residues of the NLRP3 NATCH domain. The docking results of the identified NLRP3 inhibitors from South African database are shown in Table 4. As compared to reference compound most of the newly identified compounds made strong interactions and obtained a best docking score. The most probable 3D interactions of the reference compound and the top hits from the docking simulation of both the databases are shown in (Fig. 3A–C).

Table 1
Statistical results of several ML models.

Model	Accuracy	Sensitivity	Specificity	Precision	F1score	roc_auc score
KNN	91 %	0.82	0.95	0.88	0.85	0.88
SVM	90 %	0.78	0.96	0.90	0.83	0.87
RF	94 %	0.86	0.98	0.95	0.91	0.92

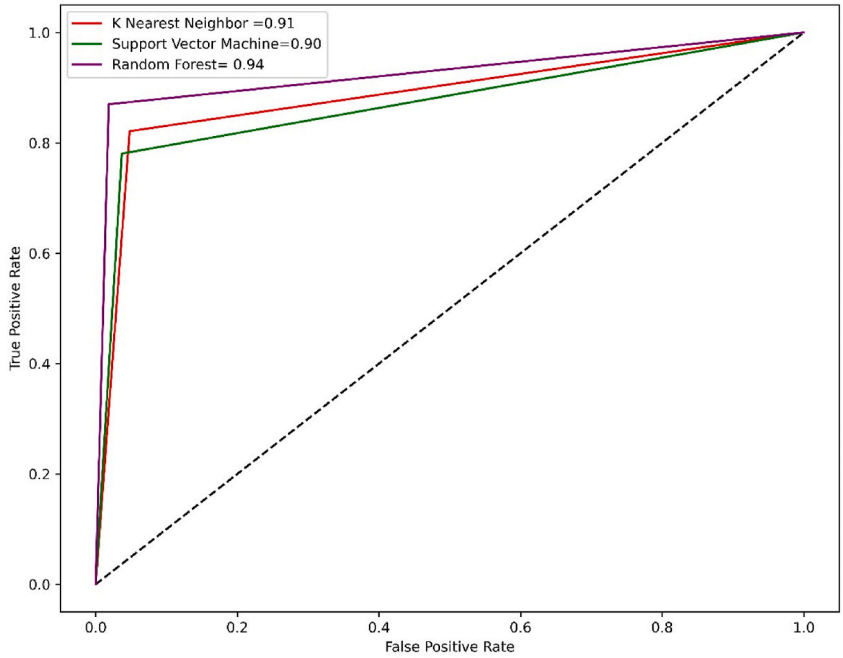

Fig. 2. Receiver operating characteristic (ROC) plot of the three machine learning models.

Table 2
Results on unseen dataset.

Database	No of compounds	No of hits predicted from ml based virtual screening
ZINC	136,564	98
South African	1012	39

3.7. Re-docking scores with PyRx

PyRx re-docked ligands (obtained from ML-based VS) with receptor protein (7ALV) to validate the docking protocol by MOE. The re-docking results of top 10 ranked compounds from ZINC and top 7 ranked compounds from South African are shown in (Table 5) and (Table 6).

3.8. Drug-likeness prediction

Lipinski's rules of five (RO5) states that a ligand could be considered as a drug-like molecule if the criteria i-e. number of hydrogen bond acceptors and donors are $\leq$10 and 5, respectively, logP value is $\leq$ 5 and molecular weight is $\leq$ 500 Da [48]. According to Lipinski's Ro5, all the tested compounds in this study were acceptable according to the drug scanning results. All the top-ranked ligands followed Rule of 5 (Table 7) and (Table 8).

3.9. Ligands with best binding energies

Ligands with the lowest docking score (S-score), binding affinity and promising binding pose were found as the best NLRP3 binders. Visual analysis of the interaction between the protein-ligands were performed in Pymol [49]. ZINC000009601348 and SANC00225, two compounds with docking scores of −10.42 and −8.97 kcal/mol, respectively, underwent further 100ns MD simulations to more rigorously examine their interaction details and stabilities in simulated physiological environment.

3.10. Molecular dynamics simulation to determine the stability of each complex

To comprehend the dynamics of each selected NLRP3-hit complex, a 100 ns MD simulation was executed. For each simulated complex, the stability was determined by calculating several parameters including RMSD, RMSF, DCCM, and RoG.

3.10.1. Root mean squares deviation (RMSD) analyses

The root means square deviation (RMSD) of the atomic positions to study the conformational fluctuations of NLRP3 in complex

Table 3
Docking analysis of ZINC compounds with NLRP3.

