# Pharmacoinformatics-based identification of chemically active molecules against Ebola virus

*Md Ataul Islam[1,2], Tahir S. Pillay[*1,3]*

*[1]Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria and National Health Laboratory Service Tshwane Academic Division, Pretoria, South Africa.*
*[2]School of Health Sciences, University of Kwazulu-Natal, Westville Campus, Durban, South Africa.*
*[3]Division of Chemical Pathology, University of Cape Town, South Africa.*

Email IDs: MA Islam: ataul.islam80@gmail.com
TS Pillay: tspillay@gmail.com

*Correspondence should be addressed to T.S. Pillay, Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria, Private Bag X323, Arcadia, Pretoria, 0007
Email: tspillay@gmail.com
Phone: +27-123192114
Fax: +27-123283600

**Abstract**

Ebola is a dangerous virus transmitted by animals and humans and to date there is no curable agent for such a deadly infectious disease. In this study, pharmacoinformatics-based methods were adopted to find effective novel chemical entities against Ebola virus. A well predictive and statistical robust pharmacophore model was developed from known Ebola virus inhibitors collected from the literature. The model explained the significance of each of hydrogen bond acceptor and donor, and two hydrophobic regions for activity. The National Cancer Institute and Asinex (Antiviral library) databases were screened using the final validated pharmacophore model. Initial hits were further screened with a set of criteria and finally eight molecules from both databases were proposed as promising anti Ebola agents. Further molecular docking and molecular dynamics studies were carried out and it was found that the proposed molecules possessed capability to interact with amino residues of Ebola protein as well as retaining equilibrium of protein-ligand systems. Finally, the binding energies were calculated using MM-GBSA approach and all proposed molecules showed strong binding affinity towards the Ebola protein receptor.

**Keywords:** Ebola; pharmacoinformatics; pharmacophore; virtual screening; molecular docking; molecular dynamics

## Introduction

The deadly Ebola virus (EV) causes an acute, serious illness and haemorrhagic fever in humans and nonhuman primates resulting in high mortality rates (Leroy et al., 2009; Ray et al., 2004; Sullivan, Sanchez, Rollin, Yang, & Nabel, 2000). The World Health Organisation (WHO) reported an outbreak of Ebola virus disease (EVD) in West Africa in 2014(Briand et al., 2014), and there was an urgent response from international agencies and collaboration to control the outbreak. As per data by WHO, in 2016 about 28 500 people were infected by EV and approximately 28 500 died in the same year (*World Health Orginization (WHO)*, 2016). It is also reported that about 900 healthcare personnel working in the war zone of the outbreak contracted EVD and there were 513 deaths(Horton, 2014). The EV is a 19-kb, single-strand, negative-sense RNA virus which is a member of Filoviridae family and Mononegavirales order(Vetter et al., 2016). There are five species of EV with major variances in response to virulence and geographical distribution. It is elucidated that out of five species of the genus, four (Zaire ebolavirus, Sudan ebolavirus, Bundibugyo ebolavirus, and Taï Forest ebolavirus) are known to cause EVD in humans, whereas Reston ebolavirus is not pathogenic in humans. It is also reported that Zaire ebolavirus is the most pathogenic form for humans with lethality rates of up to 90%(Feldmann & Geisbert, 2011). The entire genome of the EV consists of seven genes that form a nucleoprotein, virion protein (VP) 35, VP40, glycoprotein, VP30, VP24, RNA-dependent RNA polymerase. The virus-like particles (VLP) are formed with help of the matrix protein VP40 and exposed on the surface due to presence of glycoproteins and subsequently presented to the host cell(Harty, Brown, Wang, Huibregtse, & Hayes, 2000; Noda et al., 2002). Furthermore, virus fusion and entry occur through a number of processes including a complex cascade of micropinocytosis–endocytosis(Aleksandrowicz et al., 2011; Nanbo et al., 2010; Saeed, Kolokoltsov, Albrecht, & Davey, 2010), endosome trafficking(Carette et al., 2011; Lee & Saphire, 2009) and proteolytic activation(Chandran, Sullivan, Felbor, Whelan, & Cunningham, 2005; Schornberg et al., 2006). The consequence of fusion and transmission of virions are internalization and the viral genome replication in the host cell.

A number of therapeutic options for EVD are being developed including vaccines(Feldmann & Geisbert, 2011), monoclonal antibodies(Olinger et al., 2012; Qiu et al., 2012), recombinant proteins(Geisbert et al., 2003; Smith et al., 2013), antibody–interferon (IFN) combinations(Qiu et al., 2013) and small interfering (si)RNA(Geisbert et al., 2010) and these have been successfully tested in the nonhuman models of virus infection. However, to date none of them is approved for the application in humans. As there are no therapeutic agents

available to treat or control such life-threatening EVD, it has become high level of public concern worldwide, and highlights the requirement to discover promising and effective therapeutic chemical agents targeting EV.

Discovery of therapeutically effective molecules is a complex, cost and time-intensive process. Integration of computational power with pharmaceutical research known as pharmacoinformatics plays an important role in enhancing the drug discovery pipeline and also reduces significant amount of cost as well as animal sacrifice. Pharmacoinformatics approaches including pharmacophore(Kim, Kim, & Seong, 2010), virtual screening(Kim et al., 2010), molecular docking(Meng, Zhang, Mezei, & Cui, 2011) and molecular dynamics(De Vivo, Masetti, Bottegoni, & Cavalli, 2016) studies have already proven their pivotal role in identifying promising and medicinally effective small molecules(Sliwoski, Kothiwale, Meiler, & Lowe, 2014). Pharmacophore is an abstract representation of the structures of small molecules or concept of receptor cavity of the macromolecule which allows one to discover structurally diverse promising compounds. Two strategies can be adopted to develop pharmacophore models *viz.* ligand-based and structure-based. In the ligand-based method, important mutual chemical functionalities are obtained from 3D structures of a set of known inhibitors/activators involved in binding interactions between the ligands and a precise macromolecular receptor site. The receptor-based pharmacophore approach is the identification of the complementary pharmacophoric features of the receptor site with the bound or docked small molecule and their three-dimensional associations, and a successive pharmacophore model assembly with selected features(Amaravadhi, Baek, & Yoon, 2014).

The virtual screening or extraction of small molecules from the molecular databases is considered a crucial technique in drug discovery research. With the availability of three-dimensional crystal structure of receptor molecules, molecular docking is one of the crucial step to screen the molecular databases. Molecular docking predicts the preferable orientation of the ligand inside the receptor cavity and subsequently calculates the binding affinity between ligand and protein. In order to explore dynamic behaviour of the ligand-protein complex, molecular dynamics (MD) simulations is one of the best options in which both receptor molecule and ligands are treated in a flexible manner and permitted to relax the binding site about the bound ligand. Moreover, MD can directly assess the consequence of explicit water molecules. In this study, we developed a well validated pharmacophore hypothesis using the *HypoGen* module(H. Li, Sutter, & Hoffmann, 2000) with key chemical features of a set of inhibitors that selectively hinder the Ebola and Marburg glycoprotein

(GP)-mediated infection of human cells(Yermolina, Wang, Caffrey, Rong, & Wardrop, 2011). The selected pharmacophore model was corroborated using a number of approved statistical approaches including Fischer's randomization, test set prediction and decoy set validation. The virtual screening of small molecular databases was performed using the final validated model. The primary goal of the virtual screening is to decrease the large virtual chemical space of small organic molecules to a reasonable number of the compounds that activate or inhibit the target for a maximum chance to lead to a drug candidate. The potential of the work is demonstrated by the credentials of eight promising molecules for the EVD described here. The molecular docking study was performed to explore the interactions between promising ligands from databases and catalytic amino acid residues. Finally, to explore constancy and detailed binding interactions of the final proposed inhibitors inside the protein, a molecular dynamics study was carried out.

## Materials and methods

In order to develop pharmacophore model Discovery Studio 2016 (DS) was used. The virtual screening and molecular docking studies were also carried out in DS. The DS is a widely used commercial software package in pharmaceutical research comprising a number of modules(Al-Balas et al., 2013; Chhabria, Brahmkshatriya, Mahajan, Darji, & Shah, 2012; Huang et al., 2012). The *3D QSAR pharmacophore generation* module of DS was adopted for pharmacophore generation in which a set of EV inhibitors and activity data were considered. Two modules of ligand-based pharmacophore method *viz.*, *HypoGen* and *HipHop* incorporated in DS. The *HipHop* module recognises the hypotheses conjoint in the 'active' set of compounds of training set but absence in the 'inactive' compounds. While, *HypoGen* module advances with hypotheses existing both in 'active' and 'inactive' compounds. In the current work, the *HypoGen* module was adopted to develop pharmacophore models in DS.

### *Dataset*

A dataset of 46 isoxazole analogues as inhibitors of Ebola GP-mediated cell entry with infectivity values were collected from literature(Yermolina et al., 2011). The *Training Set Generation* module of DS was used to identify the training and test sets compounds. The 2-dimensional structures of training set molecules are given in the Figure 1. The infectivity value of the compounds in the dataset possessed wide range of values from 3 to 421.

