Chapter 7
Homology Modelling, Structure-Based Pharmacophore Modelling, High-Throughput Virtual Screening and Docking Studies of L-Type Calcium Channel for Cadmium Toxicity

Madhu Sudhana Saddala and A. Usha Rani

Abstract Cadmium (Cd) is a heavy metal present in air, water, soils and sediments. It is well known that long-term exposure to Cd causes various toxic effects in various organ systems such as cardiovascular, kidneys, liver, brain, lung, bones, immune/haemopoietic, endocrine and reproductive systems. Cd influx mediates voltage-gated L-type calcium channels (LCC) in excitable cells including mammalian neurons and also Cd uptake in non-excitable tissues. Therefore, LCC has been recognized as an attractive metal toxicity target. We construct a homology model of LCC in addition to the generated pharmacophore models then used to retrieve 50,500 molecules from Zinc database. There are 18 best reliable molecules mapped with core pharmacophore model of LCC. These hits were retrieved and further evaluated by molecular dynamics (MD) simulation, molecular docking and protein–ligand interactions, and binding affinity predictions as well as in silico ADMET properties were tested. Our work results focus on homology modelling, structure-based pharmacophore mapping, molecular docking, MD simulation, protein–ligand interactions and binding affinity predictions which were used in virtual screening strategy to spot new hits for blockade of LCC. Finally, the outcome results, priming the five best lead compounds, were expected to be the potential lead scaffolds for developing novel and potent blockers of LCC against metal toxicity.

Keywords Cadmium • LCC • Docking • Pharmacophore • MOE • ADMET

M.S. Saddala (✉) • A. Usha Rani
Division of Bioinformatics, DBT- Bioinformatics Centre, Department of Zoology, Sri Venkateswara University, Tirupati 517502, Andhra Pradesh, India
e-mail: madhubioinformatics@gmail.com

D.-Q. Wei et al. (eds.), *Translational Bioinformatics and Its Application*, Translational Medicine Research, DOI 10.1007/978-94-024-1045-7_7

7.1 Introduction

Cadmium (Cd) is a heavy metal present in air, water, soils and sediments (Kocak and Akcil 2006). Cd is widely used in pigments, plastic stabilizers, electroplating, alloys, nickel–Cd batteries and welding industry and is also present in tobacco (Pappas et al. 2006). It is well known that long-term exposure to Cd causes various toxic effects in various organ systems such as cardiovascular, kidneys, liver, brain, lung, bones, immune/haemopoietic, endocrine and reproductive systems (Satarug et al. 2010). Several efforts are being made to find a substance which can significantly decrease the magnitude of metal toxicity when present in the biological system during heavy metal intoxication. Membrane damage caused by the reactive oxygen species (H_2O_2 and OH^- ions) generated from the exposure of living tissues to heavy metals may allow the entry of excess calcium into the cells with a subsequent biochemical cellular degradation and necrosis. Calcium channel blockers act on ion-conducting cell membrane channels. The 1,4-dihydropyridine moiety is commonly useful as calcium channel blockers and is used most frequently as drugs such as nifedipine, diltiazem, nicardipine and amlodipine (AD), which have been found as potent cardiovascular agents for the treatment of hypertension. Hence, this class of agents may be included in the search for protectors with a more favourable therapeutic index. Therefore, the present study is an attempt to find out the detoxifying action of calcium channel blockers against cadmium-induced toxicity in albino rats through computational tools.

In recent years, high-throughput virtual screening has been emerging as a complementary to high-throughput screening in an attempt to discover novel potential lead compounds in the process of drug discovery (Lyne 2002). Thus, to identify new and potent compounds that block the L-type calcium channel (LCC) model like AD, structure-based pharmacophore modelling and virtual screening may be considered as an effective approach. This study describes the structure-based pharmacophore modelling to identify the pharmacophoric features required for simultaneous inhibition of LCC for Cd toxicity by virtual screening: molecular docking, protein–ligand interaction fingerprints (PLIFs), binding energy calculations and binding affinity predictions.

7.2 Material and Methods

7.2.1 Homology Modelling

For unknown protein structures such as membrane proteins, homology modelling was introduced to construct the three-dimensional structure of a known atomic resolution model of the protein (target) and related homologous protein (template).

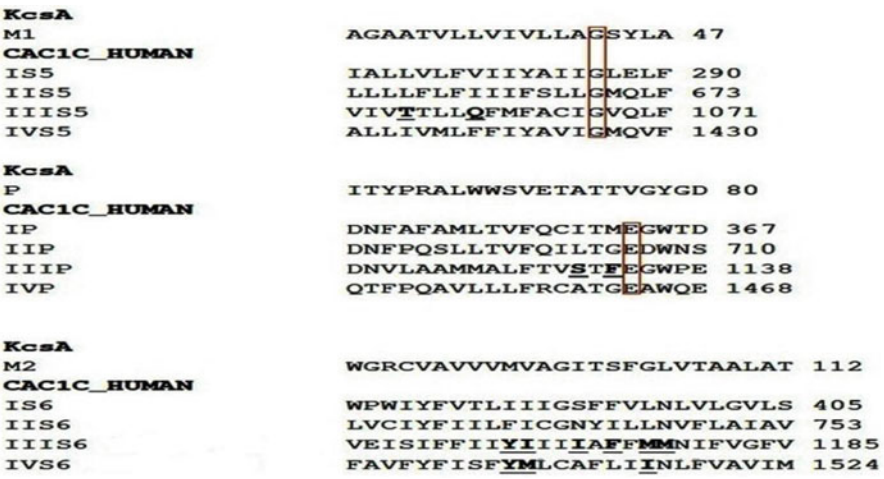

Fig. 7.1 Pairwise alignment of CAC1C_HUMAN and KcsA sequences. The conserved key residues used to align the sequences are shown in *red boxes*. Residues reported to affect DHP antagonist binding and underscored and highlighted in *bold*

7.2.1.1 Construction of the Human LCC Model

The structural model of the human LCC was built using the recently reported 3.20 Å crystal structure of KcsA (Shaldam et al. 2014) (PDB entry code 1BL8) as a structural template. The sequence of the human LCC pore region α1c subunit (Cav1.2, CAC1C_HUMAN) was retrieved from the Swiss-Prot database (Shaldam et al. 2014) and aligned as described in the Results and Discussion section (Fig. 7.1). The construction of the transmembrane region of the model was achieved by the employment of the modeller 9.13.

The protocol used to develop the LCC model is divided into three phases: sequence alignment, model construction and model evaluation.

7.2.2 Sequence Alignment

The model was constructed using amino acid sequence of voltage-dependent LCC subunit alpha-1C (CAC1C HUMAN Q13936) obtained from UniProtKB/Swiss-Prot sequence database (Reyes et al. 1990; http://www.uniprot.org/unipro/Q13936). Coordinates of potassium channel KcsA atoms in their closed conformation were downloaded from the RCSB Protein Data Bank (PDB ID: 1BL8). Amino acid sequences of S5, S6 and P-loops in between for the four repeats (I–IV) (271–405, 654–753, 1052–1185 and 1411–1524, respectively) were used for sequence alignment with the amino acid sequence of KcsA as proposed by Zhorov et al. (2001) (Fig. 7.1). In order to favour valid superimposition of the residues, the

sequence of each repeat was organized as S5, S6 and P-loop, allowing for a more flexible inspection of the results and easier corrections. The amino acid sequence of repeats I and III has a long extracellular loop which would decrease the quality of the generated model, so amino acid sequences were excluded from repeats I and III, respectively.