S. No	ZINC ID	Ligand	Receptor	Interaction	Distance	E (kcal/mol)	S-Score
1	ZINC000009601348	O 9	N ALA 228 (A)	H-acceptor	3.00	−3.2	−10.422
		O 9	CD ARG 351 (A)	H-acceptor	3.63	−0.2	
		O 9	NH1 ARG 351 (A)	H-acceptor	3.05	−0.9	
		6-ring	NH1 ARG 178 (A)	π-cation	4.07	−0.5	
		6-ring	NH2 ARG 178 (A)	π-cation	3.36	−0.3	
		6-ring	CB ALA 228 (A)	π-H	4.40	−0.5	
2	ZINC000019810420	O 9	N ALA 228 (A)	H-acceptor	3.13	−2.2	−9.9841
		O 9	CD ARG 351 (A)	H-acceptor	3.63	−0.3	
		CL 20	CD2 LEU 187 (A)	H-acceptor	3.64	−0.2	
		6-ring	CB ALA 227 (A)	π-H	4.85	−0.3	
		6-ring	CB ALA 228 (A)	π-H	4.22	−0.3	
3	ZINC000017135518	C 4	O ALA 228 (A)	H-donor	3.55	−0.1	−9.1942
		N 6	O ALA 228 (A)	H-donor	−0.7	−0.7	
		N 6	CG2 THR 524 (A)	H-acceptor	3.66	−0.1	
		O 8	CG2 VAL 414 (A)	H-acceptor	−0.1	−0.2	
		5-ring	N ALA 228 (A)	π-H	4.78	−0.1	
		6-ring	CG1 VAL 353 (A)	π-H	3.75	−0.1	
4	ZINC000003103298	N 15	OE2 GLU 527 (A)	H-donor	3.02	−2.5	−8.8764
		O 14	NH1 ARG 351 (A)	H-acceptor	3.00	−2.4	
		O 14	NH2 ARG 351 (A)	H-acceptor	3.37	−0.9	
5	ZINC000020938182	O 8	NH1 ARG 351 (A	H-acceptor	2.97	−1.9	−8.7864
		O 8	NH2 ARG 351 (A)	H-acceptor	3.41	−0.6	
		6-ring	NH1 ARG 178 (A)	π-cation	4.24	−1.3	
		6-ring	CB ALA 227 (A)	π-H	4.35	−0.3	
6	ZINC000024985355	C 3	SD MET 523 (A)	H-donor	2.96	−0.2	−8.7234
		C 21	O ALA 228 (A)	H-donor	3.30	−0.2	
		C 16	6-ring TYR 443 (A)	H-π	3.75	−0.2	
		6-ring	CB THR 524 (A)	π-H	4.15	−0.3	
7	ZINC000001376250	C 15	OE2 GLU 527 (A)	H-donor	3.31	−0.1	−8.6453
		O 11	NH1 ARG 351 (A)	H-acceptor	3.36	−0.9	
		N 24	CG2 THR 524 (A)	H-acceptor	3.52	−0.2	
		N 25	CG2 THR 524 (A)	H-acceptor	3.69	−0.2	
		6-ring	CD1 ILE 411 (A)	π-H	4.28	−0.1	
		5-ring	CG2 THR 439 (A)	π-H	3.78	−0.1	
8	ZINC000009795937	O 7	N ALA 228 (A)	H-acceptor	3.37	−0.5	−8.5632
		O 7	CD ARG 351 (A)	H-acceptor	3.51	−0.5	
		O 7	NH1 ARG 351 (A)	H-acceptor	3.12	−0.2	
		5-ring	CB ALA 227 (A)	π-H	3.53	−0.3	
		6-ring	CG2 THR 524 (A)	π-H	3.99	−0.4	
9	ZINC000033063933	N 11	OE1 GLU 527 (A)	H-donor	2.47	−0.8	−8.1431
		S 31	OE1 GLU 527 (A)	H-donor	2.92	−0.7	
		O 2	NH2 ARG 351 (A)	H-acceptor	2.85	−1.5	
		S 16	NE2 GLN 308 (A)	H-acceptor	3.79	−1.0	
		O 21	NH2 ARG 351 (A)	H-acceptor	3.24	−3.9	
10	ZINC000044168026	C 26	OE2 GLU 184 (A)	H-donor	3.70	−0.2	−8.0139
		O 14	N ALA 228 (A)	H-acceptor	3.12	−1.5	
		O 14	NH1 ARG 351 (A)	H-acceptor	3.47	−0.2	
		6-ring	CA VAL 353 (A)	π-H	4.42	−0.6	
11	Reference compound (MCC950)	O 3	NH1 ARG 351 (A)	H-acceptor	2.88	−0.8	−7.6432
		O 6	CA ALA 227 (A)	H-acceptor	3.34	−0.3	
		O 6	N ALA 228 (A)	H-acceptor	3.18	−1.9	
		5-ring	CG PRO 352 (A)	π-H	3.71	−0.5	

with hit molecules for 100ns was created to evaluate their dynamic stability. The stability of the drug-target interaction is significantly influenced by the RMSD value. A uniform RMSD plot suggests that the system's structure is in equilibrium and that there is increasing stability as the simulation duration was approaching. The NLRP3-MCC950 system's backbone RMSD graph attained equilibrium and stability at RMSD 2.0 Å and remained stable for the first 20 ns of the simulation. Nevertheless, a slight divergence in the RMSD was seen between 20 and 40 ns but the overall RMSD remained stable during the whole 100 ns of simulation as shown in Fig. 4. The NLRP3-SANC00225 complex remained stable for the first 25 ns of simulation time, after which it had a lower RMSD deviation of 3.4 Å, which eventually fell back to 2.5 Å and represented a stable RMSD graph after 30 ns (Fig. 4). Over the first 18 ns of the simulation, the NLRP3-ZINC000009601348 complex remained stable; after that, it displayed a smaller RMSD deviation of 0.25 Å and increased in stability until simulation completion. According to the RMSD interpretation, the NLRP3-ZINC000009601348 system and the NLRP3-SANC00225 system had greater intermolecular stability in terms of chemical interactions and conformation, suggesting that these two compounds may have a larger affinity for NLRP3 (Fig. 4).

