Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

Identifyingnovel Leucine-Rich Repeat Serine/Threonine-Protein Kinase 2 (LRRK2) Inhibitors Usingmolecular Modelingapproaches.

Rengarajan Kavitha[1*],Swaminathan Sivagami[1], Chandrasekaran Meganathan[2],Rajendran Nithyabalaji[3],Rajendran sri balan[4],
1.Department of Physics, New Prince Shri Bhavani College Of Engineering and Technology,Gowriwakkam, Chennai, Tamil Nadu 600073.
1.Department of Chemistry, New Prince Shri Bhavani College Of Engineering and Technology,Gowriwakkam, Chennai, Tamil Nadu 600073.
2.Department of Physics, Central Institute of Petrochemicals Engineering & Technology (CIPET), Guindy, Chennai, Tamil Nadu 600032.
3.Biochemie Innovations Lab, Tindivanam, Tamil Nadu, India
4.Department of Physics, Valliammai Engineering College, Kattankulathur, Kanchepuram Dt., Tamil Nadu, India

***Corresponding author**

E-mail:kavisri.try@gmail.com
Phone no: 8754455632

Abstract

The etiology of the Parkinson's disease (PD) is one of the world's most common age-related neurodegenerative disorders. The leucine-rich repeat kinase 2 (LRRK2)is an important target in designing drugs against PD.The R1441C mutation in the Ras-like GTPase domain (ROC) of LRRK2 havereduced GTPase activity.It changes the strength of ROC whichinturn enhanced kinase activity and caused PD.Hence, we target ROC of LRRK2 as a therapeutics strategy to mitigate brain damage in PD patients. Thus, we performed an atom based3D-QSAR studieson LRRK2 inhibitors, using a series of derivatives such as 4-alkylamino-7-aryl-3-cyanoquinoline,cinnoline-3-carboxamides, and triazolopyridazine. The statistical parameters generated by 3D-QSAR model revealed the strength of predictabilitywith the highest score.The reliability of the generated model was confirmed by internal and external validation parameters.The validated hypothesis was usedfor performing virtual screening for extraction of the potential leads from commercial databases (Maybridge, NCI, and ZINC).Further, the pharmacokinetic properties (Lipinski's, ADME) wereanalyzed for the selected hits to make more drug-likeness.Themost suitable binding orientation of selected hit compounds wasanalyzed by the docking studies. Finally,four hits were selected on the basis of Glide score, GOLD fitness score and hydrogen bonding interactions with critical residues.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

KEYWORDS:LRRK2, 3D-QSAR, Virtual Screening, Molecular docking

List of Abbreviations

3D-QSAR	-	3 Dimensional Quantitative Structure–Activity Relationship
ADME	-	Absorption, Distribution, Metabolism and Excretion
BBB	-	Blood Brain Barrier
CADD	-	Computer-Aided Drug Design
channels		
EF	-	Enrichment Factor
GH	-	Goodness of Hit
GOLD	-	Genetic Optimization for Ligand Docking
HBA	-	Hydrogen Bond Acceptor
HBD	-	Hydrogen Bond Donor
HTS	-	High Throughput Screening
HY	-	Hydrophobic

1. Introduction

A Parkinson's disease (PD), which was first defined in "An Essay on the Shaking Palsy" in 1817 by a London-based physician,James Parkinson, has probably existed for thousands of years[1]. With the aging of individuals, the importance of PD as public health issue is expected to increase and the number of persons worldwide will double in 25 years.Abnormal brain activity occurs due to the loss of dopaminergic neurons in the substantianigra leading to PDdisease. Thesymptoms of PD are a chronic progressive characterized by rigidity, bradykinesia, tremor and postural instability [2].Recently mutations of the gene coding forhuman leucine-rich repeat kinase 2 (LRRK2) werefound to be the prevalent genetic cause of PD.LRRK2 proteins are involved in many cellular processes such as regulation of cell polarity, chemotaxis, cytokinesis, cytoskeletal arrangements and programmed cell death.

The multidomain protein of LRRK2 belongs to the ROCO class of proteins.Itcontains2527 amino acids [3]. The N-terminal consistsof armadillo repeats domain (ARM),ankyrin repeats domain (ANK),leucine rich repeats (LRR) domain andRas of complex (ROC) domain.The C-terminal is characterized as ROC (COR) domain, the kinase domain and WD40 repeats domain.

The ROC of LRRK2 forms a strong dimer and acts as a GTPase to regulate kinase activity [3].The COR domain helps to conveying the signal from ROC domain to the kinase domain through a GTP/GDP bound cycle [3].The one residue of ROC domain contains multiplePD mutations (R1441C, R1441G, and R1441H) whicharelocatednear the central region of the protein. These mutations changethe ROC domain stability by increasing its binding affinity

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

for GTP, andat the same time, lowering its GTPase activity than wild-type LRRK2 [3].Several studies have shown the occurrence of GTP hydrolysis occurs as unclear.