7.2.3 Construction of the LCC Model

The modelling procedure consisted of two steps: model construction from the template and refinement of loops. The above described sequence alignment file was used as input for the MODELLER 9.13 program (Sali et al. 1995) with the high-resolution NMR structure of potassium channel KcsA available in the RSCB Protein Data Bank (PDB ID: 1BL8) as a template for the 3D structure. Molecular modelling studies were performed using the MODELLER 9.13 running on Intel Core 2 Duo CPU personal computers. The model sequence, template structure and sequence alignment were used as input files to build the model. Loops can be defined automatically from the model to a template sequence alignment. The MODELER Loop Refinement-DOPE-Loop method (Shen and Sali 2006; Shaldam et al. 2014) was used for initial refinement of the loop conformation after model generation. The model side-chain conformation was optimized based on systematic searching of side-chain conformation and CHARMm energy minimization using the ChiRotor algorithm (Spassov et al. 2007; Shaldam et al. 2014). Five models were obtained from the first step of molecular modelling. These models were subjected to a comparison based on the best scores to reveal the differences among them. The model with the lowest energy and the lowest restraint violation was selected for the second step. Secondly, the loops between helices were subjected to refinement while keeping the start and end residues constrained. This procedure is based on the idea that transmembrane helices are much less flexible than loops, thus permitting to produce a sounder core alignment if the integrity of the helices is conserved. The more unpredictable loops can bear the more important differences. A CHARMm-based protocol (Spassov et al. 2008; Shaldam et al. 2014) that optimizes the conformation of a contiguous segment (i.e. a loop) of a protein structure was used for loop refinement. It is based on systematic conformational sampling of the loop backbone and CHARMm energy minimization. This approach can be used to refine a loop structure from a homology model as well as to optimize a segment of the protein experimental structure where the structure is poorly defined. The homology modelling (HM) phase was followed by the model evaluation phase. The stereochemical quality and structural integrity of the model were tested by RAMPAGE, ERRAT, MolProbity, ProSA and Verify3D software and target–template superimposition by PyMol (Eswar et al. 2008) (Fig. 7.2).

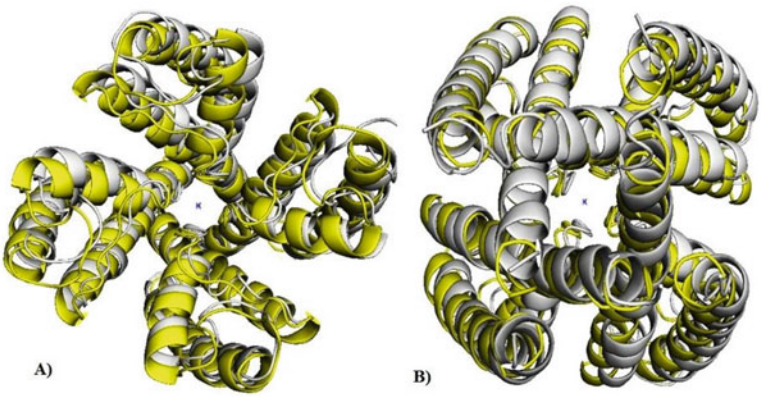

Fig. 7.2 Superimposition of the LCC model (*white*) and KcsA (*yellow*) (PDB: 3BPM). (**a**) Open conformation, (**b**) closed conformation

7.2.4 Active Site Identification

The active site of LCC model was identified using a CASTp server (Computer Atlas of Surface Topology of Proteins) (Dundas et al. 2006). A new program, CASTp, for automatically locating and measuring protein pockets and cavities, is based on precise computational geometry methods, including alpha shape and discrete flow theory. CASTp identification, measurements of surface accessible pockets as well as interior inaccessible cavities by locating, delineating and measuring concave surface regions on three-dimensional structure of proteins. The measurement includes the area and volume of pocket or void by solvent-accessible surface model (Richards' surface) and by molecular surface model (Connolly's surface), calculated analytically. It can also be used to study surface features and functional regions of proteins.

7.2.5 Generation of Structure-Based Pharmacophore Model

In the present study, the LCC modelled receptor complex with a channel blocker AD was used as starting structure for the generation of structure-based pharmacophore models (Abdul et al. 2012). LIGANDSCOUT (LS) is a tool that allows the automatic construction and visualization of 3D pharmacophore for structural data of macromolecule/ligand complexes. For the LS algorithm, chemical features include hydrogen bond donors and acceptors as directed vectors, and positive and negative ionizable regions as well as lipophilic areas are represented by spheres. Moreover, to increase the selectivity, the LS model includes spatial information regarding areas inaccessible to any potential ligand, thus reflecting possible steric restrictions. In particular, for excluded volume spheres placed in

positions that are sterically forbidden, LS may also be used to construct pharmacophore of varying degrees of sophistication, suitable for export to different programs. In the present study, Molecular Operating Environment (MOE, version 2008, Chemical Computing Group Inc.)-compatible 3D pharmacophore model was first developed by LS using default parameters, and then, it was exported and converted into a MOE, pharmacophore query for virtual screening (http://www.chemcomp.com). Prior to the screening, it was necessary to make a number of adjustments, because feature interpretation differs slightly between the two programs. Those aromatic rings that LS classified simply as hydrophobic groups were classified as either aromatic or hydrophobic in MOE, using the PPCH_All scheme (which incorporates directionality of hydrogen bond donors and acceptors and orientation of aromatic rings). As in LS pharmacophore, the aromatic ring is not directly classified as such (because of the lack of detection of π–π stacking or cation–π interactions) but, rather as a set of hydrophobic atoms, can be interpreted in MOE in a manner that is useful in the prediction of right compounds in virtual screening.

7.2.6 Pharmacophore-Based Virtual Screening

The Zinc database (http://zinc.docking.org/), which allows the user to download compounds, structures from a variety of vendors as SDF files based on the structure-based amlodipine (AD) compound (Query), was used in this preliminary screen. Using MOE, the database was washed, and the 3D structure of each compound was built using the MMFF94x force field. Then for each compound, the low-energy conformers were generated using Conformation Import methodology implemented in MOE software. After assessing the pharmacophore query, virtual screening was carried out using the software MOE against the Zinc database. Because some changes may occur when the pharmacophore is exported from LS to MOE environment, therefore, the pharmacophore queries were validated before using it for virtual screening. To reduce the data of identified hits, they were docked into the recently identified binding pocket of LCC model, and the PLIFs were developed using MOE. Binding energies and binding affinities were calculated using LIGX (Chemical Computing Group, Montreal, Quebec, Canada) implemented in MOE to prioritize the final hits.

7.2.7 Molecular Docking

Docking is a computational method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Docking has been widely used to suggest the binding modes of protein inhibitors. Most docking algorithms are able to generate a large number of possible structures; thus,

they also require a means to score each structure to identify those that of greatest interest. Docking was performed using AutoDock in PyRx Virtual Screening tool (Wolf 2009; Trott and Olson 2010).

Pharmacophoric hit compounds were docked into the active site of the refined LCC model. Lamarckian genetic algorithm was used as the number of individual population (150), max number of energy evaluation (25000000), max number of generation (27000) (Laskowski et al. 1993), gene mutation rate (0.02), crossover rate (0.8), Cauchy beta (1.0) and GA window size (10.0). The grid was set whole protein due to the multi-binding pocket at X=3.42, Y=−56.23, Z=98.32 and dimension (Å) at X=89.92, Y=98.56, Z=98.32 and exhaustiveness 8. The pose for a given ligands identified on the basis of highest binding energy. Only ligand flexibility was taken into account and the proteins were considered to be rigid bodies. The resulting complexes were clustered according to their root mean square deviation (RMSD) values and binding energies, which were calculated using the AutoDock scoring function. The PyMol molecular viewer (http://www.pymol.org/) was employed to analyse the docked structures.