Table 4
Docking analysis of South African compounds with NLRP3.

S. No	Compounds ID	Ligand	Receptor	Interaction	Distance	E (kcal/mol)	S-Score
1	SANC00225_minRM1.pdb	O 15	OE2 GLU 527 (A)	H-donor	3.03	−0.8	−8.9764
		O 19	N ALA 228 (A)	H-acceptor	3.16	−2.2	
		O 19	CD ARG 351 (A)	H-acceptor	3.65	−0.2	
		O 19	NH1 ARG 351 (A)	H-acceptor	2.99	−1.5	
		6-ring	ALA 227 (A)	π-H	3.69	−0.1	
2	SANC00529_minRM1.pdb	O 9	OE1 GLU 369 (A)	H-donor	3.01	−1.4	−8.6695
		O 12	CB ALA 227 (A)	H-acceptor	3.57	−0.2	
		O 12	CG PRO 352 (A)	H-acceptor	3.29	−0.1	
3	SANC00639_minRM1.pdb	C 1	SD MET 523 (A)	H-donor	3.72	−0.5	−8.4939
		O 24	N ALA 228 (A)	H-acceptor	3.31	−0.2	
		O 24	CB ALA 228 (A)	H-acceptor	3.11	−0.2	
4	SANC01077_minRM1.pdb	C 1	O ALA 228 (A)	H-donor	3.39	−0.1	−7.9052
		O 16	NH2 ARG 351 (A)	H-acceptor	3.45	−0.5	
		O 17	NH1 ARG 351 (A)	H-acceptor	2.99	−0.1	
		6-ring	CG1 VAL 353 (A)	π-H	4.62	−0.1	
5	SANC01125_minRM1.pdb	C 4	OE1 GLU 369 (A)	H-donor	3.48	−0.2	−7.8472
		C 22	OE1 GLU 184 (A)	H-donor	3.45	−0.1	
		O 34	N ALA 227 (A)	H-acceptor	3.04	−0.4	
6	SANC00101_minRM1.pdb	O 19	CG MET 408 (A)	H-acceptor	3.39	−0.1	−7.7239
		O 21	CB ALA 228 (A)	H-acceptor	3.37	3.37	
		6-ring	CD1 ILE 411 (A)	π-H	4.31	−0.1	
7	SANC00457_minRM1.pdb	C 23	O ALA 228 (A)	H-donor	3.83	−0.1	−7.6583
		O 13	CG1 VAL 353 (A)	H-acceptor	3.58	−0.1	
		C 27	6-ring TYR 443 (A)	H-π	4.76	−0.1	