The multiple mutations in LRRK2 signifydetermination of the activation process of the protein and its misregulation in PD. The results of in vitro studies suggest the significant role played by the guanine nucleotide-binding in the regulation of LRRK2 kinase activity. Interactions between the kinase domain and ROC domain critical to the neurotoxic action of LRRK2. In general, LRRK2 mutations play a role in cancer, crohn's disease, and leprosy are significant.

In our research , 3D-QSAR based pharmacophore hypothesis was used to find the essential features which are responsible for the biological activity of LRRK2 inhibitors.The predictability and reliability of the generated atom based 3D-QSAR model was evaluated by internal and external validation metrics. The well-validated metric was used as 3D-query to retrieve the potential lead against LRRK2. The retrieved leads were filtered by pharmacokinetic properties such as Lipinski's rule of five and ADME to make them more drug-likeness. The hit compounds were then docked with LRRK2 for analysing the intermolecular interactions and binding mode between the ligands and active site residues.

2. Materials and methods

2.1 Ligand preparation

In this study, a data set of 40 compoundssuch as, 4-alkylamino-7-aryl-3-cyanoquinoline,cinnoline-3-carboxamides, and triazolopyridazinederivatives were collected from literature [4-5]. The inhibitory concentrations (IC_{50}) of all compounds were converted into pIC_{50} value using the formula (pIC_{50} = -log IC_{50}). The chemical structures of all the molecules were drawn by using ACD/ChemSketch11.0. The selected compounds were taken to a Ligprep module of Schrodinger software for preparing the ligands using "OPLS_2005" force field [6]. The maximum number of conformers were generated for each structure and minimization of each conformers was filtered through relative energy difference of 10Kcal mol^{-1} and RMSD of 1.00Å.

2.2 Phase methodology

Ligands were imported in *PHASE module* [6] for generating an atom 3D-QSAR model. Based on the pIC_{50} values, the compounds were divided into active, inactive, and moderately active. The *PHASE module* provides default common pharmacophore features such as acceptor (A), donor (D), hydrophobic (H), negative ionizablecentres (N), positive ionizablecentres (P), and aromatic ring (R) for determining the chemical structural pattern of ligands. The common pharmacophore hypotheses (CPH) was identified from the variant list through use of the create site option. The active ligands of CPH were scored by setting the root mean square deviation (RMSD) value below 1.0 and the vector score at 0.5. The best hypothesis has been produced through application of default scoring hypothesis on the generation of CPH [6-7]. The best model was calculated by a survival score, defined as

S = WsiteSsite+WvecSvec+WvolSvol + WselSsel+Wmrew(1)

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

In equation (1) W- weights, S - scores, Ssite - alignment score, Svec - vector score, Svol - volume score. The Ssel - selectivity score. The Wsite, Wvec, Wvol, and Wrew had default values of 1.0, Wsel had a default value of 0.0. The Wmrew denoted reward weights was defined by m^{-1}, where m was the number of actives that match the hypothesis.

2.3 3D-QSAR generation

The best hypothesis was chosen by the investigation of the scores and alignment of the active ligands to the generated hypothesis. An atom based 3D-QSAR model was developed for the selected hypothesis by dividing the dataset into 30 compounds of the training set (70%) (Fig. 1) and 10 compounds of the test set (30%) (Fig. 2)in a random manner. Keepingof 70% molecules gives rise to binary-valued occupation patterns that can be used as independent variables for creating a partial least-squares (PLS) QSAR model. The atom-based QSAR is one in which each molecule is treated as the set of overlapping van der Waal's spheres. Then, each atom of spheres was placed into one of six categories such as hydrogen bond donors (D); hydrophobic/non-polar (H); negative ionic (N); positive ionic (P); electron-withdrawing (W); and all other types of atoms were categorized as miscellaneous (X). Regression analysis was made on the hypothesis with increasing partial least square (PLS) for generatingthe 3D-QSAR model.

2.4 Pharmacophore validation

The main aim was to develop the best 3D-QSAR model for accurate prediction of biological activity of new compounds which were statistically robust both internally and externally. The internal validation of the best model was done by R^2 (squared correlation coefficient), Q^2 (R^2 for test set), SD (standard deviation), Pearson's R (Pearson's correlation coefficient), P (statistical significance) and F (variance ratio). But, these parameters are not necessary to determine the robustness of a model.

Further, the external validation for the best model was performed by R^2_{pred} using the formula

$$R^2_{pred} = 1 - \frac{\sum(Y_{pred\ (test)} - Y_{test})^2}{\sum(Y_{(test)} - \overline{Y}_{training})^2} \quad (2)$$

Using equation (2) $Y_{pred\ (test)}$ and Y_{test} refer to predicted and observed activity values for test set compounds. $\overline{Y_{training}}$ Refer to mean activity value of training set compounds. The value of R^2_{pred} is mainly dependent on the sum of squared differences between observed values of test set compounds and average observed activity values of training set compounds. This difference will create, the increased value of R^2_{pred} which predicts the activity value of the test-set compounds [8].