For simplicity the infectivity values have been converted into *pInfect* = log((1/*Infectivity*)x1000). The entire dataset was categorised into highly active (*pInfec* >

1.700, +++), moderately active (1.301 < *Infectivity* ≤ 1.700, ++) and least active (*Infectivity* ≤ 1.700, +) based on infectivity activity values. In order to identify the training and test set molecules the well knows strategies by Li et al.(H. Li, Sutter, & Hoffman, 1999) was used and afterward pharmacophore model developed from the training set. The guidelines enlightened that (a) compounds present in the set should have provide strong and transitory information with chemical features and range of the biological activity, (b) minimum 16 structurally diverse compounds should be considered in the set to confirm the statistical significance and elude chance correlation, (c) the most and least active molecules of the dataset must present in the set and (d) the activity of the compounds should have spanned at least 4 orders of magnitude. By following the above rules 30 compounds were considered in the training set including most and least active compounds. The remaining 16 compounds of dataset were taken as test set molecules (Figure 4) and further used to evaluate the performance of pharmacophore model. The *2D/3D visualizer module* of DS was used to generate the three-dimensional (3D) coordinates of the compounds. The modified CHARMm force field(Brooks et al., 1983; Momany & Rone, 1992) was implemented to correct the coordinates and energy minimization of each compound of the dataset.

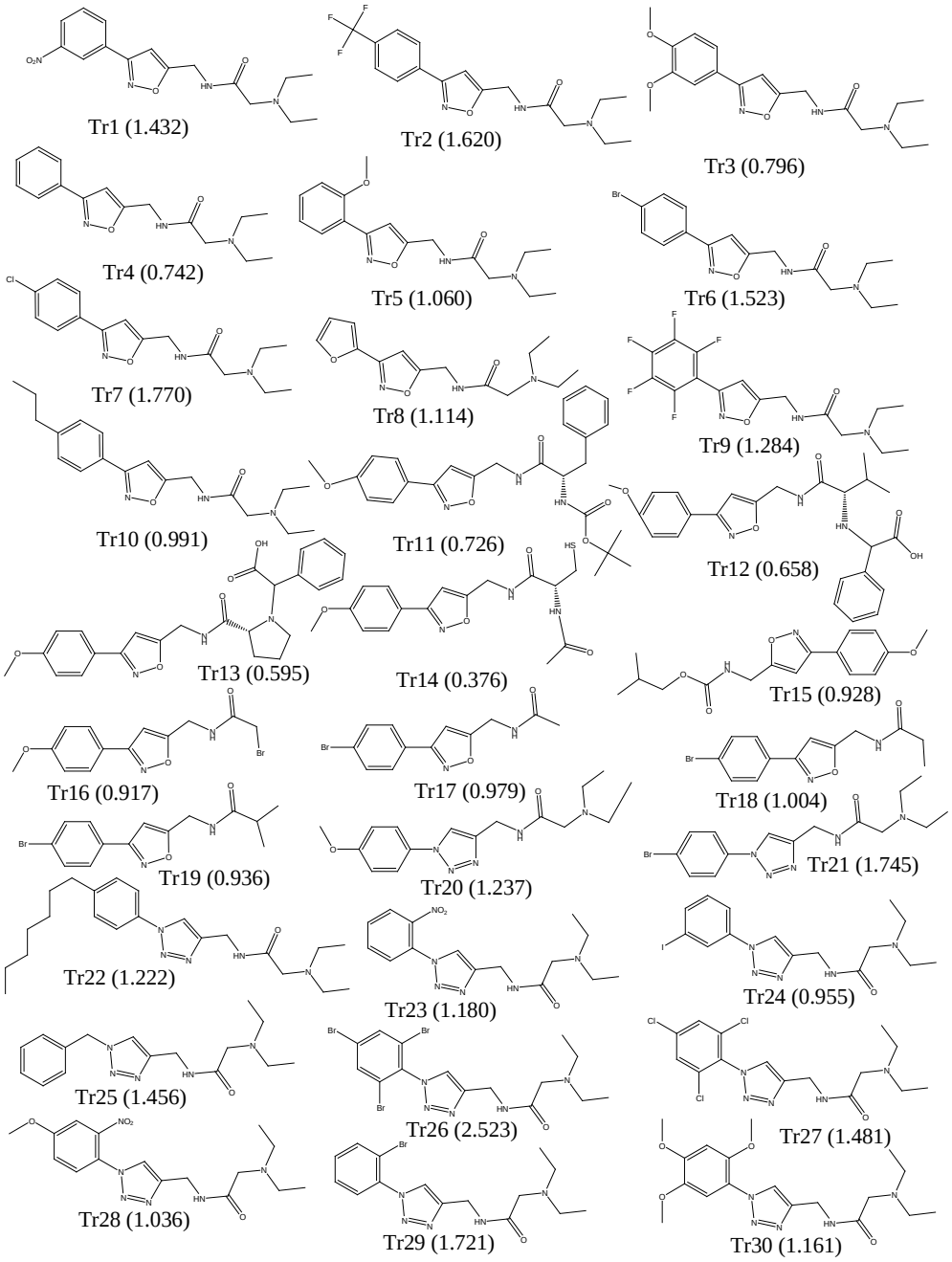


**Figure 1.** Two-dimensional structure of the training set compounds and the *pInfect* values are given in the parentheses

### *Pharmacophore model generation*

The *3D QSAR Pharmacophore Model Generation* module of DS was considered to generate pharmacophore hypotheses in which the molecules converted in 3D structure and conformations were created by *Cat-Conf* program. The BEST conformational generation method was used to find out various acceptable conformation of each molecules. The BEST method uses poling algorithm(Smellie, Teig, & Towbin, 1995) in which the molecular conformations are optimized by the rigorous energy minimization and consequently give

complete and improved coverage of conformational space. The algorithm also takes into consideration the arrangement of the chemical functionalities in space instead of simply the arrangement of atoms (Kristam, Gillet, Lewis, & Thorner, 2005). In order to identify the favourable pharmacophoric features from the highly active molecules the *Feature mapping* module of DS was used and subsequently best mapped features of the molecules considered as input pharmacophoric features for hypotheses development. With the help of conformers and chemical features the algorithm drives in two methods such as *HipHop* and *HypoGen*. In order to develop the hypotheses, only active molecules are considered by the *HipHop* approach, whereas *HypoGen* method takes both active and inactive compounds(Kristam et al., 2005). *HypoGen* approach gives the best ten hypotheses from the training set molecules by considering conformations of the molecules and pharmacophoric features through following steps: constructive, subtractive and optimization(Sadler, Cho, Ishaq, Chae, & Korach, 1998). The hypotheses are developed from those conjoint in the active molecules is the constructive step; in the subtractive step, the hypotheses are discarded from those fitted to the inactive molecules. Finally, in the optimization phase, remaining hypotheses are improved with regards to the score using the small perturbations(Kristam et al., 2005; H. Li, Sutter, & Hoffman, 2000). The best robust model was considered on the basis of high correlation coefficient (*R*), low root mean square deviation (RMSD), cost function analysis and good predictive ability.

***Validation of pharmacophore model***

Pharmacophore models validation is a crucial step to verify the predictive and robustness of the model. In the current work, the selected pharmacophore model from the training set was validated by five different approaches, (a) internal validation, (b) cost function analysis, (c) Fischer's randomization test, (d) test set prediction and (e) decoy set.

*Internal validation*

The leave-one out (LOO) cross-validation is an important approach to validate the model with the training set data. In this method, one compound from the training set was randomly deleted and the model developed with the remaining molecules with the parameters used from the original model and subsequently activity of the deleted compound predicted through the newly developed model. Similarly, the predicted activity of all molecules of the training set were recorded for further analysis. In order to assess the quality of the model two critical parameters of internal validation protocol, the LOO cross-validated correlation coefficient

($Q^2$) and error of estimation (*se*) were calculated based on estimated and experimental activity of the training set molecules. The model explained good predictive ability with high $Q^2$ (>0.5) and low *se*(Kubinyi, Hamprecht, & Mietzner, 1998). Further, the modified $r^2$ ($r^2_m$) reported by Roy *et al.*(Ojha, Mitra, Das, & Roy, 2011; K. Roy et al., 2012) was also calculated which measures the degree of deviation of the estimated activity from the experimental ones. As reported the model may be considered with $r^2_m$>0.5.

*Cost function analysis*

Several statistical parameters included spacing, uncertainty and weight variation were varied and given as input to generate the statistically robust model. The spacing parameter implies the minimum inter-features distance that may be tolerable in the final hypothesis. The order of magnitude explored by the hypothesis in which every pharmacophoric feature indicates some degree of magnitude of the molecule's activity. Generally 3.0 and 0.3 are the default values of spacing and weight variation respectively but it can vary from 4.0 to 1.0 and 1 to 2 respectively. The error of prediction represented by the uncertainty parameter which implies the standard deviation of the error cost. The default value of this parameter is 3 but in some cases it may vary from 1.5 to 4.0. Three cost factors *viz.*, weight cost, error cost, and configuration cost were minimized to explore the cost function of the hypotheses. The weight cost based on the weight variation and increases the weight of chemical feature in a hypothesis differ from input value. Error cost implies the root mean square deviation (RMSD) between the estimated and experimental biological activities for the training set molecules. The configuration cost depends on the entropy of hypothesis and it is reported that value should be <17 for an acceptable pharmacophore model. A fixed cost penalizes the complexity of the hypothesis space. The total cost is inclusive cost of all three costs factors of a hypothesis. Another cost factor known as null cost is generated during the null hypotheses development which is the postulation that there is no association in the estimated and the experimental activities. As per reports, the higher (>60) cost difference ($\Delta cost$ = null cost - total cost) specified that the hypothesis does not reflect a chance correlation.