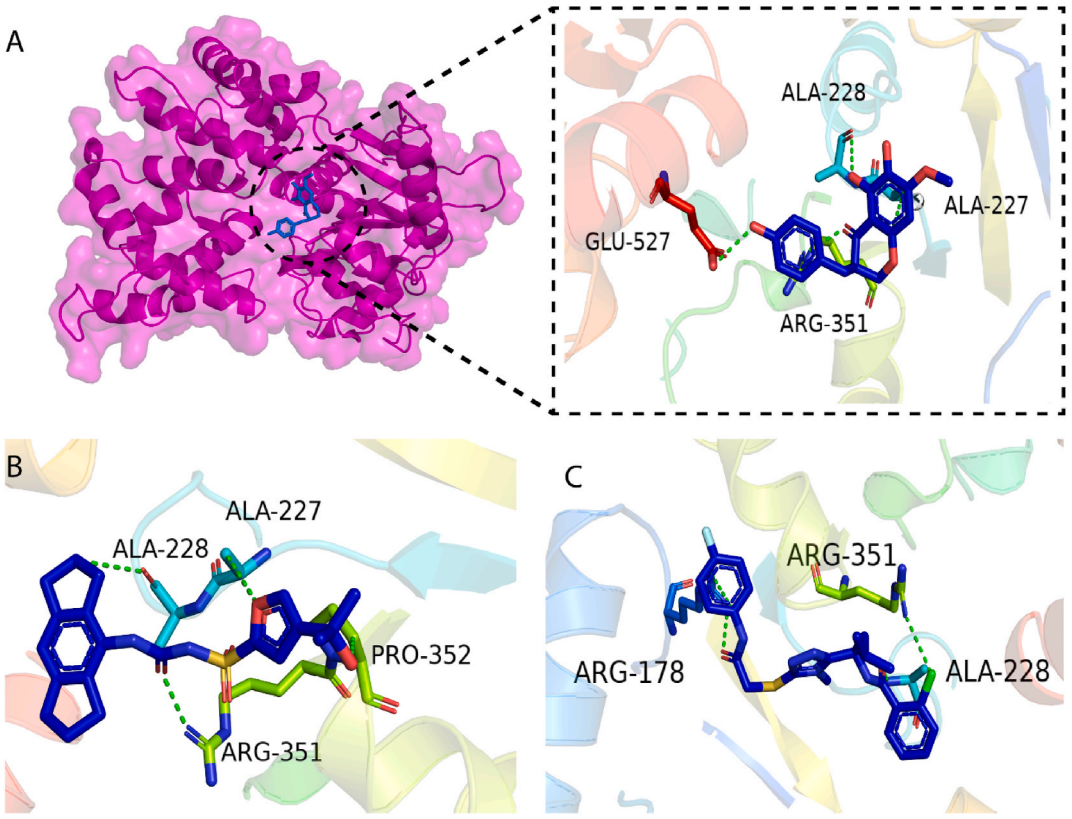

Fig. 3. Ligand interaction of hits compound (**A**) NLRP3-SANC00225, (B) NLRP3-Reference (MCC950), (C) NLRP3-ZINC000009601348. The blue colors show the ligand molecule.

Table 5
ZINC compounds Re-docking result with PyRx.

S. No	Ligands	Binding Affinity	RMSD/ub	RMSD/lb
1	ZINC000009601348_uff_E = 578.38	−6.4	2.612	1.676
2	ZINC000019810420_uff_E = 809.61	−5.8	8.051	3.899
3	ZINC000017135518_uff_E = 811.81	−6.1	6.837	3.212
4	ZINC000003103298_uff_E = 372.53	−5.5	6.034	3.619
5	ZINC000020938182_uff_E = 779.72	−4.2	1.593	1.343
6	ZINC000024985355_uff_E = 806.49	−5	2.658	1.678
7	ZINC000001376250_uff_E = 933.64	−4.2	8.572	3.268
8	ZINC000009795937_uff_E = 850.65	−5.2	2.412	1.712
9	ZINC000033063933_uff_E = 591.46	−4.2	8.768	5.036
10	ZINC000044168026_uff_E = 610.03	−4.6	8.827	3.819
11	Reference_uff_E = 1656.57	−4.5	2.886	1.852

Table 6
South African compounds Re-docking result with PyRx.

S. No	Ligands	Binding Affinity	RMSD/ub	RMSD/lb
1	SANC00225_minRM1_uff_E = 262.42	−5.4	7.45	3.017
2	SANC00529_minRM1_uff_E = 442.26	−4.6	7.356	2.359
3	SANC00639_minRM1_uff_E = 709.07	−4.9	6.628	2.496
4	SANC01077_minRM1_uff_E = 310.65	−4.7	6.425	1.489
5	SANC01125_minRM1_uff_E = 967.16	−3.7	2.144	1.551
6	SANC00101_minRM1_uff_E = 230.90	−4.4	3.214	2.301
7	SANC00457_minRM1_uff_E = 710.19	−4.2	2.215	1.727

Table 7
Pharmacokinetic properties of ZINC compounds.

S. No	ZINC ID	MW	Don	Acc	LogP	TPSA (Å)	Toxic
1	ZINC000009601348	490	2	4	5.00	88.91	No
2	ZINC000019810420	484	1	5	4,71	91.41	No
3	ZINC000017135518	498	1	3	5.31	59.06	No
4	ZINC000003103298	481	3	4	5.02	96.98	No
5	ZINC000020938182	464	1	4	5.58	84.30	No
6	ZINC000024985355	473	1	3	5.14	78.95	No
7	ZINC000001376250	497	1	4	6.02	93.26	No
8	ZINC000009795937	478	0	5	4.59	89.41	No
9	ZINC000033063933	452	2	5	4.21	93.21	No
10	ZINC000044168026	456	0	3	6.02	66.92	No

Table 8
Pharmacokinetic properties of South African compounds.

S.no	Compound ID	MW	Don	Acc	LogP	TPSA (Å)	Toxic
1	SANC00225_minRM1	316	3	6	2.01	96.22	No
2	SANC00529_minRM1	330	4	7	1.94	116.41	No
3	SANC00639_minRM1	317	2	6	0.01	71.39	No
4	SANC01077_minRM1	246	1	3	2.17	55.76	No
5	SANC01125_minRM1	486	3	5	5.50	94.83	No
6	SANC00101_minRM1	290	5	6	0.93	110.38	No
7	SANC00457_minRM1	384	3	4	1.32	77.76	No

3.10.2. RMSF analysis

Throughout the simulation, the root means square fluctuation analyses displayed per-residue fluctuations. The acquired RMSF values demonstrated that in comparison to the reference compound (MCC950), the NLRP3-ZINC000009601348 system displayed little RMSF. The reference compound's residues 1–100 showed greater fluctuations than the NLRP3-ZINC000009601348 system. However, NLRP3-SANC00225 showed higher fluctuations at residues 1–100 than the NLRP3-MCC950 system. In addition, all systems showed an average RMSF of 1.0 Å, with significant fluctuations in the 10–90, 300–320, 350–390, and 400–410 regions. As compared to our predicted inhibitors the reference compound shows higher fluctuation, but the residual flexibility pattern was comparable. Fig. 5 clearly shows the RMSF for each of the three systems. According to these findings, the chosen inhibitors' docking to the target protein stabilizes it. Hence, both correlated and non-correlated motions are impacted by ligand binding, which has a substantial impact on the