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

For establishing the proximity between the experimental and the predicted activities, data of the test set compounds was further determined by r_m^2 metrics such as $r^2_{m(test)}$ and $\Delta r^2_{m(test)}$, $r^2_{m(overall)}$ and $\Delta r^2_{m(overall)}$. The $r^2_{m(test)}$ value is calculated bythe formula as [8].

$$r^2_{m(test)} = *(1 - \sqrt{r^2 - {r_0}^2}) \quad (3)$$

In equation (3) r^2 and r_0^2 are the determination coefficients for the least-square regression line correlating between the observed and the predicted values with intercept (set to zero) and without intercept for the test set compound. The overall validation would be the developed models in terms of both internal and external predictive ability was calculated by $r^2_{m(overall)}$ statistical parameters. The site HTTP //aptsoftware.co.in/rmsquare/ was used for calculation of$r^2_{m(test)}$and $\Delta\ r^2_{m(test)}$for external validation, $r^2_{m(overall)}$and $\Delta\ r^2_{m(overall)}$for overall validation in both scaled and unscaled versions.
The $r^2_{m(rank)}$) metric is calculated on the basis of the acceptability of a quantitative model in order to determine the rank-order predictions that could quantify the model stability based on a comparison between the observed and predicted response ranking by using the formula [9].

$$r^2_{m(rank)} = r^2 * (1 - \sqrt{r^2 - {r_0}^2}) \quad (4)$$

Where, in equation (4), $r^2_{(rank)}$ and $r_0^2{}_{(rank)}$ is the determination coefficient for the least squares regression line correlating (with and without intercept)the experimental and the predicted rank values for the test set compounds.

2.5 Virtual screening and ADME analysis

The virtual screening methodwas used for finding the potential and the novel leads for LRRK2. The selected hypothesis was implemented as a 3D-query for searching for active molecules from Maybridge (50,000), NCI (238,819) and ZINC (35,000) databases using the search for matches' option in *phase module* (Schrodingersuite 2015).To obtain the best 3D similarity search, we have used limitations that included the maximum of 0.7 RMSD, obeying 10 rotatable bond cutoffs and molecular weight range of 180–500 Dalton [10]. After screening, the hit molecule which fits well with the pharmacophoric features of the best hypothesis was retrieved as a hit compound. Further,drug-like predictions for the retrieved hits were carried out by using *Ligand filtering*and *Qikprop module*for Lipinski's rule of fiveand ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties using in Maestro software[11].

3. Molecular docking

3.1 Glide Docking

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

Molecular docking was carried out using *Glide module* in the Schrödinger suite 2015 for analysing the detailed intermolecular interactions between the virtual database hits and LRRK2. The crystal structure of LRRK2 was downloaded from the protein data bank (PDB code: 2ZEJ), having the resolution of 2.0 Å [3]. The structure was imported into the protein preparation wizard. The protein structure was prepared through removal of water molecules,addition of hydrogen atoms, creation of zero-orders bonds to metals, and convertion selenomethionines to methionines. The minimization of protein was done by OPLS 2005 force field. After preparation of protein, the protein active site was defined through the use of default parameters of receptor grid generation. Hence, we carried out docking calculations from the database hits using Glide module by the step-wise process [12-14]. The Glide protocols were run using the default parameters. In the initial step, the database hits were sent to thehigh throughput virtual screening (HTVS) mode. In the next step, the highest scoring ligands from HTVS were taken to the Glide standard precision mode(SP). Then the top-scoring ligands from Glide SP were retained and then docked into Glide Extra Precision (XP). The results of the best docking poses were quantified based on on the basis of Glide score.

3.2 GOLD docking

Molecular docking studies were performed on Genetic Optimization for Ligand Docking (GOLD) [15]for further evaluation of the molecular interaction of the retrieved hits. In the preparation of protein, all the water molecules were deleted, hydrogen atoms were added, and bond orders for crystal protein were adjusted. The active site of ROC domain of LRRK2 was defined within a 10 Å radius around the ligand presents in the crystal structure.The Guanosine-5'-Diphosphate and Magnesium ion (reference compound) was docked into the active site of ROC domain to check the ability the selected parameters to yield the most suitable binding orientation. During docking process,ten poses were generated for each ligand and ranked by the GOLD fitness score [16]. Further, the hit compounds were docked into ligand binding site in the same way asthe reference compound. The final hits were selected on the basis of thehighest GOLD fitness score and binding affinity of the active site of ROC.