*Fischer's randomization test*

The *CatsScramble*(H. Li et al., 2000) is one of the important approaches to verify the quality of the selected model and it is based on Fischer's randomization test. This approach confirms the robust relationship between the molecule and the biological activity of the training set compounds. For this purpose, the biological activities of the training set molecules were scrambled and the new pharmacophore hypotheses generated with same set of parameters as

the original hypothesis. If the statistical parameters of randomized hypothesis were found to be better than the original hypothesis then the original hypothesis may have developed by chance. Based on the statistical significance a number of spreadsheets are produced. The statistical significance is given by following equation.

$$Significance = [1-(1+a)/b] \quad (1)$$

Where, *a* implies the number of hypotheses developed with low total cost compared to the original hypothesis, and *b* indicates total number of *HypoGen* and random runs. For example, total number of spreadsheets are obtained as 19 (*b* = 20) with 95% confidence level when individual developed spreadsheet is submitted to *HypoGen* using the original restrictions as the initial run.

*Test set prediction*

External predictivity of the pharmacophore model outside the compounds involved in the model formation is a crucial and essential approach. In this approach, the test set molecules were mapped to the best pharmacophore model with help of *Ligand Pharmacophore Mapping* protocol in DS and estimated activity recorded. A number of statistical parameters included $R^2_{pred}$ (correlation coefficient) and $s_p$ (error of prediction)(Golbraikh & Tropsha, 2002; Mitra, Saha, & Roy, 2010) were calculated to verify the quality of prediction of the model. Further the modified $r^2$ [$r^2_{m(test)}$](P. P. Roy, Paul, Mitra, & Roy, 2009; P. P. Roy & Roy, 2008) value was calculated (threshold value=0.5).

***Virtual screening***

Pharmacophore-based virtual screening can be used to identify novel promising compounds from the small molecule databases that can interact with the receptor site cavity to block or trigger activity. The validated pharmacophore model was adopted to screen the NCI (National Cancer Institute) (https://cactus.nci.nih.gov/ncidb2.2/) and Asinex antiviral (http://www.asinex.com/libraries-html/) databases to retrieve novel molecular entities for EVD. The NCI and Asinex (anti viral) databases contain 265,242 and 8,722 compounds respectively. Several measures were used to accomplish final potential molecules for the EVD. Subsequently, the molecular docking and molecular dynamics studies were performed to explore binding interactions and stability between ligands and protein molecules respectively.

***Molecular docking***

The molecular docking is one of the important and effective approaches to screen the promising molecules from the molecule database. In order to identify promising anti-EV molecules, several criteria were executed on initial hits. Among these criteria molecular docking was one of them. In order to select the molecules with better dock score than most active compound (**Tr26** in Figure 1)of the dataset and to comprehend how the promising drug-like virtual hits molecules bind to the receptor of Ebola protein molecule, the molecular docking was carried out with the help of the *LigandFit* protocol of DS. The *LigandFit* protocol primarily identifies the cavity to recognise and select the portion of the protein as the receptor site followed by docking the molecules in the selected site. The 3D coordinates of crystal structure of the EVD were downloaded from RCSB Protein Data Bank (RCSB-PDB) for the molecular docking study. In order to select the best protein molecules the receptor size, resolution and date of submission were explored and finally PDB ID: 5JQB(Zhao et al., 2016) was considered for further study. Protein and ligands were prepared using the *Prepare Protein* and *Prepare Ligand* tools of DS respectively, while the CHARMm force field(Vanommeslaeghe et al., 2010) was implemented for minimization of both protein and ligand. For preparation of the protein the 'Build Loop' and 'Protonate' parameters were given to 'True' and, dielectric constant, pH, ionic strengths and energy cut-off were kept as default. In case of preparation of ligands, the 'Change ionization', 'Generate Tautomers' and 'Generate isomers' were considered as 'False', and 'Generate Coordinates' was selected as '3D'. The volume occupied by the co-crystallize ligand was used to find out the binding site for the molecular docking. False positive output of molecular docking study were overcome by validating the selected parameters. In this regard, the co-crystal small molecule present in the receptor site was regenerated and afterward docked into the same receptor cavity of EV (PDB ID: 5JQB) protein. The binding interactions of the best pose were analysed and further the same pose superimposed on the co-crystal small molecule. The RMSD value from the superimposed structures was calculated. As per the report RMSD <2Å explains the production of comparable conformation to co-crystal ligand(Taha et al., 2011). In the current work RMSD <2Å was observed between docked and co-crystal ligand and hence the same parameters as used in the co-crystalized docking study were considered in molecular docking studies of proposed compounds retrieved from databases. To explore the binding interactions and dock score values, the best ten poses for each docked-ligand were analysed to explore the binding interactions and dock score values

***Molecular dynamics***

It is difficult to explore static complexes to completely understand protein-ligand complexes as one needs to grasp dynamic information generated by simulating their internal motions or dynamic processes. For this purpose, MD simulations were performed on the best docked poses between EV protein and the most active compound, and also with final screened compounds from the databases. MD simulation study was performed in the Amber14(D.A. Case et al., 2012) using the ff99SB force field[35]. The explicit TIP3P water model(Lindorff-Larsen et al., 2010) box was used and minimum distance of 8 Å between the solute and each face was considered. The general amber force field (GAFF)(J. Wang, Wang, Kollman, & Case, 2006; J. Wang, Wolf, Caldwell, Kollman, & Case, 2004) with Antechamber was used to parametrize the ligands. With the help of Propka3.1(Olsson, Sondergaard, Rostkowski, & Jensen, 2011; Sondergaard, Olsson, Rostkowski, & Jensen, 2011) missing hydrogen atoms were predicted and added. $Na^+$ ions were used to neutralize the system with the help of the Leap program of Amber14. Propka allocates the protonation state of the amino acid in a protein based on empirical *pKa* prediction of titratable residues while considering its microenvironment. In order to treat the long range electrostatic force the particle mesh Ewald (PME) method was used(Harvey & De Fabritiis, 2009), through space and vdW (van der Waals) cut off of 12 Å. Before starting the MD runs, the systems were partially minimized with a restrained force of 500 kcal/mol on the solute molecule using 750 cycles of the steepest descent, followed by 2500 cycles of the conjugate gradient method. Furthermore, the conjugate gradient method was used to minimize the system for 1500 cycles. After this the systems were heated gradually from 0 to 300 K with a harmonic restraint of 10 kcal/mol to hold the solute fixed. In order to control the temperature using a collision frequency of 1.0 $ps^{-1}$ the Langevin dynamics was used. The systems were equilibrated for 2 ns at 300 K and pressure was kept constant of 1 bar. With the help of the SHAKE algorithm[38] bonds were constrained involving hydrogen atoms. The MD study was run for 40 ns with a time step of 2 fs using GPU version of Amber 14(Gotz et al., 2012). The MD trajectories were analysed for RMSD, RMSF and Rg for which the ptraj and cpptraj(Roe & Cheatham, 2013) module of Amber14 was used.

***MM-GBSA analysis***

The Molecular Mechanics Poisson-Boltzmann Surface Area (MMGBSA) method(Genheden & Ryde, 2015) in Amber14 was used to quantitatively measure the binding strength between

the receptor and ligands. The binding free energy can be expressed as the difference between bound and unbound states of two solvated protein molecules as per following equation.

$$[PL]_{aq} \rightleftharpoons [L]_{aq} + [P]_{aq} \quad (2)$$

where, [PL], [L] and [P] are the concentration of protein-ligand complex, ligand and protein respectively. The binding energy calculation based on equation (2) is not ideal for practical purposes(J. Wang, Deng, & Roux, 2006). Therefore the binding energy calculation may be divided and rewritten as given below.

$$\Delta G_{(bind,aq)} = \Delta G_{(bind,vaccum)} + \Delta G_{(bind,complex)} - \Delta G_{(bind,ligand)} - \Delta G_{(bind,receptor)} \quad (3)$$

The electrostatic component of the solvation free energy in MM-GB method can be calculated by solving Generalized Born (GB) equation and an empirical term for hydrophobic contribution is added.

$$\Delta G_{aq} = \Delta G_{GB} + \Delta G_{hydrophobic} \quad (4)$$

The average interactions energies of final screened molecules and most active compound (**Tr26** in Figure 1) of the dataset, and receptor were calculated from 50ns of MD trajectories.

**Results and discussion**

The pharmacophore model from the training set was generated using the *HypoGen* module of DS. The training set molecules are given in Figure 1 with their *pInfect* value in the parentheses. The *Feature mapping* protocol of DS was used to find common chemical features in the molecules of the dataset and were considered as inputs to the *3D QSAR pharmacophore generation* module.