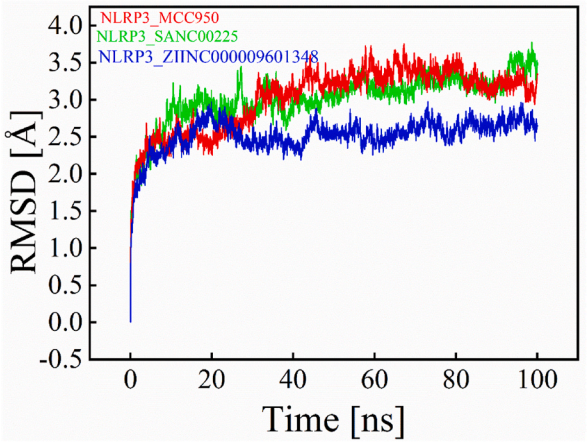

Fig. 4. RMSD of the ZINC, South African and Reference compounds. Time in ns is shown on the X-axis, while the Y-axis displays the RMSD value.

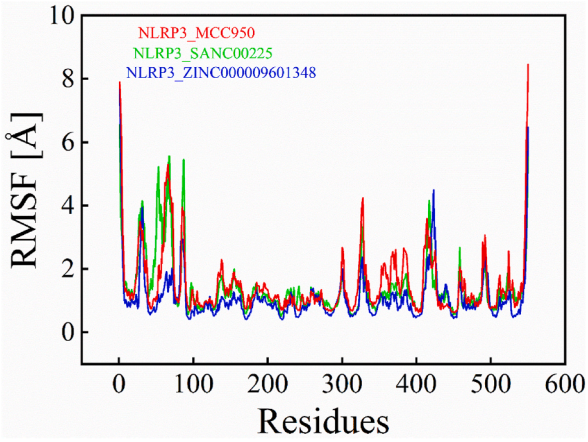

Fig. 5. RMSF of the ZINC, South African and Reference compounds. X-axis represents the total number of residues, while the Y-axis displays the RMSF value.

residue fluctuation that results from the internal residues being disturbed by the binding of various ligands.

3.10.3. Dynamic cross-correlation (DCCM) analyses

A dynamic cross-correlation matrix was developed to determine correlated or anti-correlated motions in all systems. Positive correlations show the movement of amino acid residues are in a parallel direction (same way), while negative correlations show that residues are moving in the anti-parallel (opposite direction). Based on the colors yellow, green, and blue, respectively, the predicted map results were examined. Fully correlated pairs are denoted by the yellow color, whereas anti-correlated pairs are denoted by the blue color. Yet, a greenish color highlights the moderately and uncorrelated zones. The stronger the positive correlation, the deeper the color, and vice versa (Fig. 6A-**C**).

3.10.4. Radius of gyration (RG) analyses

Radius of gyration reflects the protein system's compactness during the MD simulation. Rg of both the ligands as well as the reference compound are plotted in Fig. 7. According to the mean values, which range from 39.80 Å to 40.05 Å, no significant change in Rg values was often observed for the reference and the hit molecules, suggesting that the hit molecules remained stable during the entire simulation period.

3.10.5. Binding free energy calculation

Understanding protein-ligand binding affinity is crucial to comprehending molecular stability. One of the most used methods for calculating binding energy is MMPBSA [50]. The relative binding free energy of MCC950, ZINC000009601348, and SANC00225 with the NLRP3 NATCH domain was calculated. We used 500 frames from the simulation trajectories to determine the binding affinity of these ligands with the NLRP3 NATCH domain. Our results suggested that the binding free energy of NLRP3 in complex with