4 Result and discussion

4.1 Generation of pharmacophore model

In our present study, we have considered 40 compounds for the generation of the atom based 3D-QSAR model.Of the 40 compounds, eighteen moleculeswere active ($pIC_{50} \geq 6.924$), ten molecules inactive ($pIC_{50} \leq 6.848$) andthe rest of the molecules moderately active. The CPH were generated with 40 different combinations of variants. All the 40 CPH were analyzed based on the survival scoring function. The function of the scoring method helpedthe identification and ranking of the entire hypothesis.One ofthe 40 CPH i.e.ADRRR.42 shows a very good survival score. Hence, we consideredADRRR.42 for generatingthe atom based 3D-QSAR model (Table 1).In order to analyze the statistical significance and predictive power, the regression analysis was carried out on ADRRR.42

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

with increasing PLS factor up to 6. Further, increase in the number of PLS factorsdid not improve the predictability of 3D-QSAR model [10]. The features represented by best ADRRR.42 hypothesis are one acceptor, one donor and three aromatic rings (Fig .3a). The distance details (Fig. 3b) of ADRRR.42 are shown in Table 2.

The compound activity mainly depends on the fitness score.The prediction of the larger activity was represented by the highest fitness score. The distance between the features of the ligand to the centroid of the hypothesis was calculated by the fitness scoring function by superimposing each ligand on ADRRR.42. It also checks whether the features are mapped or not, hence ADRRR.42 yielded the best alignment of active ligands (Fig.3c) andinactive ligands (Fig. 3d).

4.2 Pharmacophore validation

The predictive power of the ADRRR.42 hypothesis was validated by the statistical parameters (Table 3). The results of the ADRRR.42hypothesis has shown the highest R^2 value of 0.9509.This was also regressed against the experimental versus predicted LRRK2 inhibitory activities of the training set molecules (Fig.4a) (Table 4). The reliable 3D-QSAR model was found to be significant whenthe high value of predictive squared correlation coefficient should exceeded 0.60 [17]. Hence, we have obtained the highest Q^2 value of 0.8729 for the test set prediction (Table 5). The correlation plot ofexperimental versus predicted activity of the test set molecules showed good predictability of AARRR.42 hypothesis (Fig.4b). Further, the large value of Fisher ratio (F=138) clearly indicated a statistically regression model. The minimum value of variance ratio (P= 1.435E-25), indicates the high degree of confidence. Apart from this, the highest values of Pearson-R (0.844) and lowest values of standard deviation (0.0772), root mean square error (RMSE= 0.2458) creates the consequent generation of the QSAR analysis.

The robustness and predictability of ADRRR.42 were further assessed by statistical parameters such as predictive correlation coefficient (R^2_{pred}), r^2_m metrics (Table 6). The predictive ability of the ADRRR.42 was performed on test set molecules and the value of R^2_{pred} was found to be 0.967 which has shown as above the threshold value of 0.5 [8]. Close correlation between predicted and observed activity value is not a satisfactory indication.

The quality of ADRRR.42 hypothesis was further examined by traditional r^2_m metrics such as $\overline{r^2_{m(test)}}$, $\Delta r^2_{m(test)}$, for external validation $\overline{r^2_{m(overall)}}$& $\Delta r^2_{m(overall)}$ for overall validation (both scaled and unscaled) version. The value of $\overline{r^2_m}$ should be > 0.5 & Δr^2_m should be < 0.2 [18] for an acceptable QSAR model. Hence we have obtained the results for test set $\overline{r^2_{m(test)}} = 0.8682$& $\Delta\ r^2_{m(test)}$=0.0589 (scaled) and $\overline{r^2_{m(test)}} = 0.8785$&$\Delta r^2_{m(test)}$= 0.0664 (unscaled). The overall validation indicates the predicted activity values of both the training and test sets can be calculated as $\overline{r^2_{m(overall)}}$= 0.8870&$\Delta r^2_{m(overall)}$= 0.065 (scaled) and $\overline{r^2_{m(overall)}}$ = 0.8918 & $\Delta r^2_{m(overall)}$= 0.06793 (unscaled). Hence, the obtained results for overall validation specifies r^2_m metrics as exceeding the minimum cutoff value of 0.5 [19].

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

Further, our aim to incorporate rank-order predictions ($r^2_{m(rank)}$) could qualify the stability of the selected 3D-QSAR model. The threshold value of $r^2_{m(rank)}$ should be greater than 0.5 [9].Hence, in our study, we have obtained the correlation of the ranks between the observed and the predicted data as 0.7621. The obtained results both from internal and external source for all the qualitative validation metrics were within the acceptable range. The predictive efficacy of the developed QSARmodel was also significant.

4.3 Virtual screening and ADME analysis

The ADRRR.42 was applied as a 3Dquery to retrieve molecules from Maybridge (50,000), NCI (23 8819) and ZINC (35,000) databases for identifying novel scaffolds of LRRK2 inhibitors.All queries were made using the *phase find matches* search method in the phase module of Schrodingersuite. We applied maximum RMSD 0.7, cutoff limit of 10 rotatable bonds, and molecular weight 180–500 Dalton rangeandto fit with all the chemical features of the ADRRR.42 hypothesis. This was done for preventing incorrect predictions of the hits from virtual screening. Totalsof 18400hits fromMaybridge (6000 compounds), NCI (5400 compounds), and ZINC databases (7000 compounds) (Fig. 5) were retrieved respectively, based on above criteria. The selected hitswere further filtered by applying the Lipinski's rule of five. For predicting the drug-likeness, Lipinski's rule of fivestatesthat the molecular weights were < 500 Daltons, no more than 5 hydrogen bond donors, no more than 10 hydrogen bond acceptors, an octanol-water partition coefficient log P not greater than 5.Hence, there were, 5775 hits, shown drug-like properties based on the rule of five.