Minimum and maximum feature values were set to ‘0’ and ‘5’ respectively. Based on excellent statistical parameters the top ten hypotheses were considered for further analysis. The statistical parameters along with correlation coefficient were noted and are depicted in Table 1. Debnath’s analysis(Debnath, 2002, 2003) was used to select the best hypothesis which explains that model can be considered for further evaluation if it possessed low RMSD, high correlation coefficient, low cost value and high cost difference. For a robust model the overall cost of the hypothesis should be distant from the null cost and adjacent to the fixed cost. It is stated that differences between null and total cost (*Δcost*) in the range of 40–60 bits suggests the possibility of the predictive correlation of 75–90%, while the *Δcost* ≥ 60 bits defines the hypothesis and has a correlation probability of more than 90%(Sakkiah, Thangapandian, John, Kwon, & Lee, 2010). In the current work, the cost difference of *Hypo*

*1* (Table 1) was found to be 76.657 which clearly explained that selected hypothesis has more than 90% chance of being able to select EV inhibitors.

**Table 1.** Statistical parameters of top ten hypotheses

| Hypo No. | Spacing | [1]Unc. | [2]Wt. Var. | [3]*R* | Rmsd | Costs | | | | | Features |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Total | Null | Fixed | [4]Δ | [5]Config. | |
| 1 | 1.5 | 1.5 | 0.3 | 0.882 | 1.150 | 113.276 | 189.933 | 90.161 | 76.657 | 16.731 | a, d, 3xp |
| 2 | 1.5 | 1.5 | 0.3 | 0.855 | 1.263 | 117.617 | 189.933 | 90.161 | 72.316 | 16.731 | a, p, 2xr |
| 3 | 1.5 | 1.5 | 0.3 | 0.836 | 1.338 | 120.091 | 189.933 | 90.161 | 69.842 | 16.731 | 2xa, p, r |
| 4 | 1.5 | 1.5 | 0.3 | 0.799 | 1.465 | 126.304 | 189.933 | 90.161 | 63.629 | 16.731 | 2xa, p, r |
| 5 | 1.5 | 1.5 | 0.3 | 0.784 | 1.514 | 131.913 | 189.933 | 90.161 | 58.020 | 16.731 | d, 3xp, r |
| 6 | 1.5 | 1.5 | 0.3 | 0.764 | 1.574 | 132.113 | 189.933 | 90.161 | 57.820 | 16.731 | a, 2xp, r |
| 7 | 1.5 | 1.5 | 0.3 | 0.788 | 1.502 | 132.301 | 189.933 | 90.161 | 57.632 | 16.731 | 2xa, 3xp |
| 8 | 1.5 | 1.5 | 0.3 | 0.782 | 1.519 | 132.445 | 189.933 | 90.161 | 57.488 | 16.731 | a, 3xp, r |
| 9 | 1.5 | 1.5 | 0.3 | 0.781 | 1.523 | 132.616 | 189.933 | 90.161 | 57.317 | 16.731 | a, 3xp, r |
| 10 | 1.5 | 1.5 | 0.3 | 0.772 | 1.551 | 133.197 | 189.933 | 90.161 | 56.736 | 16.731 | d, 3xp, r |

[1]Uncertainty; [2]Weight variation; [3]Correlation coefficient; [4](Null cost – Total cost); [5]Configuration cost

From the Table 1 it can be seen that the best hypothesis (*Hypo1*) was found to have a high correlation coefficient value ($R$ = 0.882), which clarifies robustness of the selected model. The total cost and fixed cost were found to be 113.276 and 90.161 respectively, along with the *Δcost* of 76.657. The top 10 hypotheses were selected to analyse the results and it was found that only *Hypo1* possessed high *R*, less RMSD, highest *Δcost* and minimum error values in contrast to other hypotheses. Consequently, *Hypo1* was selected as the best pharmacophore model for further analyses.

The best model is portrayed in Figure 2 and revealed the importance of one of each hydrogen bond acceptor (HB) and HB donor along with two hydrophobic regions in three dimensional space. *Hypo1* mapped with the most active molecule (**Tr26** in Figure 1) of the dataset and inter-feature distances is depicted in Figure 2.

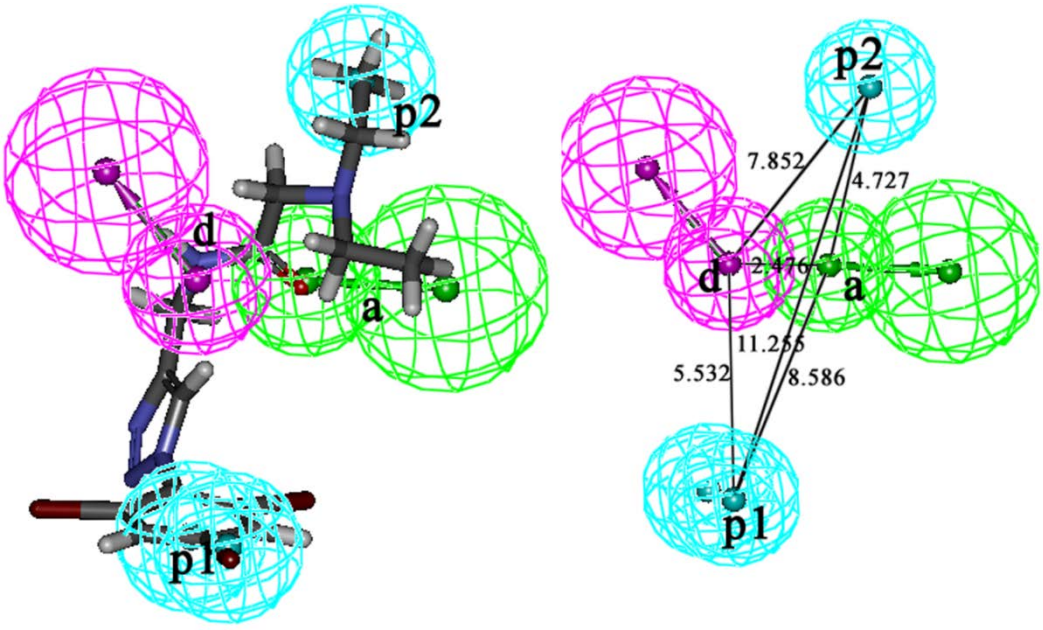


**Figure 2.** Best pharmacophore model mapped with most compound and inter-feature distances

The Figure 2 explained that oxo group present in the alkyl chain attached to the five member ring behaves as HB acceptor. The amine group present in between five membered ring and oxo group was found to act as HB donor. The ethyl group attached to the nitrogen atom and the benzene ring in the molecular system imparts the hydrophobicity of the molecule. Therefore the best model suggested that HB acceptor and donor along with hydrophobic regions are crucial to form potential interactions between the ligand and catalytic amino residues present in the active cavity of Ebola enzyme. The inter-feature distances may play crucial roles for the arrangement of the key pharmocophoric features being a potent and saver chemical entity for the treatment of Ebola. Hence the new and potent chemical entities can be designed using the above crucial pharmacophoric features with derived inter-feature distances.

***Validation of pharmacophore model***

It is obligatory to validate the pharmacoinformatics models to check robustness of models and their wide applications to new or unknown compounds. In our study the best selected model was validated using a) internal, b) cost value analysis, c) test set, d) Fischer's randomization test and e) decoy set.

*Internal validation*

The infectivity value of the training set molecules were predicted by fitting all molecules on the best selected model. The experimental and predicted activity along with error values which represents ratio between observed and predicted *pInfect* are given in Table 2 and Figure 3.

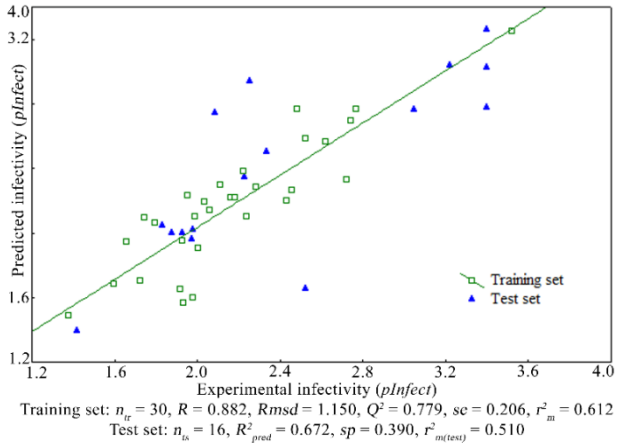


**Figure 3.** Observed vs predicted infectivity values (*pInfect*) of training and test set as per *Hypo1*

The error value reflects closeness between observed and predicted infectivity values. Table 2 clearly explained that error value of all training set molecules are in considerable range. The cross-validated correlation coefficient ($Q^2$) was calculated based on procedure explained in Materials and Methods section. The $Q^2$ and *se* value of best model were found to be 0.779 and 0.206 respectively. Further, $r^2_m$ and $\Delta r^2_m$ were calculated and found the values of 0.612 and 0.127 respectively. Therefore, high $Q^2$ and $r^2_m$, and low *se* and $\Delta r^2_m$ undoubtedly explained that selected model is robust in nature and has excellent predictive ability of training set molecules.