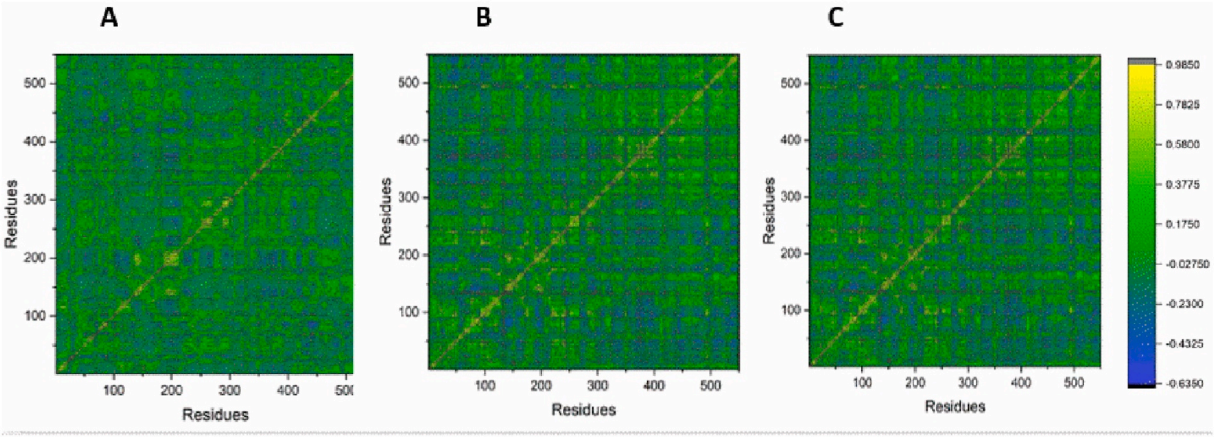

Fig. 6. Dynamical cross-correlation maps show collective motions of positive and negative correlative relationships between the residues in (A) Reference compound (MCC950) (B) SANC00225 (C) ZINC000009601348.

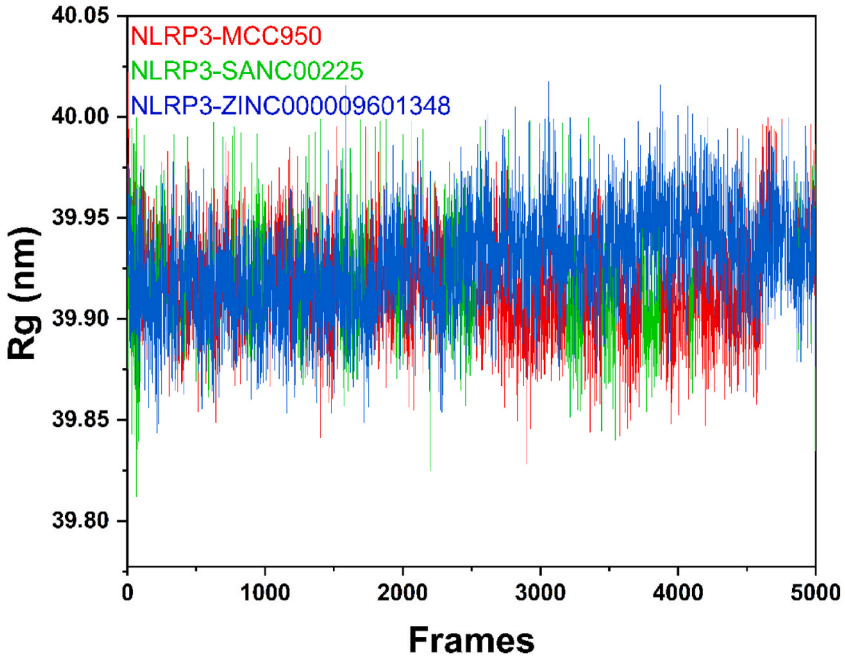

Fig. 7. The Rg values of all the studied complexes. The Rg of the Reference compound is shown in Red, complex 2 (SANC00225) is shown in Green and complex 3 (ZINC000009601348) are shown by Blue color.

ZINC000009601348 was higher than that of SANC00225 and MCC950 (Table 9). The SANC00225 also showed higher binding free energy of −30 kJ/mol than the reference compound (−27 kJ/mol). These data showed that both the predicted compounds had a stronger binding affinity for NLRP3 inflammasome.

Table 9
MMPBSA analysis.

Complex	VDW	EEL	EPB	ENPOLAR	EDISPER	ΔPBSA
Reference compound (MCC950)	−27.9643	−4.9479	14.5772	−16.7174	31.5368	−3.5156
SANC00225	−30.4509	−20.8062	33.1200	−19.8752	32.5442	−5.4682
ZINC000009601348	−31.3798	0.3502	8.7654	−19.4179	34.5251	−7.1571

4. Conclusion

The NLRP3 inflammasome is an important target in the field of drug design due to its role in the development and progression of epilepsy as well as in the progression of a variety of neurodegenerative diseases including Parkinson's disease and Alzheimer's disease. Hence, it is crucial for managing epilepsy and lowering seizures to negatively regulate the activity of the NLRP3 inflammasome. In this research, we used a machine learning (ML)-based virtual screening (VS) approach to find new potential NLRP3 inhibitors. Models are created using k-NN, SVM and RF machine learning algorithms. It was found that models built using 2D descriptor based on RF algorithm give better results compared to the other models and hence used for further screening of large chemical libraries. In addition to screening from the ZINC database, we also performed the same task using the South African database. We further narrowed the candidate range using a molecular docking approach. Lastly, using molecular dynamic simulation, we confirmed the stability of the complexes once more. Both the target protein complexes and the final candidate ligands displayed stable binding over the course of the simulation. The proposed method indicates that machine learning-based virtual screening has application prospects. The application of machine learning in drug screening relies on molecular interaction relationships and provides improvements and complements to traditional drug screening approaches. Further in-vitro and in-vivo studies are recommended to evaluate the inhibitory potential of the hits identified in this study.

Funding

This research was funded by the Deputyship of Graduate Studies and Scientific Research Jazan University, Saudi Arabia (Project number: RG24-L05).

Additional information

No additional information is available for this paper.