Further, the pharmacokinetics of hit molecules was analyzed by ADME parameters using *Qikprop* module of Schrödinger software. The Qikprop properties were such as, water partition co-efficient (QPlogPo/w), 1.828 to 3.719; water solubility (QPlogS), -5.244 to -2.47; IC_{50} value for blockage of HERG K+ channels (QPlogHERG) -2.478 to -1.895; Caco-2 cell permeability in nm/s (QPPCaco), 96.479 to 565.57 ; brain/blood partition co-efficient (QPlogBB), -2.54 to- 1.32; percentage human oral absorption, 79–98 %, (Table 7) [26]. The ADME properties were calculated for the hit molecules within a suitable range.Thishas shown good oral absorption. Hence, we obtained 5567 hit compounds for selection based onQikprop properties.

4.4 Molecular docking

The crystal structure of ROC domain in LRRK2 (2EZJ) was obtained from protein data bank. Further, the co-crystal GDP was extracted and redocked into the binding site of ROC domain for validating the docking reliability. We obtained theroot mean square deviation (RMSD) value as0.93Å (Fig. 6)indicates the reproducing ability of Glide program.

In the first step, thedatabase hit molecules were subjected to Glide HTVS. Almost all the screened compounds (5567)were retained and passed on to the second stage of SP docking. The resultant compounds of 2220 obtained from SP were then docked on to the third stage

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

of the XP mode(Fig. 5).The hit compounds (90) obtainedfrom XP wereranked as thebest compoundsand the important interactions were analyzed by XP Glide score and XP visualizer.The resultant SP Hits were further evaluated by the GOLD program for confirming the potency and binding mode of ligands with ROC domain of LRRK2.Hence, the four hits (ZINC00959174, ZINC02136788, ZINC38545694, ZINC02138190) were selected based on Glide score and GOLD fitness score greater than reference compound (-13.95)for furtheranalysis (Table 8).

4.4.1 Docking analysis of hit compounds

The first hit compound ZINC00959174 has shown a Glide score of -13.46 and a GOLD score 56.76. The carboxyl group of ZINC00959174 has shown hydrogen bonding interactions with the amide group of amino acid residue Gly1346, and Lys1347 (Fig. 7a&8a). The donor (D7) features of our hypothesis are clearly mapped in that region (Fig. 9a). In addition, the hydrogen bonding interactions observed between the oxide ion of hit compound and Thr1348 as well as Mg1. The ring aromatic feature (R10) of the hypothesis mapped on phenyl ring has shown hydrogen bonding and π-π stacking interactions with His1453 and as well as hydrophobic interactions observed in Ala1490.

The second hit ZINC02136788 has gained hydrogen bonding between the hydroxyl group of phenol ring and amide group of Ala1367 (Fig. 7b&8b). The acceptor features (A4) of thehypothesis accepts the hydrogen atom from Ala1367. The oxypropanoylamino group of the compound mapped on donor feature (D7) of Hypothesis formed hydrogen bonding interaction with Mg1, and Lys1347 as well as Gly1346 (Fig. 9b). The hydrogen bonding interactions observed between the carboxyl group of compound ZINC02136788 and the side chain of Thr1349. The hit compound showed methyl chromen formed π-π stacking and hydrogen bonding interactions with His1453 as well as hydrophobic interactions with Ala1490. The compound has scored a glide score of -13.36 and GOLD fitness score of 53.21.

The aminobutonoic group of third hit compound ZINC38545694 which was mapped on the hydrogen bond acceptor (A4) feature of hypothesis showed hydrogen bonding interaction with Ala1349. The compound was found to have hydrogen bond interactions between the carboxyl group and the backbone of amide group of residues Gly1346, Lys1347, and Mg1 (Fig. 7c&8c). The hydrogen bond donor (D7) features were overlapped clearly in that region. Moreover, the aromatic features (R10) of hypothesis overlaid on the phenyl group showedπ-π stacking and hydrogen bonding interactions with His1453 (Fig. 9c). It has shown hydrophobic interaction with the amide group of Ala1490. The compound showed maximum Glide score -13.14 and GOLD fitness score of 50.11.

The fourth compound ZINC02138190 showed hydrogen bonding interaction with the active site residue Gly1346, Lys1347, and Mg1(Fig. 7d&8d). It has showed hydrogen bond donor (D7) feature of hypothesis as mapped and specifiedits donation of hydrogen atom from Gly1346, Lys1347 (Fig. 9d). The π-π stacking and strong hydrogen bonding interactions with His1453 which is an important residue in the active site of ROC domain of LRRK2.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

The Glide docking and GOLD fitness scores of the compound was found to be -13.04 and 48.44 respectively.