**Table 2.** Observed, predicted activities values of the training and test sets obtained using the pharmacophore model *Hypo 1*

| Mol. No. | *pInfect* | | Error | *pInfect* Scale | | Mol. No. | *pInfect* | | Error | *pInfect* Scale | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | [1]Obs | [2]Pred | | [1]Obs | [2]Pred | | [1]Obs | [2]Pred | | [1]Obs | [2]Pred |
| **Tr1** | 1.432 | 1.196 | +1.197 | ++ | + | **Tr24** | 0.955 | 1.230 | -0.776 | + | + |
| **Tr2** | 1.620 | 1.565 | +1.035 | ++ | ++ | **Tr25** | 1.456 | 1.261 | +1.155 | + | + |
| **Tr3** | 0.796 | 1.065 | -0.747 | + | + | **Tr26** | 2.523 | 2.250 | +1.121 | +++ | +++ |
| **Tr4** | 0.742 | 1.097 | -0.676 | + | + | **Tr27** | 1.481 | 1.765 | -0.839 | ++ | +++ |
| **Tr5** | 1.060 | 1.143 | -0.927 | + | + | **Tr28** | 1.036 | 1.196 | -0.866 | + | + |
| **Tr6** | 1.523 | 1.585 | -0.961 | ++ | ++ | **Tr29** | 1.721 | 1.332 | +1.292 | + | + |
| **Tr7** | 1.770 | 1.770 | +1.000 | +++ | +++ | **Tr30** | 1.161 | 1.222 | -0.950 | + | + |
| **Tr8** | 1.114 | 1.297 | -0.859 | + | + | **Ts1** | 1.252 | 1.947 | -0.643 | + | +++ |
| **Tr9** | 1.284 | 1.283 | +1.001 | + | + | **Ts2** | 2.398 | 2.261 | +1.061 | +++ | +++ |
| **Tr10** | 0.991 | 1.103 | -0.898 | + | + | **Ts3** | 0.879 | 1.002 | -0.877 | + | + |
| **Tr11** | 0.726 | 0.703 | +1.033 | + | + | **Ts4** | 1.523 | 0.656 | +2.323 | ++ | + |
| **Tr12** | 0.658 | 0.944 | -0.697 | + | + | **Ts5** | 2.222 | 2.041 | +1.089 | +++ | +++ |
| **Tr13** | 0.595 | 0.684 | -0.870 | + | + | **Ts6** | 1.229 | 1.352 | -0.909 | + | ++ |
| **Tr14** | 0.376 | 0.488 | -0.770 | + | + | **Ts7** | 0.419 | 0.397 | +1.056 | + | + |
| **Tr15** | 0.928 | 0.951 | -0.976 | + | + | **Ts8** | 2.398 | 1.778 | +1.349 | +++ | +++ |
| **Tr16** | 0.917 | 0.652 | +1.406 | + | + | **Ts9** | 0.975 | 0.963 | +1.013 | + | + |
| **Tr17** | 0.979 | 0.599 | +1.634 | + | + | **Ts10** | 0.830 | 1.052 | -0.789 | + | + |
| **Tr18** | 1.004 | 0.906 | +1.108 | + | + | **Ts11** | 0.928 | 1.003 | -0.925 | + | + |
| **Tr19** | 0.936 | 0.567 | +1.651 | + | + | **Ts12** | 1.086 | 1.750 | -0.621 | + | + |
| **Tr20** | 1.237 | 1.102 | +1.123 | + | + | **Ts13** | 2.398 | 2.032 | +1.180 | +++ | +++ |
| **Tr21** | 1.745 | 1.695 | +1.029 | +++ | ++ | **Ts14** | 0.979 | 1.022 | -0.958 | + | + |
| **Tr22** | 1.222 | 1.381 | -0.885 | + | + | **Ts15** | 1.337 | 1.508 | -0.887 | ++ | ++ |
| **Tr23** | 1.180 | 1.222 | -0.966 | + | + | **Ts16** | 2.046 | 1.768 | +1.157 | +++ | +++ |

[1]Observed; [2]Predicted

*Cost value analysis*

In order to assess the pharmacophore models, *HypoGen* gives a number of parameters during the hypotheses generation including different cost functions and RMSD. The cost functions included *Δcost*, total cost, null cost, fixed cost and configuration cost. The *Δcost* was found to be 76.657 and this explains that the model was not developed by chance. As per elucidation that for a statistically robust model, difference between fixed and total cost should be as close as possible and in the current findings it was found to be 23.115. For a consistent and well validated pharmacophore model the configuration cost should be <17. For the *Hypo1* (Table 1) the configuration cost was observed to be 16.731. The RMSD between observed and predicted infectivity for *Hypo1* was obtained as 1.150. Therefore the cost value analysis indisputably favour towards the robustness of the model.

*Test set prediction*

Checking the predictive ability of molecules not involved in model generation is one of the crucial approaches of validation. For this purpose, the infectivity values of test compounds (Figure 4) were predicted using the *Ligand Pharmacophore Mapping* protocol of DS and depicted in Table 2 and Figure 3. Both observed and predicted infectivity values were converted into logarithm value [*pInfect* = log(1/infectivity)x10000]. The test set compounds were divided according to their infectivity values i.e. highly active (*pInfect* > 2.700), moderately active (2.301 < *pInfect* ≤ 2.700) and least active (*pInfect* ≤ 2.301). On analysis of test set predicted *pInfect* values, it was observed that one and two least active molecule were underestimated as moderately and highly active respectively. One moderately active molecule was overestimated as least active molecule. The rest of the test set molecules were estimated within their range. Moreover the correlation coefficient ($R^2_{pred}$) between observed and predicted infectivity and error of prediction (*sp*) were calculated. The value of $R^2_{pred}$ and *sp* were obtained as 0.672 and 0.390 respectively. Therefore the above finding clearly explains that the selected hypothesis is proficient enough to predict the infectivity of molecules outside the compounds used in model development.

Additionally the selected model was taken into consideration to check better predictive ability and calculated $r^2_{m(test)}$ and $\Delta r^2_{m(test)}$ from the predicted and experimental infectivity values. These parameters explain how the predicted infectivity values are contiguous to the actual experimental values because high $R^2_{pred}$ value of test set cannot always put forward a low residual between actual and predicted activity data. Both $r^2_{m(test)}$ and $\Delta r^2_{m(test)}$ were calculated and values found as 0.510 and 0.033 respectively. According to reports for good predictive

model values $r^2_{m(test)}$ and $\Delta r^2_{m(test)}$ should be >0.5 and <0.2 respectively. Therefore the above findings can clearly explain that selected hypothesis (*Hypo1*) has sufficient predictive capability for new molecules. It is always recommended to validate the *in silico* models with large test set. Due to unavailability of suitable compounds sometimes test set may be smaller. Although several successful pharmacophore models were validated using the smaller test set compounds(Fu et al., 2017; Kandakatla & Ramakrishnan, 2014; Madan et al., 2018; F. Wang & Chen, 2013) but it may give false positive results. Therefore the test set is not only sufficient to validate the pharmacophore model and subsequent use for virtual screening. Other validation protocols must be employed and checked the superiority of the model. The current study was used a small test set to validate the pharmacophore model and subsequent several validation approaches as well.

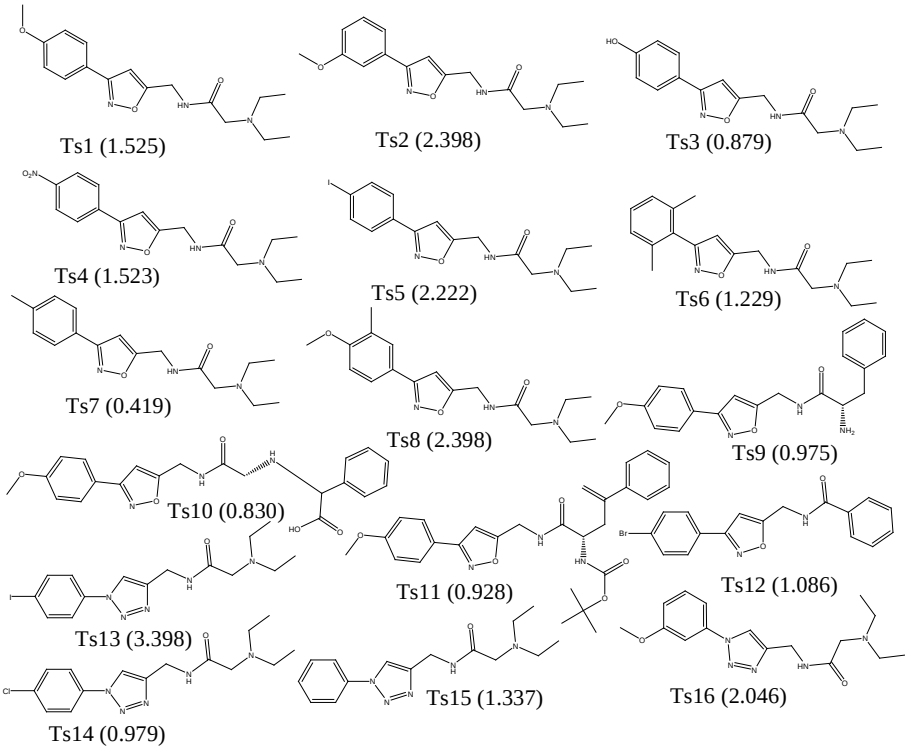


**Figure 4.** Two-dimensional structure of test compounds. The *pInfect* values are given in the parentheses

*Fischer's randomization test*

The Fischer's randomization test was carried to check the quality of the selected hypothesis by assigning a particular confidence level. The selected hypothesis (*Hypo1*) in the study was considered for 95% confidence level in which observed infectivities of training set compounds were randomized and produced into 19 random spreadsheets by creating a hypothesis on each spreadsheet. Lowest total cost and highest correlation coefficient of each

spreadsheet were collected and depicted in Figure 5 and Table 3 respectively. Significance of the model was evaluated by using the equation (1).