7.2.8 Analysis of Drug Likeness

MolSoft Drug Likeness explorer (http://www.molsoft.com/mprop/) was used to analyse the drug likeness as per "Lipinski rule of 5" (Lipinski et al. 1997). According to "Lipinski rule of 5", a compound is more likely to be membrane permeable and easily absorbed by the body if its molecular weight is less than 500, its lipophilicity expressed as a quantity known as log P is less than 5, the number of groups in the molecule that can donate hydrogen atoms to hydrogen bonds is less than 5 and the number of groups that can accept hydrogen atoms to form hydrogen bonds is less than 10 (Leeson 2012).

7.2.9 ADMET Properties

The in silico pharmacokinetic properties and ADMET (absorption, distribution, metabolism, elimination and toxicity) analysis were predicted using OSIRIS property explorer (http://www.organic-chemistry.org/prog/peo/; Access date: September 23, 2014) which uses Chou and Jurs algorithm, based on computed atom contributions.

7.3 Results and Discussion

7.3.1 Sequence Alignment

Besides the choice of the reference, the accuracy of the alignment is the most crucial step in assuring the quality of the homology modelling. An accurate sequence alignment between the model and the template proteins is essential to achieve high-quality models. Voltage-gated LCC are members of a gene superfamily of transmembrane ion channel proteins that includes voltage-gated K^+ and Na^+ channels. LCC share structural similarities with K^+ and Na^+ channels in that they possess a pore-forming $\alpha 1$ subunit in four repeats of a domain with six transmembrane-spanning segments that include the voltage-sensing S4 segment and the pore-forming (P) region. As no atomic resolution images of calcium channel structures exist, much has been learnt about their structure since the recent determination of crystal structures of a number of potassium channels (Jiang et al. 2003; Long et al. 2005; Shaldam et al. 2014). The $\alpha 1$ subunit contains four repeated domains (I–IV), each of which includes six transmembrane segments (S1–S6) and a membrane-associated loop (the "P-loop") between segments S5 and S6. The four repeated domains are also remarkably similar to those known to form the voltage-gated potassium channels. However, potassium $\alpha 1$ subunit is homotetramer and calcium channel is heterotetramer. Potassium channel KcsA (PDB code 1BL8) has been selected to be the template. Amino acid sequences of S5, S6 and P-loops in between the four repeats (I–IV) (271–405, 654–753, 1052–1185 and 1411–1524, respectively) of voltage-dependent LCC subunit alpha-1C (CAC1C HUMAN Q13936) were used for sequence alignment with the amino acid sequence of KcsA as proposed by Zhorov et al. (2001), where S6 segments of LCC are aligned with M2 segments of KcsA in a manner similar to the alignment of the Na^+ channel with KcsA described by Lipkind and Fozzard (2000) and S5s were aligned with the M1 segment of KcsA as proposed by Huber et al. (2000) and the P-loops were aligned using MULTALIN server (Corpet 1988; Shaldam et al. 2014) (Fig. 7.1). Proteins that fold into similar structures can have large differences in the size and shape of residues at equivalent positions. These changes are tolerated not only because of replacements or movements in nearby side chains, as explored by Ponder and Richards, but also as a result of shifts in the backbone (Bowie et al. 1991; Shaldam et al. 2014). For a more flexible inspection, the sequence of each repeat was organized as S5, S6 and P-loop, allowing easier corrections. The amino acid sequence of repeats I and III has long extracellular loop which would reduce the quality of the generated model, so amino acid sequences were excluded from repeats I and III, respectively. Since the template is 88 residues shorter than the target protein, gaps were inserted to achieve best sequence similarity and identity without affecting sequence alignment proposed by Zhorov et al. (2001). The greatest attention was thus paid to the careful construction of transmembrane helices S5 and S6 and P-loop as well.

7.3.2 Construction of the LCC Model

Although the two proteins have low sequence identity of 9.5% and sequence similarity of 29.2%, the MODELLER program was applied to generate satisfactory models. As an integral process of model building, initial refinement of the loop conformation after model generation was automatically performed by MODELER Loop Refinement-DOPE-Loop method during the process. The model achieved from the alignments by Zhorov et al. (2001) was subjected to extensive loop optimization. This procedure is based on the idea that transmembrane helices are much less flexible than loops, thus permitting to produce a sounder core alignment if the integrity of the helices is conserved. On the contrary, the more volatile loops can bear the more important difference between the coordinates of the reference and the model. When a homology model is created, there are parts of the model sequence which are not aligned to any template structures. For these sections, no homology restraints (such as Cα–Cα distance restraints) can be applied. These parts of the structure generally have greater errors compared to the regions which are modelled based on a template structure. In attempts to reduce these errors, a CHARMm-based protocol that optimizes the conformation of a contiguous segment (i.e. a loop) of a protein structure called loop refinement was applied (Spassov et al. 2008; Shaldam et al. 2014). This is based on systematic conformational sampling of the loop backbone and CHARMm energy minimization. The algorithm goes through three stages: construction and optimization of loop backbone, construction of loop side chain and optimization of loop followed by reranking of the conformations. The model was then checked after a thorough energy minimization designed to reduce the steric clashes of the side chains without modifying the backbone of the protein to solve these contacts. To avoid modification of the backbone of the protein, the optimization of the geometry of side chain was performed with constraining the backbone. After the optimization, models were checked to assess the quality of their structure.

7.3.3 Model Evaluation

To assess stereochemical quality and structural integrity of the model, RAMPAGE (Lovell et al. 2003) (Fig. 7.3), ERRAT (Colovos and Yeates 1993; Shaldam et al. 2014), ProSA (Sippl 1993; Wiederstein and Sippl 2007) and Verify3D (Luthy et al. 1992; Shaldam et al. 2014) software were used. For comparison, these methods were also used to evaluate the template structure, and then each repeat was examined separately by means of ProSA. RAMPAGE is an offshoot of RAPPER which generates a Ramachandran plot using data derived by Lovell et al. (2003). It is recommended that it be used for this purpose in preference to PROCHECK, which is based on much older data. The Ramachandran diagram plots phi versus psi dihedral angels for each residue in the protein. The diagram is divided into

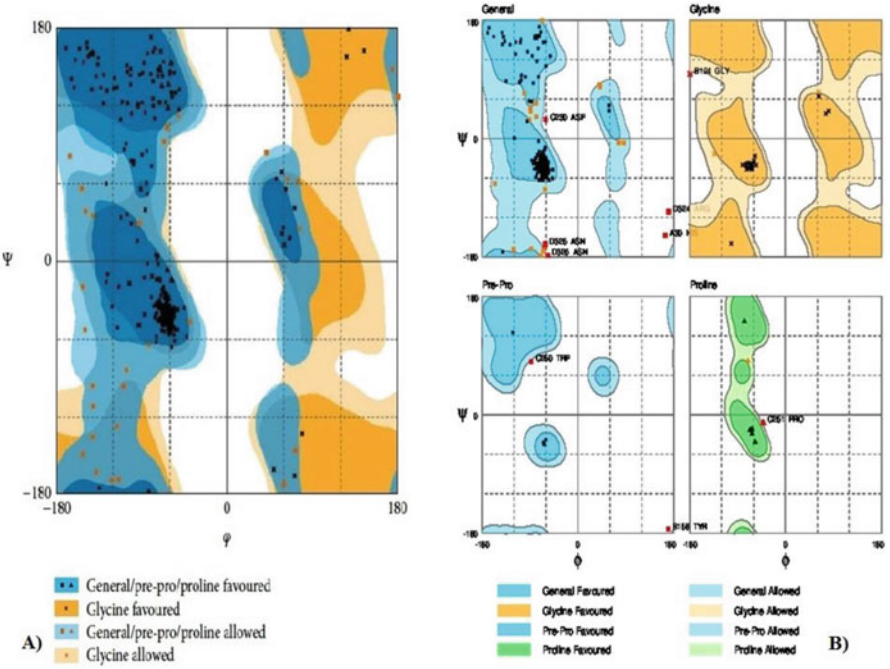

Fig. 7.3 Ramachandran plot. (**a**) The plot of LCC model shows that 92.3% of residues were found in the favoured, 7.7% in the allowed and none in the outlier regions. (**b**) The plot shows general, glycine, pre-proline and proline for LCC model

favoured, allowed and disallowed regions, whose contouring is based on density-dependent smoothing for 81,234 non-glycine, non-proline residues with B < 30 from 500 high-resolution protein structures. Regions are also defined for glycine, proline and pre-proline as shown in Fig. 7.3.