4.4.2 Overview of docking analysis

The guanine ring of the reference compound showed hydrogen bonding with amide chain of His1453 and Asn1455. The ROC was stabilized by unique stacking interactions with His1453, and Thr1491. The other GTPase structure showedoccupation of lysine residue at the position of His1453. This residue presented the sequence structural mark for the ROC GTPase superfamily. In addition, the β phosphate and α phosphate group of reference compound showed hydrogen bond interactions with backbone amide of the Gly1344, side chain hydroxyl group of Thr1349, Gly1346, Lys1347, and Mg1.

The four (ZINC00959174, ZINC02136788, ZINC38545694, ZINC02138190) hit compounds were selected based on the highest Glide scores along with GOLD fitness score,and mapped with five features of the ADRRR.42 as well (Fig. 9). The docking results of the four hit compounds were compared with the reference compound. The most important π-π stacking and strong hydrogen bonding interactions with amide group of His1453 were observed in all hit compounds. Based on the prediction of contour analysis, thefour hits showedthe presence of hydrogen bond donor D7 features favours the ROC of LRRK2 inhibitory activity. This was confirmed by hydrogen bonding interactions with critical residues such as Gly1346, Lys1347 and Mg1. After comparison of docking result with the reference compound, we strongly expressed the usefulness of the the selected hits in on inhibiting the LRRK2 activity.

5 Conclusion

In our study, we have generated an atom based pharmacophore model with highest survival score. The developed 3D-QSAR model was validated by internal and external methods for the predictability of hypothesis. The important structural features were identified by contour cubes for the developed pharmacophore model. Further, the selected model was used for screening the NCI, Maybridge, and ZINC databases. The retrieved hits from virtual screening were taken to molecular docking studies. The four hits (ZINC00959174,ZINC02136788,ZINC38545694,and ZINC02138190) were chosen based on good Glide score, GOLD fitness score, and molecular interaction. Further we applied DFT method to calculate HOMO, LUMO and energy gap for confirming the inhibitory potential for the four hits. Finally,ZINC00959174, ZINC02136788 were selected based on highest Glide score and GOLD fitness score with lowest energy gap and it could be helpful to design new classes of LRRK2 inhibitors.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

References

1. ParkinsonJ, *J Neuropsychiatry Clin Neurosci*,14(1817) 223 .
2. George DeMaagd, PharmD, BCPS and Ashok Philip, *Pharmacy & Therapeutics*, 40(2015) 504.
3. Junpeng Deng, Patrick A. Lewis, Elisa Greggio, Eli Sluch, Alexandra Beilina, and Mark R. Cookson, *Proceedings of national academy of science*,105 (2008) 1499.
4 M.Franzini, , X. M. Ye, M. Adler, D. L. Aubele, A. W. Garofalo, S. Gauby, E. Goldbach, G. D. Probst,K. P. Quinn, P. Santiago, H. L. Sham, D. Tam, A. Truong , Z, *Ren,Bioorganic & Medicinal Chemistry Letters*23(2013) 1967.
5 A. W Garofalo, M. Adler, D. L. Aubele, S. Bowers, M. Franzini, E. Goldbach, C. Lorentzen, R. J. Neitz, G. D. Probst, K. P. Quinn, P. Santiago, H. L. Sham, D. Tam, A. P. Truong, X. M. Ye , Z,*Ren Bioorganic & Medicinal Chemistry* Letters,23(2013) 71.
6 S. L.Dixon, , A. M. Smondyrev, E. H. Knoll, S. N. Rao, D. E. Shaw and R. A. Friesner,*J Comput Aided Mol Des*,20(2006),647.
7 D.Sen, and T. K. Chatterjee, *J Adv Pharm Technol* Res,4(2013) 50.
8 P.Pratim Roy, S. Paul, I. Mitra and K. Roy, *Molecules* 14(2009) 1660.
9 K.Roy, I. Mitra, P. K. Ojha, S. Kar, R. N. Das and H. Kabir,*Chemometrics and Intelligent Laboratory Systems* 118(2012) 200.
10 M. K Teli, and R. G. K ,*Organic and Medicinal Chemistry Letters*,2(2012) 25.
11 C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, *Adv Drug Deliv Rev* 46(2001) 3.
12 R. A., Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis , P. S. Shenkin, *J Med Chem*,47(2004), 1739.
13 Friesner, R. A., R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin , D. T. Mainz, *J Med Chem*,49(2006), 6177.
14 T. A.Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard , J. L. Banks,*J Med Chem*, 47(2004) 1750.
15 G.Jones , P. Willett , R. C. Glen,*J Mol Biol*,245(1995) 43.
16 S.Sakkiah, C. Meganathan, Y.S. Sohn, S. Namadevan , K. W. Lee, *International Journal of Molecular Sciences* 13(2012), 5138.
17. I. Goyal, V. Kumar, *Der Pharma Chem*, 3(2011),448.
18. P. K. Ojha, , K. Roy,*Biosystems*,113(2013), 177.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

Figure captions:

Fig. 1. Chemically diverse 30 compounds used as training set in Phase 3D-atom based QSAR pharmacophore model.