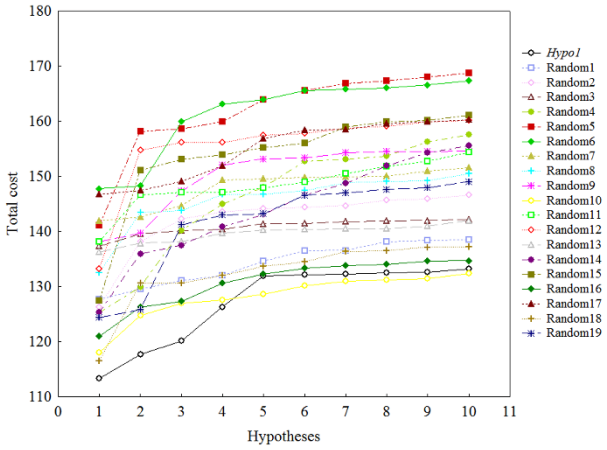


**Figure 5.** Total cost of 19 randomized runs and *Hypo1* in Fischer's randomization test

**Table 3:** Highest correlation coefficient of 19 randomized runs and *Hypo1* in Fischer's randomization test

| **Hypotheses** | ***R*** | **Hypotheses** | ***R*** | **Hypotheses** | ***R*** | **Hypotheses** | ***R*** |
|---|---|---|---|---|---|---|---|
| *Hypo1* | 0.882 | Random5 | 0.657 | Random10 | 0.827 | Random15 | 0.721 |
| Random1 | 0.746 | Random6 | 0.580 | Random11 | 0.719 | Random16 | 0.810 |
| Random2 | 0.770 | Random7 | 0.616 | Random12 | 0.721 | Random17 | 0.610 |
| Random3 | 0.697 | Random8 | 0.687 | Random13 | 0.690 | Random18 | 0.838 |
| Random4 | 0.782 | Random9 | 0.746 | Random14 | 0.766 | Random19 | 0.787 |

From Figure 5 and Table 3 it can be seen that not a single randomized run perceived predictive capability comparable to or improved over that of *Hypo1*. Highest correlation coefficient among all 19 trials was found to be 0.838 while average value of 0.725. Both highest and average correlation values of 19 trials are much lower compare to *Hypo1*. Further, the total cost of all 19 trails were found be higher than *Hypo1*. Therefor on analysis of correlation coefficient and total costs of *Hypo1* and 19 trials it can be conclude that selected hypothesis is superior in nature and was not generated by chance.

*Decoy set*

A set of 450 decoy molecules were retrieved from ZINC database using the DecoyFinder2.0 to explore the screening capability of the selected model. Total 9 active molecules from the dataset were combined with decoy molecules and amalgamated set used to screen through the *Hypo1*. The screening results explained accuracy of 0.660 and, the true positive (TP), false

positive (FP), True negative (TN) and false negative (FN) as 9, 157, 293 and 0 respectively. The ROC plot was generated by plotting true positive rate of active vs. false positive rate of decoys and depicted in Figure 6.

The area under curve (AUC) was calculated and the value found to be 0.960. The ROC curve and AUC value undoubtedly explained that the selected hypothesis has sufficient potential to discriminate between the active and inactive molecules, and more biased towards selection of active molecules rather inactive. The enrichment factor (EF 1%) was found to be 40.80 which acknowledged active compound very well. The average BEDROC value was found to be 0.620, which indicates that the top hits were not only enriched with active molecules but also ranked higher than inactive. Therefore above findings and explanations clearly indicated that pharmacophoric features present in the *Hypo1* are well capable to screen active molecules from the molecular databases.

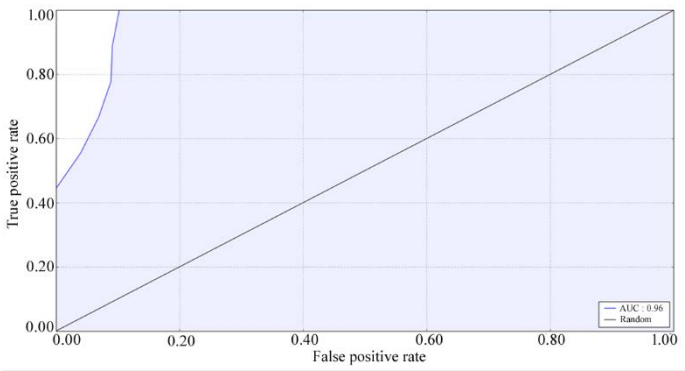


**Figure 6.** ROC curve for Hypo 1, derived from true positive rate of actives vs. false positive rate of decoys

*Virtual screening*

The validated pharmacophore model was used to screen NCI and Asinex antiviral databases to identify promising molecules for therapeutic application in EVD. The '*Search Database*' protocol of '*Pharmacophore*' module of DS package was adopted to screen both databases. The 'Search Method' and 'Limit Hits' were set to 'Best' and 'All' respectively.

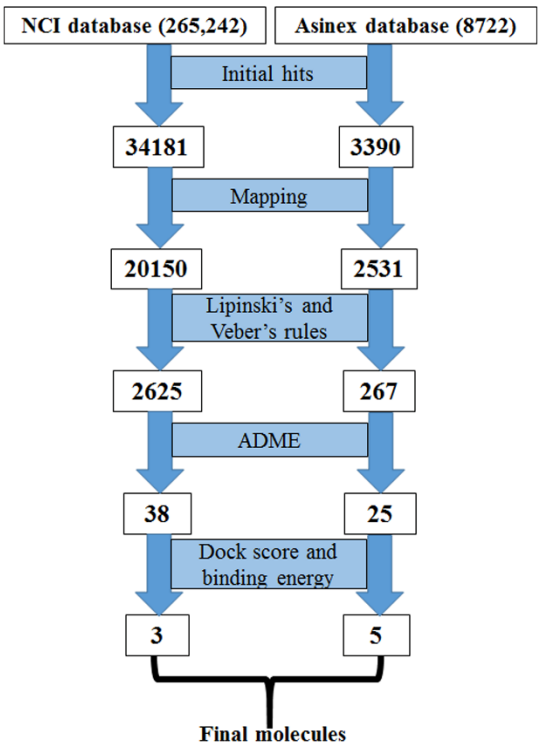


**Figure 7.** Flow diagram of virtual screening of NCI and Asinex databases

The best model was able to retrieve initial hits of 34181 and 3390 compounds from NCI and Asinex respectively. Initial hits were further sorted out separately using a number of criteria. Flow diagram of the virtual screening of both databases is given in Figure 7. The estimated activities of the compounds were calculated after mapping on the best model using the "*Ligand Pharmacophore Mapping*" protocol of DS with "Maximum Omitted Feature" set to '0'. Furthermore, Lipinski's rule of Five and Veber's rule were checked. The compounds were only considered if those possessed activity within the range of activity of the dataset, and passed both Lipinski's rule of Five and Veber's rules. It was observed that 2625 and 267 compounds from NCI and Asinex respectively satisfied the above criteria and were considered for further screening. The ADMET descriptors were calculated using the *ADMET Descriptor* protocol of DS. The human intestinal absorption (HIA), aqueous solubility and blood brain barrier (BBB) were analysed and found that 38 and 25 compounds from NCI and Asinex respectively show good absorption, aqueous solubility and penetration values. Finally, above compounds along with most active molecule (**Tr26** in Figure 1) of the dataset were taken into consideration for molecular docking study. Molecular docking study revealed that 14 and 13 molecules from NCI and Asinex respectively were docked successfully inside

the receptor cavity. Docking analysis portrayed that most active compounds gave dock scores and binding energy of 39.603 and -105.480 respectively.