ERRAT is a protein structure verification algorithm, that is, especially well-suited for differentiating between correctly and incorrectly determined regions of protein structures based on characteristic atomic interactions (Colovos and Yeates 1993; Shaldam et al. 2014). Different types of atoms (C, N and O) are distributed nonrandomly with respect to each other in proteins because of energetic and geometric effects. Errors in model building lead to more randomized distributions of the different atom types, which can be distinguished from correct distributions by statistical methods. The program works by analysing the statistics of nonbonded interactions between different atom types. A single output plot is produced that gives the value of the error function versus position of a nine-residue sliding window. In comparison with statistics from highly refined structures, the error values have been calibrated to give confidence limits. The program provides an "overall quality factor" value which is defined as the percentage of the protein for which the calculated error value falls below the 95% statistical rejection limit. The ERRAT overall quality factor of the model is given in Table 7.1. This is not

Table 7.1 Assessment scores for the LCC receptor model

S. no.	Item	Model	Comment
1	ProSA	−3.89	ProSA Z-score as average of the four repeats
2	ERRAT	79.72	ERRAT overall quality factor
3	Verify3D	67.58% (W)	Percentage of residues with Verify3D average score > 0.2; verify3D overall assessment of the structure (P = pass, W= warning or F = fail) shown in parentheses

surprising since the model has longer loops than template. This method provides a useful tool for model building, structure verification and making decisions about reliability. A more reliable discrimination of incorrect regions would likely be obtained by combining the present analysis with others (Fig. 7.5).

ProSA and Verify3D are two methods that are sensitive in distinguishing between overall correct fold and those with an incorrect fold (Bhattacharya et al. 2008; Shaldam et al. 2014). ProSA (Protein Structure Analysis) program is a diagnostic tool that is based on the statistical analysis of all available protein structures (Wiederstein and Sippl 2007; Shaldam et al. 2014). It is a tool widely used to check 3D models of protein structures for potential errors. Its range of application includes error recognition in experimentally determined structures (Teilum et al. 2005; Llorca et al. 2006; Shaldam et al. 2014), theoretical models (Petrey and Honig 2005; Ginalski 2006; Shaldam et al. 2014) and protein engineering (Beissenhirtz et al. 2006; Mansfeld et al. 2006; Shaldam et al. 2014). The energy of the structure is evaluated using a distance-based pair potential and a potential that captures the solvent exposure of protein residues. From these energies, two characteristics are derived and displayed: Z-score and a plot of residue energies. The Z-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. Z-scores outside a range characteristic of native proteins indicate erroneous structures. The overall quality score calculated by ProSA for a specific structure is displayed in a plot that shows the scores computed from all experimentally determined protein chains currently available in the Protein Data Bank (PDB). Structures which contain errors are likely to have Z-score outside the range of values characteristic of native proteins. Table 7.1 lists the Z-score calculated by ProSA (as average of the four repeats Z-score) for the model and compared against the template. The Z-scores for the model and template are much closer to the middle region of scores observed for experimentally determined protein structures in the PDB including the template structure. The energy plot shows the local model quality by plotting energies as a function of amino acid sequence. In general, positive values correspond to problematic or erroneous parts of the model (Fig. 7.4).

Verify3D analyses the compatibility of an atomic model (3D) with its own amino acid sequence (1D) and hence tests the accuracy of the model (Fig. 7.5). Each residue is assigned a structural class based on its location and environment. The environments are described by the area of the residue buried in the protein and inaccessible to solvent, the fraction of side chain area that is covered by polar atoms

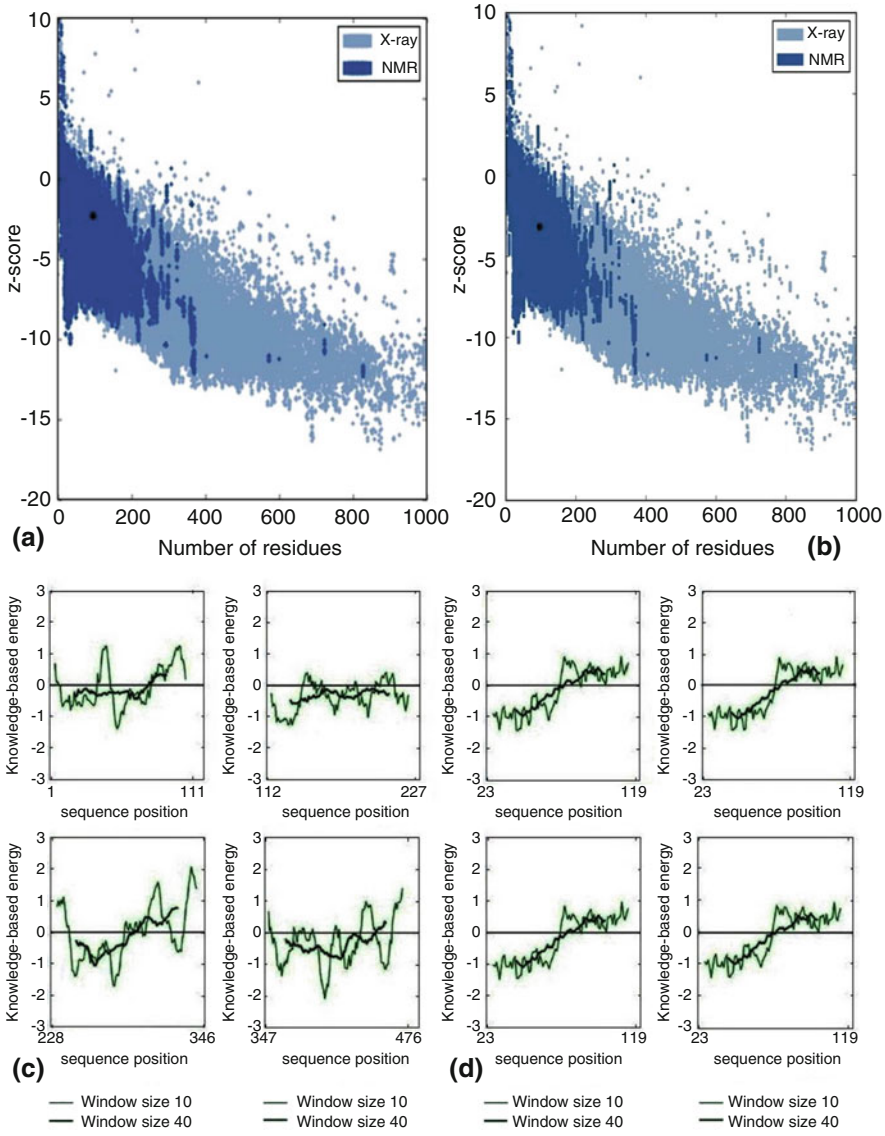

Fig 7.4 ProSA plot. Each repeat was examined separately. (**a**) ProSA Z-scores for LCC model; (**b**) ProSA Z-scores for template (KcsA) and *blue* and *sky blue dots* are Z-scores of PDB structures determined by X-ray crystallography and NMR, respectively; (**c**) ProSA energy profiles for LCC model (four repeats); (**d**) ProSA energy profiles for template (four repeats). Negative scores indicate a high-quality model