Availability of data and materials

Dr. Abdul Wadood should be contacted for data and material or other information.

Data availability statement

No data was used for the research described in the article.

CRediT authorship contribution statement

Maryam Zulfat: Writing – original draft, Methodology. **Mohammed Ageeli Hakami:** Visualization, Methodology, Investigation. **Ali Hazazi:** Validation, Formal analysis. **Arif Mahmood:** Investigation, Data curation. **Asaad Khalid:** Software, Investigation, Funding acquisition, Conceptualization. **Roaya S. Alqurashi:** Software, Resources. **Ashraf N. Abdalla:** Writing – review & editing, Methodology. **Junjian Hu:** Writing – review & editing, Methodology. **Abdul Wadood:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Xiaoyun Huang:** Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors gratefully acknowledge the funding of the Deputyship of Graduate Studies and Scientific Research Jazan University, Saudi Arabia through the project number: RG24-L05. The auhors also acknowledge the support of Guangdong Basic and Applied Research Foundation (2021B1515140026).

References

[1] M.E. Edye, et al., Epilepsy and the inflammasome: targeting inflammation as a novel therapeutic strategy for seizure disorders, Inflammasome 1 (1) (2014) 36–43.
[2] V.L.L. Lee, M.F. Shaikh, Inflammation: Cause or Consequence of Epilepsy. Epilepsy-Advances in Diagnosis and Therapy, 2019.
[3] I.E. Scheffer, et al., ILAE classification of the epilepsies: position paper of the ILAE commission for classification and terminology, Epilepsia 58 (4) (2017) 512–521.
[4] D.O.V.A. O'Brien, N. Tj Jette, I.E. Scheffer, M. de Curtis, et al., Epilepsy. Nat. Rev. Dis. Primers 4 (1) (2018) 18024.
[5] A. Rana, et al., No gains in long-term survival after liver transplantation over the past three decades, Ann. Surg. 269 (1) (2019) 20–27.
[6] C.-C. Tan, et al., NLRP1 inflammasome is activated in patients with medial temporal lobe epilepsy and contributes to neuronal pyroptosis in amygdala kindling-induced rat model, J. Neuroinflammation 12 (1) (2015) 1–12.