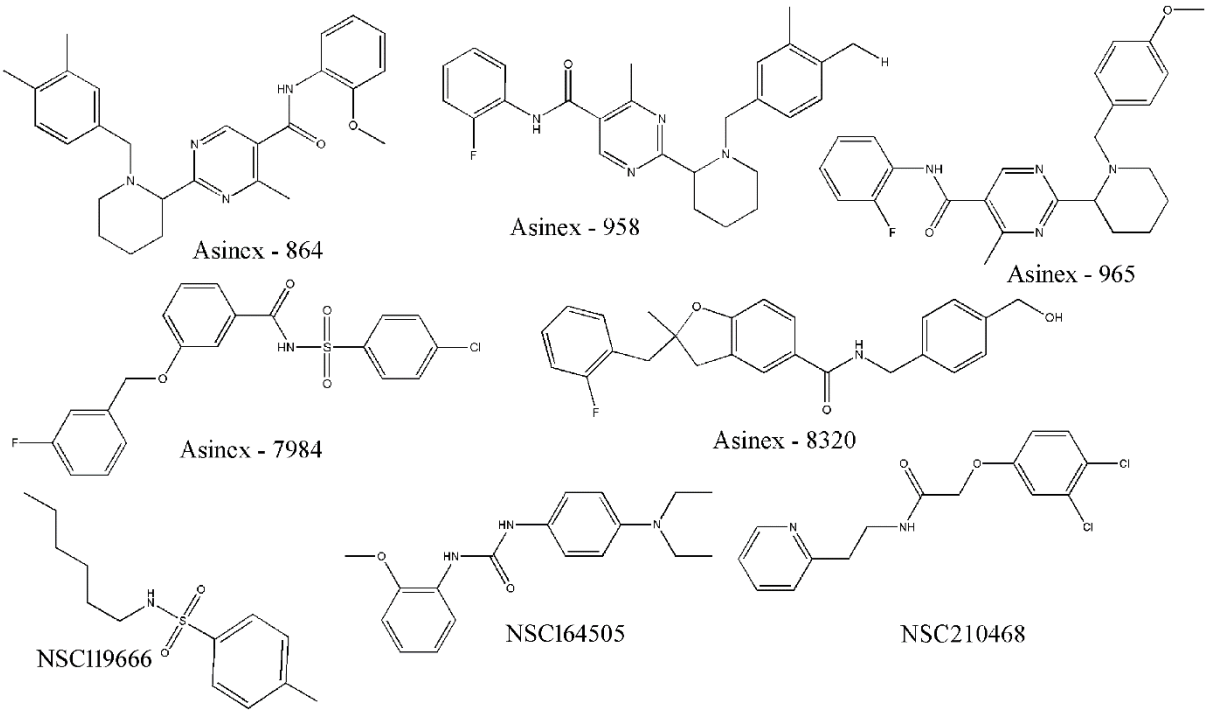


**Figure 8.** Final proposed molecules from NCI (**NSC164505**, **NSC210468** and **NSC119666**) and Asinex (**Asinex - 864**, **Asinex - 958**, **Asinex - 965**, **Asinex - 2835** and **Asinex - 8320**)

All docked screened compounds were analysed and considered only those give dock score more than 39.603 and binding energy less than -105.480. It was observed that 3 and 5 molecules from NCI (**NSC164505**, **NSC210468** and **NSC119666**) and Asinex (**Asinex – 864**, **Asinex – 958**, **Asinex – 965**, **Asinex – 7984** and **Asinex – 8320**) respectively satisfied above criteria and hence these eight molecules (Figure 8) proposed as promising anti Ebola molecules.

**Molecular docking**

Molecular docking is one of the crucial approaches to determine favourable orientations of the small molecules at the receptor site cavity. The final screened eight molecules were considered for molecular docking and the crystal structure of the protein molecule (PDB ID: 5JQB) downloaded from RCSB Protein Data Bank. In order to validate the docking study, self-docking approach was used in which bound ligand of crystal structure re-docked at the same active site and original bound conformation superimposed on the best docked pose to calculate RMSD value. As per report, the RMSD <0.2 Å between original ligand and docked

pose validate the docking procedure. In the current study, RMSD value was found to be 0.134 Å which directs that selected parameters in the molecular docking study were validated.

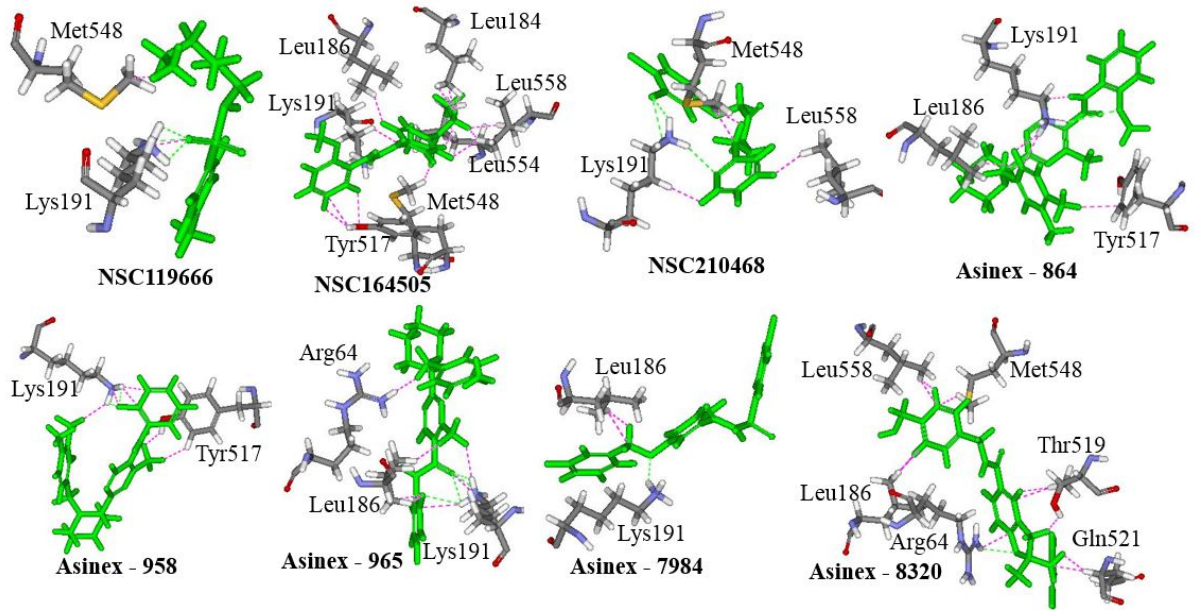


**Figure 9.** Binding modes of final screened compounds from NCI and Asinex databases

The best pose of final screened molecules are depicted in Figure 9. The docking study revealed that Lys191 was found to be critical amino residue to form number of hydrogen bond interactions to all screened compounds except **NSC164505**, **Asinex – 864** and **Asinex – 8320**. It was also observed that Arg64 and Leu186 established hydrogen bonds with **Asinex – 965** and **Asinex – 8320** respectively. The bump interaction which is defined as any bonding interactions except hydrogen bonding plays an important role to stabilize the ligand-receptor complex. The proposed molecules formed a number of bump interactions with catalytic amino residues. Leu186 formed several bump interactions with **NSC164505**, **Asinex – 864**, **Asinex -964** and **Asinex -7984**. Lys191 and Tyr517 successfully established bump interactions with **NSC164505**, **Asinex – 864** and **Asinex -958**. **NSC164505**, **NSC210468** and **Asinex – 8320** were connected with both Met548 and Leu558 through bump interactions. **NSC119666** was found to be crucial to form a number of binding interactions with Lys191 and Met548 in the form of hydrogen bonds and bump interactions. Three critical hydrogen bonds were formed between Lys191 and **NSC119666**, while each of Lys191 and Met548 established bump interactions with **NSC119666**. Beyond the above, several catalytic amino residues including Arg64, Leu184, Thr519 and Gln521 were anticipated to create number of bump interactions with the proposed promising compounds. The above observations

undoubtedly indicate that the final proposed molecules are capable of forming number of interactions with the catalytic amino residues present at the active site of the EV protein.

**Molecular dynamics**

Molecular dynamics study of 50ns time span was performed on complexes of final screened molecules and most active compound (**Tr26** in Figure 1) with the Ebola protein. The RMSD, RMSF and radius of gyration from the MD trajectories were analysed to explore dynamic stability of the protein-ligand complex.

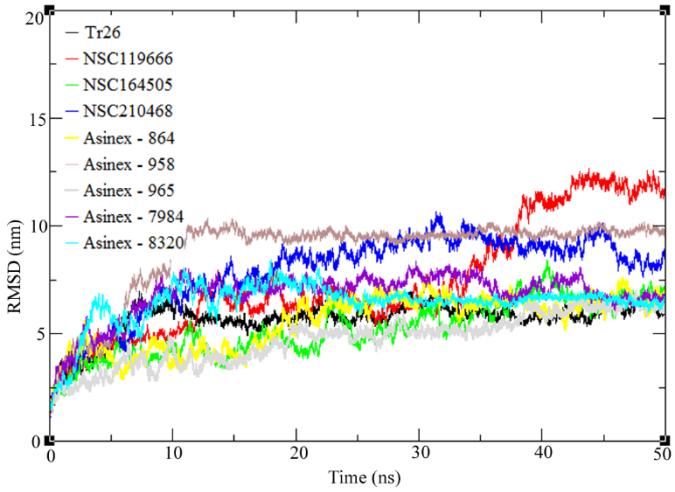


**Figure 10.** RMSD vs time of protein-ligand complexes

The RMSD values of protein backbone were calculated and plotted in the Figure 10. It was observed that after 10ns of time span all systems were equilibrated except for the complex with **NSC119666** and **NSC210468**. The complex with **NSC119666** initially increases the RMSD value and after about 12ns it achieved stability but again the compactness was disturbed during about 34ns to 40ns. Afterwards the system accomplished constancy. The complex between Ebola protein and **NSC210468** showed a slow rise in the RMSD and finally equilibrated after about 34ns. Detailed analysis of the individual system explained that complex with **Tr26**, **NSC164505**, **Asinex – 965**, **Asinex – 7984** and **Asinex – 8320** achieved equilibration at the RMSD range of 6.0 to 6.5nm. RMSD trajectories of Ebola protein complex with **NSC119666**, **NSC210468**, **Asinex – 864** and **Asinex – 958** were found to be at steady state at 10.4, 8.0, 7.0 and 10.0nm respectively. It can be worth noting that **NSC119666** fluctuated much higher when compared to others and this explained that the flexibility of this compound is much higher with respect to others. The average RMSD values of entire trajectories can give idea of the fluctuation of the backbone of the system during the MD

simulation. The mean RMSD values were found to be 5.666, 7.442, 5.203, 7.861, 5.494, 8.875, 4.713, 6.775 and 6.429 nm for protein-ligand system with **Tr26**, **NSC164505**, **NSC210468**, **NSC119666**, **Asinex – 864**, **Asinex – 958**, **Asinex – 965**, **Asinex – 7984** and **Asinex – 8320** respectively.