Fig. 2. Ten compounds used as test set compounds to validate the hypothesis.

Fig. 3. a) The best ADRRR hypothesis model. b) Hypothesis ADRRR, all distances stated in A ˚. c) Alignment of the most active compound 5 (highest fitness value = 3) applied to the ADDRR hypothesis. d) Alignment of the less active compound 39 (least fitness value = 0.75) applied to the ADRRR hypothesis.

Fig. 4. a) Scatter plot of experimental activity (pIC_{50}) verses predicted activity for training set compounds with correlation coefficient b) the correlation coefficient of experimental activity (pIC_{50}) verses predicted activity for test set compounds c) The correlation coefficient of ADDRR.46 superimposed with training set and test set.

Fig. 5. Virtual screening techniques followed during identification of the LRRK2 inhibitors.

Fig. 6. Schematic diagram of the LRRK2ROC domain active site;a) a) Superposition of the co-crystal ligand 2EZJ (ash) with its docked pose (pink) b)Yellow lines show the hydrogen-bond interactions between the substrate and cofactor and the surrounding protein residues.

Fig. 7. The binding modes and molecular interactions of 2D representation of hit compounds at the binding site of LRRK2 protein. a) ZINC00959174 b) ZINC02136788c) ZINC38545694
d) ZINC02138190

Fig. 8. The binding modes and molecular interactions of 3D representation of hit compounds at the binding site of LRRK2 protein. a) ZINC00959174 b) ZINC02136788c) ZINC38545694
d) ZINC02138190
.

Fig. 10. The ADDRR.42 hypothesis mapping of four hit compounds. a) ZINC00959174 b) ZINC02136788 c) ZINC38545694 d) ZINC02138190

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

Table 1 Score of different parameters of ADRRR hypothesis

Hypothesis	Survival	Survival -inactive	Post-hoc	Site	Vector	Volume	Selectivity	Matches	Activity	Inactive
ADRRR.42	**3.046**	**1.537**	**5.269**	**0.65**	**0.803**	**0.592**	**1.646**	**7**	**7.409**	**1.51**
AAADR.117	3.046	1.722	5.241	0.57	0.938	0.534	1.479	8	7.319	1.323
AAHHR.32	3.025	1.63	5.21	0.51	0.897	0.617	1.886	8	7.284	1.395
AAHHR.53	3.021	1.662	5.243	0.48	0.898	0.639	1.893	8	7.409	1.358
AAAAR.58	3.018	1.626	5.219	0.56	0.927	0.535	1.526	8	7.337	1.392
AAADR.195	2.997	1.72	5.151	0.53	0.87	0.599	1.56	8	7.18	1.277

Table 2 Distance between different sites of ADRRR.42

Entry	Site1	Site2
ADRRR.42	**A4**	**D7**
ADRRR.42	**A4**	**R11**
ADRRR.42	**A4**	**R10**
ADRRR.42	**A4**	**R12**
ADRRR.42	**D7**	**R11**
ADRRR.42	**D7**	**R10**
ADRRR.42	**D7**	**R12**
ADRRR.42	**R11**	**R10**
ADRRR.42	**R11**	**R12**
ADRRR.42	**R10**	**R12**

Table 3 PLS statistical parameters of the selected 3D-QSAR model

Hypothesis	PLS Factors	SD	R-squared	F	P	RMSE	Q-squared	Pearson-R
ADRRR.42	1	0.2856	0.2748	17.4	1.31E-03	0.3305	0.3409	0.6756
	2	0.1909	0.683	48.5	5.96E-12	0.2873	0.502	0.7416
	3	0.1497	0.8093	62.3	7.12E-16	0.2448	0.6384	0.831
	4	0.1271	0.8658	69.3	3.49E-18	0.2284	0.6851	0.8536
	5	0.0921	0.9312	113.6	2.79E-23	0.2408	0.7501	0.8382
	6	0.0772	0.9509	138	1.44E-25	0.2458	0.8729	0.844

Abbreviation: SD, standard deviation of the regression; R, squared value of R^2 for the regression; F, variance ratio. Large values of F indicate a more statistically significant regression, P, significance level of variance ratio. Smaller values indicate a greater degree of confidence; RMSE,root-mean-square error, Q, squared value of Q^2 for the predicted activities, Pearson-R, Pearson R value for the correlation between the predicted and observed activity for the test set