(O and N) and the local secondary structure. Based on these parameters, each residue position is categorized into an environmental class. In this manner, a 3D protein structure is converted into a 1D string, like a sequence, which represents the

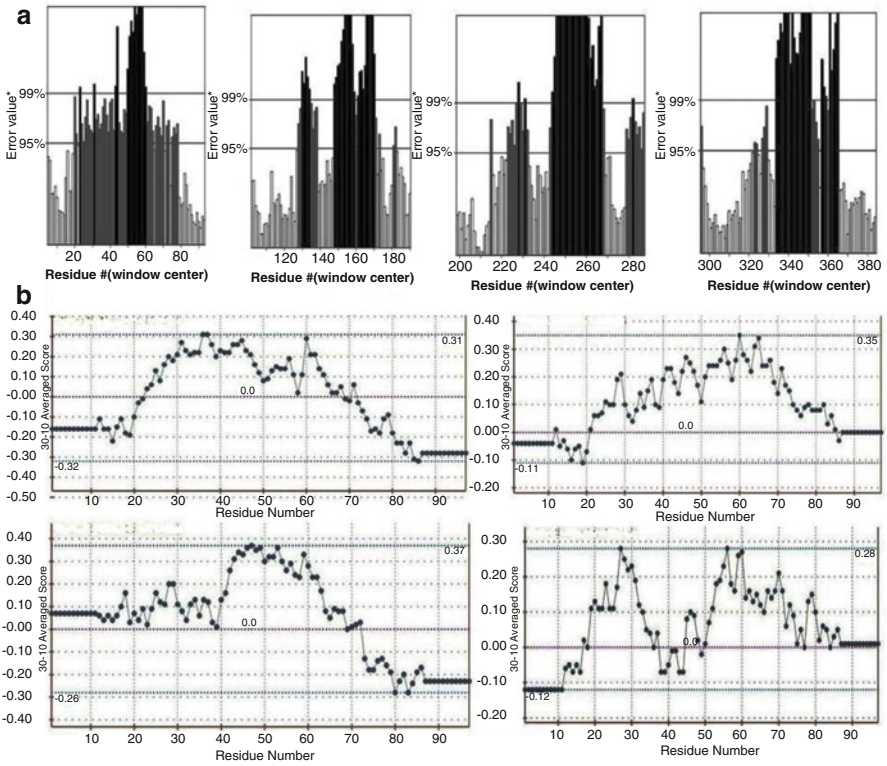

Fig. 7.5 (**a**) ERRAT score of the LCC model (four repeats). (**b**) Verify3Dscore profile calculated for LCC model. Scores over 0.2 indicate a high-quality model

environmental class of each residue in the folded protein structure. A collection of good structures is used as a reference to obtain a score for each of the 20 amino acids in this structural class. The scores of a sliding 21-residue window are added and plotted for individual residues. This method evaluates the fitness of a protein sequence in its current 3D environment. It can be applied to assess the quality of a theoretical model or to examine the characteristics of an experimental structure (Luthy et al. 1992; Shaldam et al. 2014). Table 7.1 shows the percentage of residues that had an average score > 0.2 and the Verify3D assessment of the structure (pass, warning or fail) for the model and template. Figure 7.5 shows the Verify3D profile for the model structure. Residues with a score over 0.2 should be considered reliable and the sequences exhibiting lower scores are those of extracellular loops.

Taken together, all of the above data indicate that the quality of the model is similar to that of the template. The model can be used for further computer-aided drug design (CADD) and it can be used in understanding how DHP work at the molecular level.

7.3.4 Generation of Structure-Based Pharmacophore Model

As shown in Fig. 7.6, the pharmacophore model automatically generated by the LS program includes four features: two hydrogen bond donors (HBD) (green colour) and three hydrophobic groups (yellow colour). Besides, the program automatically generated several excluded volumes (grey colour) in the model. The two HBD feature points are the amino group hydrogen atoms of the ligand towards the SER-78 and ILE-51, respectively. The three hydrophobic groups are located on the benzene group, chlorine atom on benzene and carboxy ethyl group of the ligand. The developed pharmacophore model was exported into MOE. Prior to screening, it was necessary to make a number of adjustments, as feature interpretation varies slightly between the two programs. As in LS pharmacophore, the aromatic ring of the compound in the complex was not classified as aromatic or hydrophobic features; thus, these were interpreted in MOE, using the PPCH_All scheme. Two modifications were made on this model to obtain appropriate model for virtual screening. The first modification is about the chlorobenzyl ring. It is clear that it is an aromatic group, but the LS could not interpret this ring as an aromatic group automatically. In MOE, additional features were developed using the MOE pharmacophore query editor. First, an aromatic feature was developed on the chlorobenzyl ring, and a hydrophobic feature was developed on the carboxy ethyl group of the ligand. This modified pharmacophore model was then validated by screening the test database. In the test database, we kept the compound (i.e. AD) present in complex structure. First, the AD was extracted, and then, hydrogen atoms were added and energy minimized by using MOE. The minimized structure of AD was added to the test database. After screening, the test compound was correctly mapped by the modified pharmacophore model as shown in Fig. 7.6. The result verified the validity of our modified pharmacophore model that can be used for the screening of large databases.

7.3.5 Pharmacophore-Based Virtual Screening

The modified validated pharmacophore model was then used as in silico filter to screen the Zinc database (http://zinc.docking.org/) of commercially available compounds. The Zinc database compounds in SDF format were loaded into MOE environment where the 3D structure of each compound was modelled using MMFF94x force field. The Conformation Import methodology was applied to generate low-energy conformations for each compound. All these compounds and their respective conformations were saved in MOE database. The conformers of each compound were then filtered by the pharmacophore model. To be considered as hit, the compound has to fit all the features of the pharmacophore. From the pharmacophore-based virtual screening, 18 hits (Fig. 7.7) were identified that mapped on the developed pharmacophore model (i.e. having the specified