[7] J. Mahendra, et al., The expression of allele changes in NLRP3 (rs35829419) and IL-1β (+ 3954) gene polymorphisms in periodontitis and coronary artery disease, Materials 14 (17) (2021) 5103.
[8] M. Haneklaus, L.A. O'Neill, R.C. Coll, Modulatory mechanisms controlling the NLRP3 inflammasome in inflammation: recent developments, Curr. Opin. Immunol. 25 (1) (2013) 40–45.
[9] X.-F. Meng, et al., Inhibition of the NLRP3 inflammasome provides neuroprotection in rats following amygdala kindling-induced status epilepticus, J. Neuroinflammation 11 (1) (2014) 1–12.
[10] H. Zhang, et al., NLRP3 inflammasome activation enhances ADK expression to accelerate epilepsy in mice, Neurochem. Res. (2022) 1–10.
[11] L. Gong, et al., LncRNA ZNF883-mediated NLRP3 inflammasome activation and epilepsy development involve USP47 upregulation, Mol. Neurobiol. 59 (8) (2022) 5207–5221.
[12] L. Palumbo, et al., The NLRP3 inflammasome in neurodegenerative disorders: insights from epileptic models, Biomedicines 11 (10) (2023) 2825.
[13] S.L. Cassel, F.S. Sutterwala, Sterile inflammatory responses mediated by the NLRP3 inflammasome, Eur. J. Immunol. 40 (3) (2010) 607–611.
[14] F.G. Bauernfeind, et al., Cutting edge: NF-κB activating pattern recognition and cytokine receptors license NLRP3 inflammasome activation by regulating NLRP3 expression, J. Immunol. 183 (2) (2009) 787–791.
[15] Y.A. Kaabi, The NLRP3 inflammasome rs35829419 C> A polymorphism is associated with type 2 diabetes mellitus in Saudi Arabia, Saudi Med. J. 44 (8) (2023) 745.
[16] J. Mahendra, et al., Genetic polymorphisms of NLRP3 (rs4612666) and CARD8 (rs2043211) in periodontitis and cardiovascular diseases, Biology 10 (7) (2021) 592.
[17] H.-D. Liu, et al., Expression of the NLRP3 inflammasome in cerebral cortex after traumatic brain injury in a rat model, Neurochem. Res. 38 (2013) 2072–2083.
[18] R.M. Ransohoff, M.A. Brown, Innate immunity in the central nervous system, J. Clin. Invest. 122 (4) (2012) 1164–1171.
[19] S. Jha, et al., The inflammasome sensor, NLRP3, regulates CNS inflammation and demyelination via caspase-1 and interleukin-18, J. Neurosci. 30 (47) (2010) 15811–15820.
[20] K. Rajamäki, et al., Extracellular acidosis is a novel danger signal alerting innate immunity via the NLRP3 inflammasome, J. Biol. Chem. 288 (19) (2013) 13410–13419.
[21] D.Y.-W. Fann, et al., Pathogenesis of acute stroke and the role of inflammasomes, Ageing Res. Rev. 12 (4) (2013) 941–966.
[22] S. Kate, T. Jurg, The inflammasomes, Cell 140 (6) (2010) 821–832.
[23] C. Dekker, et al., Crystal structure of NLRP3 NACHT domain with an inhibitor defines mechanism of inflammasome inhibition, J. Mol. Biol. 433 (24) (2021) 167309.
[24] V. Patel, M. Shah, Artificial intelligence and machine learning in drug discovery and development, Intelligent Medicine 2 (3) (2022) 134–140.
[25] A. Wadood, et al., Machine learning-based virtual screening for STAT3 anticancer drug target, Curr. Pharmaceut. Des. 28 (36) (2022) 3023–3032.
[26] M. Tahir ul Qamar, et al., Target-specific machine learning scoring function improved structure-based virtual screening performance for SARS-CoV-2 drugs development, Int. J. Mol. Sci. 23 (19) (2022) 11003.
[27] S. Nakapraves, et al., Prediction of mefenamic acid crystal shape by random forest classification, Pharmaceut. Res. 39 (12) (2022) 3099–3111.
[28] J. Bac, et al., Scikit-dimension: a python package for intrinsic dimension estimation, Entropy 23 (10) (2021) 1368.
[29] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[30] Y. Liu, Y. Wang, J. Zhang, New machine learning algorithm: random forest, in: Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, Springer, 2012. September 14-16, 2012. Proceedings 3.
[31] K.-Z. Huang, et al., Machine learning: modeling data locally and globally, Springer Science & Business Media, 2008.
[32] H. Franco-Lopez, A.R. Ek, M.E. Bauer, Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method, Rem. Sens. Environ. 77 (3) (2001) 251–274.
[33] Y. Cui, M. Cai, H.E. Stanley, Comparative analysis and classification of cassette exons and constitutive exons, BioMed Res. Int. (2017) 2017.
[34] S. Vilar, G. Cozza, S. Moro, Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery, Curr. Top. Med. Chem. 8 (18) (2008) 1555–1572.
[35] A. Ajala, et al., Virtual screening, molecular docking simulation and ADMET prediction of some selected natural products as potential inhibitors of NLRP3 inflammasomes as drug candidates for Alzheimer disease, Biocatal. Agric. Biotechnol. 48 (2023) 102615.
[36] M. Uzzaman, T. Mahmud, Structural modification of aspirin to design a new potential cyclooxygenase (COX-2) inhibitors, Silico Pharmacology 8 (2020) 1–14.
[37] R. Salomon-Ferrer, D.A. Case, R.C. Walker, An overview of the Amber biomolecular simulation package, Wiley Interdiscip. Rev. Comput. Mol. Sci. 3 (2) (2013) 198–210.
[38] M. Junaid, et al., Structural-dynamic insights into the H. pylori cytotoxin-associated gene A (CagA) and its abrogation to interact with the tumor suppressor protein ASPP2 using decoy peptides, J. Biomol. Struct. Dyn. 37 (2018) 4035–4050.
[39] R.D. Skeel, D.J. Hardy, J.C. Phillips, Correcting mesh-based force calculations to conserve both energy and momentum in molecular dynamics simulations, J. Comput. Phys. 225 (1) (2007) 1.
[40] J.-P. Ryckaert, G. Ciccotti, H.J. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (3) (1977) 327–341.
[41] R. Salomon-Ferrer, et al., Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald, J. Chem. Theor. Comput. 9 (9) (2013) 3878–3888.
[42] D.R. Roe, T.E. Cheatham III, PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data, J. Chem. Theor. Comput. 9 (7) (2013) 3084–3095.
[43] M. Piñon-Espitia, et al., Electronic structure comparison of Cu 2p and O 1s X-Ray photoelectron spectra for CuxO nanofibers (x= 1, 2,i), Mater. Chem. Phys. 272 (2021) 124981.
[44] E. Gawehn, J.A. Hiss, G. Schneider, Deep learning in drug discovery, Molecular informatics 35 (1) (2016) 3–14.
[45] J. Zhu, et al., Integrating machine learning-based virtual screening with multiple protein structures and bio-assay evaluation for discovery of novel GSK3β inhibitors, Front. Pharmacol. 11 (2020) 566058.
[46] T.B. Kimber, Y. Chen, A. Volkamer, Deep learning in virtual screening: recent applications and developments, Int. J. Mol. Sci. 22 (9) (2021) 4435.
[47] A.S. Rifaioglu, et al., Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases, Briefings Bioinf. 20 (5) (2019) 1878–1912.
[48] S.R. Martinez, M.S. Gay, L. Bddcs, The rule of 5 and drugability. Adv. Drug deliv, Rev. E. 176 (2016) 139–148.
[49] W.L. DeLano, Pymol: an open-source molecular graphics tool. CCP4 Newsl, Protein Crystallogr 40 (1) (2002) 82–92.
[50] S. Uddin, et al., Comparing different supervised machine learning algorithms for disease prediction, BMC Med. Inf. Decis. Making 19 (1) (2019) 1–16.