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

Table 4 Fitness and predicted activity data for the trainings set of compounds

Sl.no	Ligand name	QSAR set	Activity	PLS factor	Predicted activity	Pharm set	Fitness
1	**Compound 1**	Training	7.523	6	7.49	active	1.41
2	**Compound 2**	Training	7.509	6	7.52	active	1.89
3	**Compound 3**	Training	7.495	6	7.29	active	0.99
4	**Compound 4**	Training	7.432	6	7.51		1.86
5	**Compound 5**	Training	7.409	6	7.36	active	3
6	**Compound 6**	Training	7.319	6	7.31	active	1.86
7	**Compound 7**	Training	7.31	6	7.34	active	1.89
8	**Compound 8**	Training	7.284	6	7.17	active	2.11
9	**Compound 9**	Training	7.276	6	7.28	active	1.48
10	**Compound 10**	Training	7.18	6	7.19	active	2.13
11	**Compound 11**	Training	7.125	6	7.17	active	1.46
12	**Compound 12**	Training	7.056	6	7.13	inactive	0.93
13	**Compound 13**	Training	7.009	6	7.04	active	1.53
14	**Compound 14**	Training	6.987	6	6.96	active	1.36
15	**Compound 15**	Training	6.963	6	6.95	inactive	1.65
16	**Compound 16**	Training	6.932	6	7	active	1.46
17	**Compound 17**	Training	6.924	6	6.95	active	2.4
18	**Compound 18**	Training	6.86	6	6.98		1
19	**Compound 19**	Training	6.854	6	6.83	active	1.33
20	**Compound 20**	Training	6.848	6	6.83		2.44
21	**Compound 21**	Training	6.799	6	7.02		1.01
22	**Compound 22**	Training	6.785	6	6.81		0.84
23	**Compound**	Training	6.775	6	6.78		1.77

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

	23						
24	**Compound 24**	Training	6.737	6	6.84	inactive	1.39
25	**Compound 25**	Training	6.616	6	6.62	inactive	1.54
26	**Compound 26**	Training	6.577	6	6.53	inactive	1.57
27	**Compound 27**	Training	6.551	6	6.55		2.1
28	**Compound 28**	Training	6.503	6	6.52		1.54
29	**Compound 29**	Training	6.471	6	6.44		1.04
30	**Compound 30**	Training	6.422	6	6.4	inactive	1.38

Table 5 Fitness and predicted activity data for the test set compounds

Sl.no	Ligand name	QSAR set	Activity	PLS factors	Predicted activity	Pharm set	Fitness
31	**Compound 31**	Test	7.456	6	7.37	active	2.13
32	**Compound 32**	Test	7.337	6	7.31		1.88
33	**Compound 33**	Test	7.131	6	6.94	active	0.78
34	**Compound 34**	Test	7.119	6	7.08	active	1.36
35	**Compound 35**	Test	6.996	6	6.93		1.56
36	**Compound 36**	Test	6.917	6	7.09	inactive	1.73
37	**Compound 37**	Test	6.839	6	6.89		1.75
38	**Compound 38**	Test	6.772	6	6.72	inactive	1.41
39	**Compound 39**	Test	6.676	6	7	inactive	0.75
40	**Compound 40**	Test	5.888	6	5.9	inactive	1.4

Table 6 Results of the external validation for the atom based 3D-QSAR model

R^2pred	r_m^2 (test)		Δr_m^2 (test)		r_m^2 (overall)		Δr_m^2 (overall)		r_m^2 (rank)
0.967	**Scaled**	**Unscaled**	**Scaled**	**Unscaled**	**Scaled**	**Unscaled**	**Scaled**	**Unscaled**	0.7621
	0.8682	0.8785	0.0589	0.0664	0.887	0.8918	0.065	0.06793	

Table 7 ADME properties of hit compounds

ZINC ID	QPlogPo/w [a]	QPlogS [b]	QPlogHERG [c]	QPPCaco [d]	QPlogBB [e]	Percent human oral absorption [f]
ZINC00959174	**1.828**	**-2.47**	**-1.895**	**565.57**	**-1.658**	**98.986**

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 8994 - 9009
Received 25 April 2021; Accepted 08 May 2021.

ZINC02136788	**2.626**	**-4.278**	**-2.343**	**510.902**	**-2.457**	**92.901**
ZINC38545694	**3.367**	**-4.168**	**-2.47**	**422.271**	**-1.32**	**88.772**
ZINC02138190	**2.185**	**-3.633**	**-2.385**	**436.946**	**-2.54**	**79.507**

[a]Predicted octanol/water partition co-efficient log p (acceptable range: −2.0 to 6.5).
[b]Predicted aqueous solubility; S in mol/L (acceptable range: −6.5 to 0.5).
[c]Predicted IC_{50} value for blockage of HERG K+ channels (acceptable range: below −6.0).
[d]Predicted Caco-2 cell permeability in nm/s (acceptable range , <25 is poor and >500 is great)
[e]Predicted blood brain barrier permeability(acceptable range: −3 to 1.2).
[f]Percentage of human oral absorption (.25% is poor and .80% is high).

Table 8 Glide and GOLD docking

Hit compounds	Glide score	GOLD fitness score
ZINC00959174	-13.46	56.76
ZINC02136788	-13.36	53.21
ZINC38545694	-13.14	50.11
ZINC02138190	-13.04	48.44