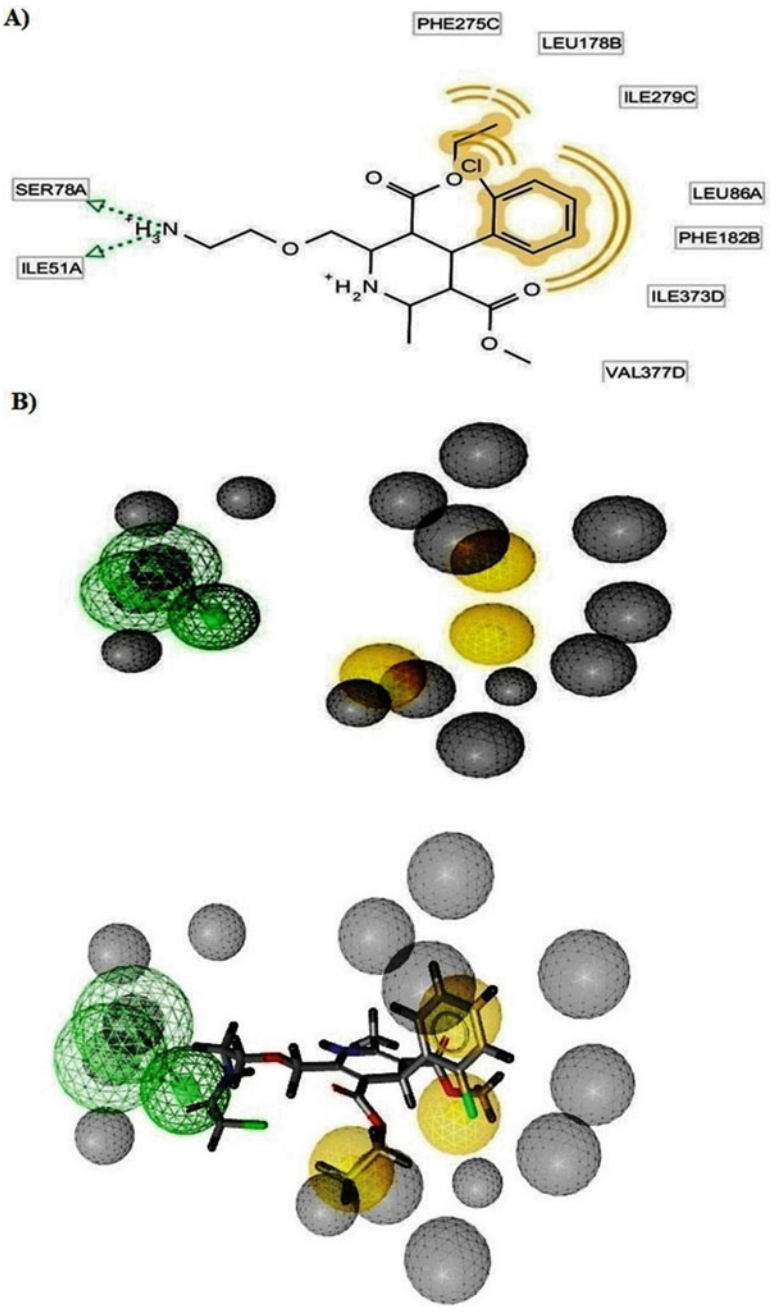

Fig. 7.6 (**a**) Two-dimensional pharmacophore model generated by LIGANDSCOUT from the complex structure of LCC and AD. The *dotted arrows* indicated the hydrogen bond donor (HBD) features. (**b**) The *yellow sphere* represented the HBD; the *yellow sphere* represented the hydrophobic feature in the ligand, whereas the *grey* colour spheres represented the excluded volumes

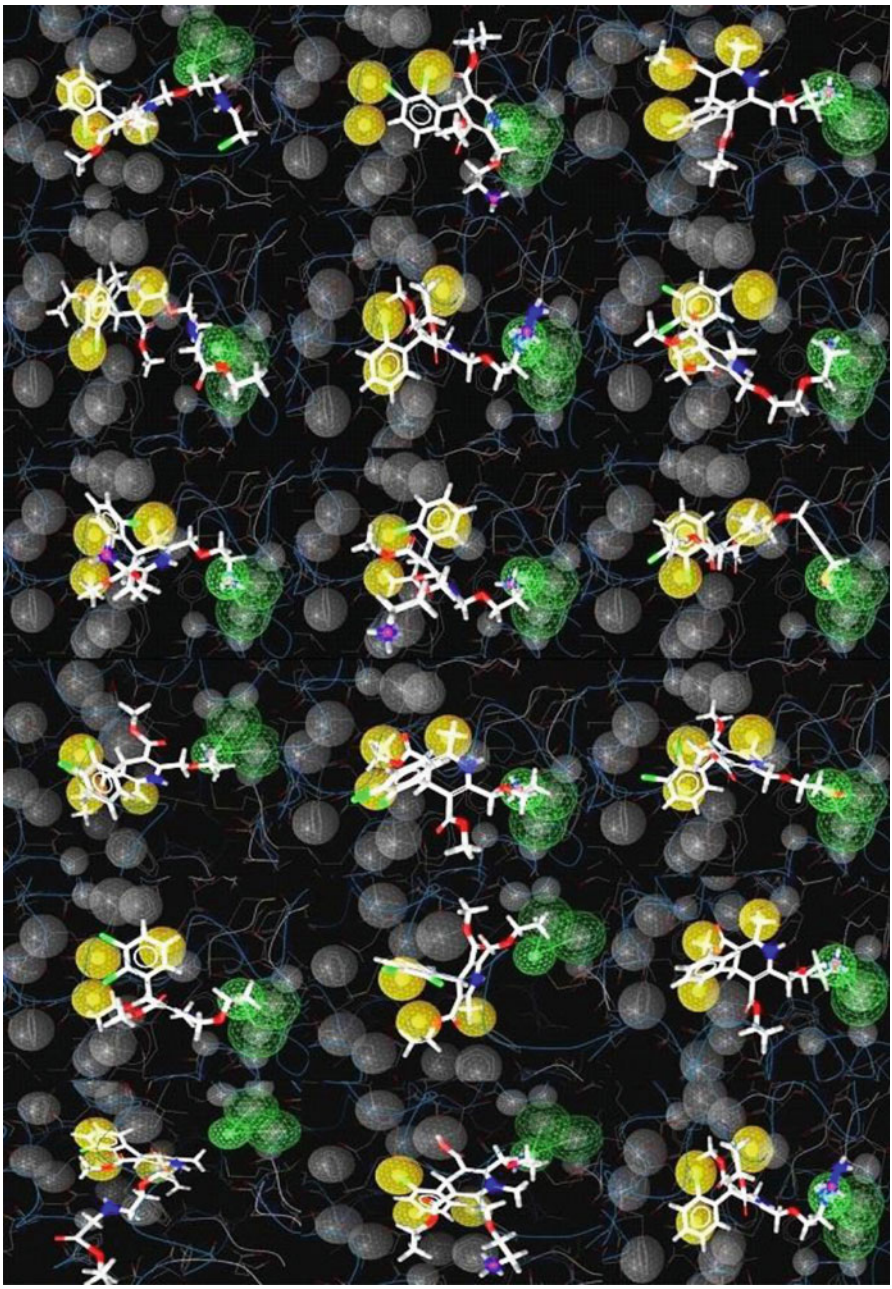

Fig. 7.7 The pharmacophore screened 18 hits from Zinc database

Table 7.2 AD and AD analogue compounds along with their respective interaction energies and their surrounding residues

S. no.	Compound	BE (kcal/mol)	Surrounding residues
1	Zinc59347487	−8.4	ILE-360 (**IP**), SER-393 (**IS6**), ASN-398 (**IS6**), ASN-740 (**IIS6**), ASN1517 (IVS6)‘
2	Zinc20267861	−8.1	LEU-704 (**IIP**), ASN-740 (**IIS6**), ASN-1517 (**IVS6**)
3	Zinc59486248	−7.9	THR-361 (**IP**), MET-362 (**IP**)
4	Zinc59494792	−7.6	THR-361 (**IP**), ASN-740 (**IIS6**)
5	Zinc67664832	−7.2	THR-361 (**IP**), ILE-360 (**IP**), SER-393 (**IS6**), ASN-1178 (**IIIS6**), ASN-740 (**IIS6**)
6	Zinc19796039 (AD) (Query)	−5.4	ILE-360 (**IP**), SER-393 (**IS6**), ASN-1178 (**IIIS6**), ASN-740 (**IIS6**)

requirements). These initially identified hits were selected for further evaluation using docking studies.

7.3.6 Molecular Docking

In order to shed light on the molecular basis of the interactions between LCC and its ligands, docking simulations were undertaken on pharmacophoric hits of DHPs (dihydropyridines) on LCC model. Such calculations were conducted employing the automated docking program AutoDock which has proven to be really effective in reproducing the experimentally found posing of ligands into their binding site. As shown in Table 7.2, the predicted free energy of binding top five compounds. Docking of hits into active site of LCC model gave comparable binding solutions with the dihydropyridine ring fitting in the cleft formed by IIIS6, IIIS5 and IVS6 segments. The Zinc59347487 compounds bound -8.4 binding energy with ILE-360 (IP), SER-393 (IS6), ASN-398 (IS6), ASN-740 (IIS6) and ASN-1517 (IVS6) active site residues, respectively. The Zinc20267861 compounds bound -8.1 binding energy with LEU-704 (IIP), ASN-740 (IIS6) and ASN-1517 (IVS6) active site residues, respectively. The Zinc59486248 compounds bound -7.9 binding energy with THR-361 (IP) and MET-362 (IP) active site residues, respectively. The Zinc59494792 compounds bound -7.6 binding energy with THR-361 (IP) and ASN-740 (IIS6) active site residues, respectively. The Zinc67664832 compounds bound -7.2 binding energy with THR-361 (IP), ILE-360 (IP), SER-393 (IS6), ASN-1178 (IIIS6) and ASN-740 (IIS6) active site residues, respectively. The Zinc19796039 (AD) compounds bound -5.4 binding energy with ILE-360 (IP), SER-393 (IS6), ASN-1178 (IIIS6) and ASN-740 (IIS6) active site residues, respectively. The LCC model and best five screened compound interaction residues are shown in Table 7.2 and the graphical representation also shown in Fig. 7.8. The docking results showed that five compounds have best binding energies than AD compound (Table 7.2).

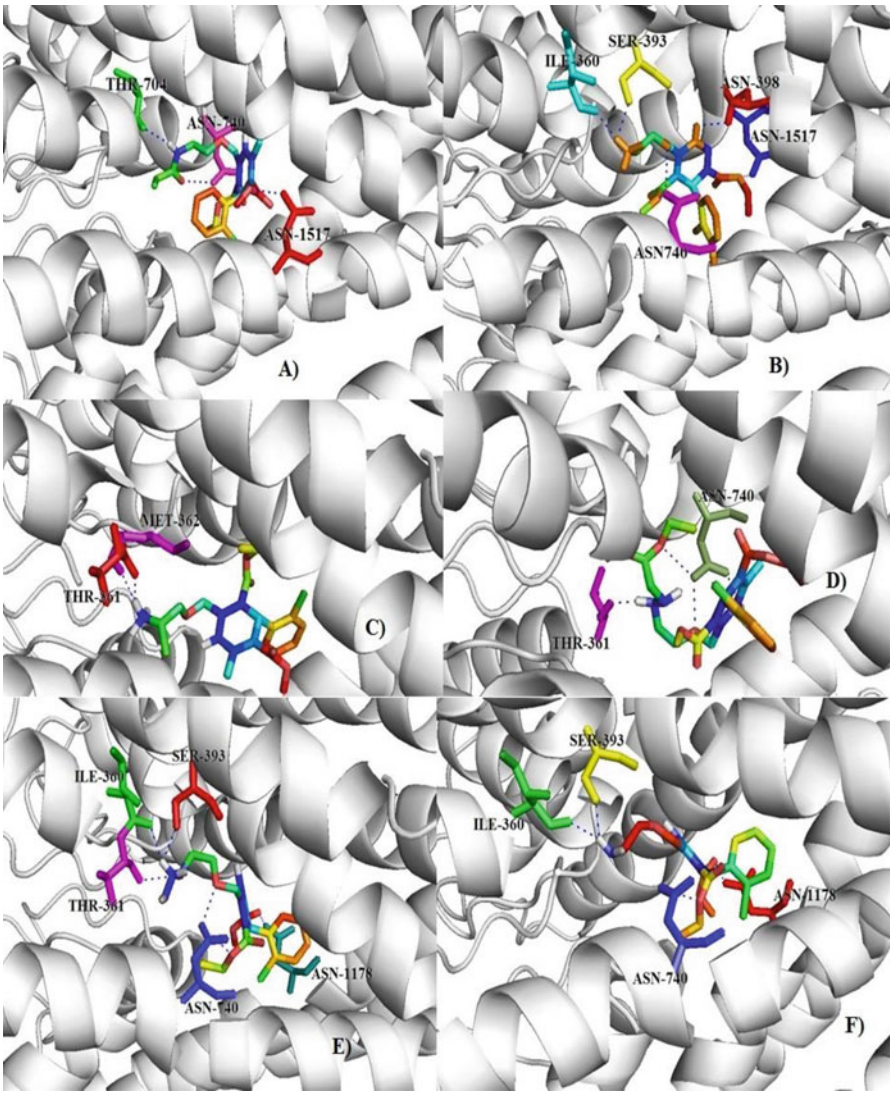

Fig. 7.8 Docked structures of Zinc20267861 (**a**), Zinc59347487 (**b**), Zinc59486248 (**c**), Zinc59494792 (**d**), Zinc67664832 (**e**) and AD (**f**) in model of LCC. DHPs are displayed as rainbow sticks, and key binding site residues are shown in *green*, *yellow*, *red*, *pink* and *blue*. Hydrogen bonds as represented with *dashed blue lines*

7.3.7 Analysis of Drug Likeness

All the compounds were tested for Lipinski "rule of 5", i.e. "drug-like" molecules have log P $\leq$5, molecular weight $\leq$500, number of hydrogen bond acceptors $\leq$10

Table 7.3 Molecular properties of compounds satisfying the Lipinski "rule of 5" by

Zinc ID	log S (moles/L)	Lipinski "rule of 5"			
		Molecular weight ≤500	log P ≤5	HB acceptors ≤10	HB donors ≤5
Zinc59347487	−4.06	429.10	3.14	5	4
Zinc20267861	−4.18	484.12	3.03	6	2
Zinc59486248	−4.90	423.17	3.61	5	4
Zinc59494792	−4.30	499.19	2.43	5	3
Zinc67664832	−4.24	437.16	2.21	9	3
Amlodipine (Query)	−4.19	409.15	3.46	7	4

and number of hydrogen bond donors ≤5 (Table 7.3). Lipinski rule of 5 is a rule of thumb for evaluating the drug likeness or determining whether a chemical compound with a certain pharmacological or biological activity has properties that would make it a likely orally active drug in humans. Results showed that five compounds, i.e. (Zinc59347487), (Zinc20267861), (Zinc59486248), (Zinc59494792) and (Zinc67664832), satisfied the Lipinski "rule of 5". Their respective drug likeness properties are shown in Table 7.3.

7.3.8 ADME Predicting Activity

Although Lipinski "rule of 5" describes the molecular properties important for a drug's pharmacokinetics in the human body, including its ADME, it does not predict if a compound is pharmacologically active. Therefore, pharmacokinetic properties and toxicities were predicted for all the four compounds using OSIRIS property explorer. Results of pharmacokinetic properties and toxicity analysis are shown in Table 7.4. Solubility and partition coefficient were calculated for pharmacokinetic property, whereas mutagenicity, tumorigenicity, irritation effect and risk of reproductive effect were predicted for toxicity study. To determine the hydrophilicity, log P value was predicted. It is suggestive that a high log P value is associated with poor absorption or permeation and it must be less than 5 (Vyas et al. 2013). Results showed that all the five compounds confirmed to this limit, and among the five compounds, Zinc67664832 has a better cLog P value than others (Table 7.3). In general, a poor solubility is associated with bad absorption, and the aqueous solubility (log S) of a compound significantly affects its absorption and distribution characteristics. Results showed that Zinc67664832 has a better log S value than others (Table 7.3). In order to consider the compound overall potential as a drug candidate, drug score is calculated which combines drug likeness, cLog P, TPSA, molecular weight and toxicity risk parameters as shown in Table 7.4. Drug score showed that the compounds, Zinc59347487 and Zinc20267861, have higher scores of 0.56 and 0.47 compared to the others.

Table 7.4 In silico ADMET prediction by OSIRIS property explorer

Properties	Zinc59347487	Zinc20267861	Zinc59486248	Zinc59494792	Zinc67664832	Amlodipine
Mutagenic	−	+	−	−	−	−
Tumorigenic	−	+	−	−	−	−
Irritant	−	−	−	−	−	−
Reproductive effective	−	+	−	−	−	−
cLog P	1.18	1.71	2.43	0.98	1.99	2.07
Solubility	−3.08	−3.10	−3.68	−3.53	−3.18	−3.30
MW	429.0	485.0	422.0	495.0	438.0	408.0
TPSA	99.88	102.9	99.88	135.20	123.0	99.88
Drug likeness	−7.9	−5.54	−4.60	−2.14	−4.94	−6.2
Drug score	0.56	0.47	0.37	0.38	0.38	0.39

7.4 Conclusions

The point of present study was to produce a pharmacophore model to recognize vitally assorted lead hits. The recognized hits may be utilized for creating novel and strong inhibitors for VP-3. A structure-based pharmacophore was created situated in light of the complex structure of VP-3 and leupeptin. The created pharmacophore model was utilized for the screening of PubChem database. The recognized hits were further assessed by docking, MD simulation and binding energy forecast. Subsequently, five lead hits were accounted for that satisfied all the criteria for the outline of compounds that may go about as great leads for advancement of novel, intense and structurally diverse compounds for VP-3 inhibition. From the binding mode, anticipated by docking, it was observed that there are some particular groups that mimic the binding method of leupeptin and fit well to active site area of VP-3. The five leads likewise demonstrated the best binding energies among screened compounds. The MD simulations for the VP-3 five lead docking complexes were performed to comprehend conformational dependability, structural flexibility and molecular dynamics of the interaction in physiological environmental condition. RMSD investigation demonstrated that the molecular system was exceptionally steady in all trajectories. Therefore, five leads are proposed as the best potential inhibitor to begin with investigation acceptance towards outlining against VP-3 inhibitors.

Acknowledgements Author, Madhu Sudhana Saddala, is especially grateful to University Grants Commission, New Delhi, for their financial assistance with the award of BSR Meritorious Fellowship. This work was carried out in DBT – Bioinformatics Infrastructure Facility (BIF), Department of Zoology, Sri Venkateswara University, Tirupati (BT/BI/25/001/2006).

References

Abdul W, Abid Ali S, Sattar R, Arif Lodhi M, Ul-Haq Z. A novel pharmacophore model to identify leads for simultaneous inhibition of anti-coagulation and anti-inflammatory activities of snake venom phospholipase A_2. Chem Biol Drug Des. 2012;79:431–41.

Beissenhirtz MK, Scheller FW, Viezzoli MS, Lisdat F. Engineered superoxide dismutase monomers for superoxide biosensor applications. Anal Chem. 2006;78(3):928–35.

Bhattacharya A, Wunderlich Z, Monleon D, Tejero R, Montelione GT. Assessing model accuracy using the homology modeling automatically (HOMA) software. Proteins. 2008;70(1):105–18.

Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science. 1991;253(5016):164–70.

Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993;2(9):1511–9.

Corpet F. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 1988;16 (22):10881–90.

Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 2006;34:W116–8.

Eswar N, Eramian D, Webb B, Shen MY, Sali A. Protein structure modeling with modeller. Methods Mol Biol. 2008;426:145–59.
Ginalski K. Comparative modeling for protein structure prediction. Curr Opin Struct Biol. 2006;16 (2):172–7.
Huber I, Wappl E, Herzog A, et al. Conserved Ca2+- antagonist-binding properties and putative folding structure of a recombinant high-affinity dihydropyridine-binding domain. Biochem J. 2000;347(3):829–36.
Jiang Y, Lee A, Chen J, et al. X-ray structure of a voltage-dependent K+ channel. Nature. 2003;423(6935):33–41.
Kocak M, Akcil E. The effects of chronic cadmium toxicity on the hemostatic system. Pathophysiol Haemost Thromb. 2006;35:411–6.
Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereo chemical quality of protein structures. J Appl Crystallogr. 1993;26:283–91.
Leeson P. Drug discovery: chemical beauty contest. Nature. 2012;481(15):455–6.
Lipinski CA, Lombardo F, Dominy BW, Feeny PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 1997;23:3–25.
Lipkind GM, Fozzard HA. KcsA crystal structure as framework for a molecular model of the Na+ channel pore. Biochemistry. 2000;39(28):8161–70.
Llorca O, Betti M, Gonźalez JM, Valencia A, Marquez AJ, Valpuesta JM. The three-dimensional structure of an eukaryotic glutamine synthetase: functional implications of its oligomeric structure. J Struct Biol. 2006;156(3):469–79.
Long SB, Campbell EB, MacKinnon R. Crystal structure of a mammalian voltage-dependent Shaker family K^+ channel. Science. 2005;309(5736):897–903.
Lovell SC, Davis IW, Arendall III WB, et al. Structure validation by Cα geometry: φ, and Cβ deviation. Proteins. 2003;50(3):437–50.
Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature. 1992;356(6364):83–5.
Lyne PD. Structure-based virtual screening: an overview. Drug Discov Today. 2002;7 (20):1047–55.
Mansfeld J, Gebauer S, Dathe K, Ulbrich-Hofmann R. Secretory phospholipase A2 from Arabidopsis thaliana: insights into the three-dimensional structure and the amino acids involved in catalysis. Biochemistry. 2006;45(18):5687–94.
Pappas RS, Polzin GM, Zhang L, Watson CH, Paschal DC, Ashley DL. Cadmium, lead, and thallium in mainstream tobacco smoke particulate. Food Chem Toxicol. 2006;44:714–23.
Perez-Reyes E, Wei X, Castellano A, Birnbaumer L. Molecular diversity of L-type calcium channels. “Evidence for alternative splicing of the transcripts of three non-allelic genes”. J Biol Chem. 1990;265(33):20430–6.
Petrey D, Honig B. Protein structure prediction: inroads to biology. Mol Cell. 2005;20(6):811–9.
Sali A, Potterton L, Yuan F, Van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. Proteins. 1995;23(3):318–26.
Satarug S, Scott Garrett H, Sens MA, Donald Sens A. Cadmium, environmental exposure, and health outcomes. Environ Health Perspect. 2010;118:182–90.
Shaldam MA, Elhamamsy MH, Esmat EA, El-Moselhy TF. 1,4-dihydropyridine calcium channel blockers: homology modeling of the receptor and assessment of structure activity relationship. ISRN Med Chem. 2014;2014(203518):1–14.
Shen M-Y, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci. 2006;15(11):2507–24.
Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins. 1993;17 (4):355–62.
Spassov VZ, Yan L, Flook PK. The dominant role of side-chain backbone interactions in structural realization of amino acid code. ChiRotor: a side-chain prediction algorithm based on side-chain backbone interactions. Protein Sci. 2007;16(3):494–506.

Spassov VZ, Flook PK, Yan L. LOOPER: a molecular mechanics-based algorithm for protein loop prediction. Protein Eng Des Sel. 2008;21(2):91–100.

Teilum K, Hoch JC, Goffin V, Kinet S, Martial JA, Kragelund BB. Solution structure of human prolactin. J Mol Biol. 2005;351(4):810–23.

Trott O, Olson AJ. AutoDockVina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31:455–61.

Vyas VK, Ghate M, Goel A. Pharmacophore modeling virtual screening docking and *in silico* ADMET analysis of protein kinase B (PKB beta) inhibitors. J Mol Graph Model. 2013;42 (13):17–25.

Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35:W407–10.

Wolf LK. PyRx. Chem Eng News. 2009;87:31.

Zhorov BS, Folkman EV, Ananthanarayanan VS. Homology model of dihydropyridine receptor: implications for L-type Ca^{2+} channel modulation by agonists and antagonists. Arch Biochem Biophys. 2001;393(1):22–41.