The role of individual amino residues of the protein-ligand complex system RMSF were calculated and depicted in the Figure 11. The figure of RMSF explained that all most all systems shared similar distribution with some differences in the case of **NSC119666**. RMSF trajectories revealed that fluctuations of amino residues were found to be higher in around the Leu20, Gly60, Pro92, Val110, Glu170, Gly240, Ile270, Trp310 and Asp400. The possible reason of deviation of these amino acids may be region of flexible loop and lack of interactions with corresponding ligands.

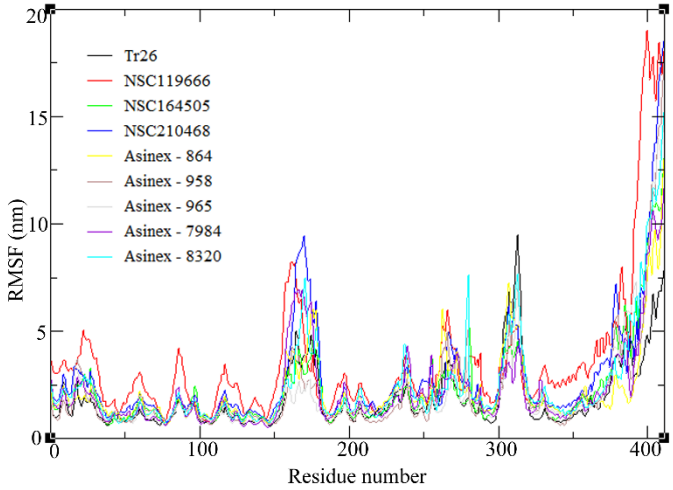


**Figure 11.** RMSF vs number of residues of complexes of most active compound of the dataset and final screened molecules

On detailed analysis of trajectories, the difference between maximum and minimum, and average value of RMSF were found to be 8.932 and 1.893; 18.256 and 3.539; 12.458 and 2.224; 17.772 and 2.800; 12.365 and 2.206; 16.904 and 2.143; 14.148 and 2.140; 11.152 and 2.234; and, 14.514nm and 2.438nm for the system of **Tr26**, **NSC164505**, **NSC210468**, **NSC119666**, **Asinex – 864**, **Asinex – 958**, **Asinex – 965**, **Asinex – 7984** and **Asinex – 8320** respectively.

In order to explore the compactness of protein-ligand complex systems the radius of gyration were calculated and plotted in the Figure 12. The trajectories explained that all the systems except **NSC119666** and **NSC164505** were almost equilibrated after about 25ns of time span.

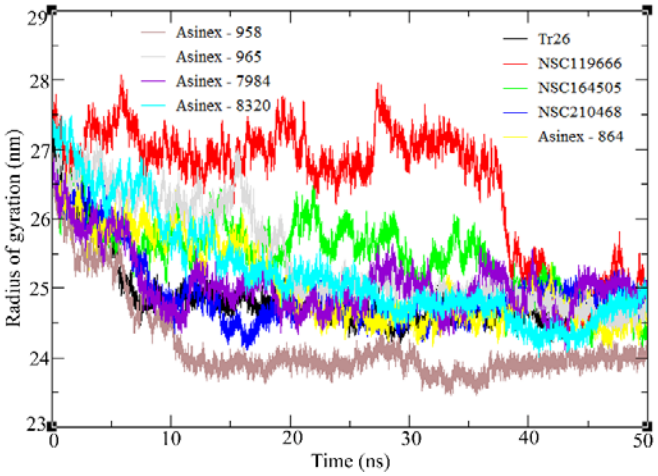


**Figure 12.** Radius of gyration of Cα atoms of EV protein over the simulation time

Initially the trend of **NSC119666** was found to stable but it reduces around 39ns and finally equilibrated after 40ns. Similarly, system of **NSC164505** was fluctuating up to 38ns and thereafter achieved steady state. All the systems were found to be equilibrated within the Rg value of 24 to 26nm. Therefore the above observations of MD simulation study indicated that all molecules well equilibrated and correlate with findings of the molecular docking study.

### Binding free energy using MM-GBSA

The binding affinity towards the receptor cavity of respective ligands were calculated from the 50ns trajectories using the MM-GBSA approach in AMBER14. The total binding free energy ($\Delta G_{bind}$), van der Waals interactions energy ($\Delta G_{vdW}$), electrostatics energy ($\Delta G_{ele}$), polar solvation energy ($\Delta_{Gele,sol}$) and nonpolar solvation energy ($\Delta G_{nonpol,sol}$) are given in the Table 4.

**Table 4.** Binding free energies and its components for **Tr26** and final screened compounds

| Complex | $\Delta G_{vdW}$ | $\Delta G_{ele}$ | $\Delta G_{ele,sol}$ | $\Delta G_{nonpol,sol}$ | $\Delta G_{bind}$ |
|---|---|---|---|---|---|
| **Tr26** | -49.093 | -0.249 | 8.974 | -5.927 | -46.295 |
| **NSC119666** | -22.804 | -35.842 | 41.116 | -3.843 | -21.373 |
| **NSC164505** | -46.499 | -4.406 | 16.430 | -6.057 | -40.533 |
| **NSC210468** | -37.792 | -3.370 | 13.376 | -5.209 | -32.996 |
| **Asinex – 864** | -54.917 | -4.680 | 17.253 | -6.650 | -48.997 |
| **Asinex – 958** | -46.135 | -1.050 | 14.457 | -6.042 | -38.773 |
| **Asinex – 965** | -52.072 | -2.716 | 16.111 | -6.840 | -45.517 |
| **Asinex – 7984** | -41.698 | -29.184 | 44.179 | -6.166 | -32.869 |
| **Asinex – 8320** | -41.961 | -6.428 | 20.450 | -5.823 | -33.748 |

The $\Delta G_{bind}$ data (Table 4) of all systems explained that each and every molecule was successful in binding to the receptor site of Ebola protein. The complexes with **NSC164505**, **Asinex – 864**, **Asinex - 958** and **Asinex – 965** were found to be possess similar affinity in respect to the $\Delta G_{bind}$ as complex with **Tr26**. The systems of **NSC210468**, **Asinex – 7984** and **Asinex – 8320** were found to show comparable binding affinity with $\Delta G_{bind}$ value of -32.996, -32.869 and -33.748 kcal $mol^{-1}$ respectively. The lowest binding energy was found for complex with **NSC119666** with -21.373 kcal $mol^{-1}$. Individual components of the binding free energies including $\Delta G_{vdW}$, $\Delta G_{ele}$, $\Delta G_{ele,sol}$ and $\Delta G_{nonpol,sol}$ were analysed to explore detailed binding process. In can be observed that $\Delta G_{vdW}$ and $\Delta G_{nonpol,sol}$ mainly contributed to the protein-ligand complexes and are responsible for the imbedded hydrophobic regions of the molecules.

From the Table 4 it can be seen that for all the systems $\Delta G_{vdW}$ were higher than $\Delta G_{bind}$ indicating the major contribution of $\Delta G_{vdW}$ on binding affinity. With exception of the complex with **NSC119666,** in all systems $\Delta G_{vdW}$ was greater when compared to $\Delta G_{ele}$. In case of **NSC116999** system the $\Delta G_{vdW}$ was much lower and $\Delta G_{ele}$ relatively higher compared to other systems, leading to achieve lower binding energy. The value of $\Delta G_{nonpol,sol}$ was found to contribute consistently to binding energies with some fluctuations. The above discussion clearly suggest that the final screened molecules might have potential to be therapeutic inhibitors for the EV.

**Conclusions**

A set of isoxazole analogues as inhibitors of Ebola GP-mediated cell entry with infectivity were used to developed ligand-based pharmacophore model. Among several developed pharmacophore models, initially best one was selected based on statistical parameters. Further selected models was validated using test set, cost functional analysis, Fischer's randomization test and decoy set validation. The best model described the importance of HB acceptor and donor along with two hydrophobic regions in the three dimensional space. Furthermore, the model was used to screen the NCI and Asinex databases to identify promising molecules for the EVD. Several criteria were used to sort out the initial hits of about 38000 compounds and finally eight compounds (**NSC164505**, **NSC210468**, **NSC119666**, **Asinex – 864**, **Asinex – 958**, **Asinex – 965**, **Asinex – 7984** and **Asinex – 8320**) proposed as potential inhibitors for EV. The binding interactions between proposed molecules and catalytic amino residues were explored and we observed that screened

molecules were capable of forming a number of interactions. The complexes between EV and ligands were subjected to molecular dynamics study of 50ns of time span. The RMSD, RMSF and Rg values clearly demonstrated that all molecules were well equilibrated and correlated with findings of the molecular docking study. Finally, binding affinity of the molecules inside the receptor cavity of EV were analysed. The data of binding free energies and its components substantiated the evidence that the proposed molecules represent possible promising chemical entities for therapeutic application in EVD.

**Acknowledgment**

MA Islam and TS Pillay were funded by the National Research Foundation (NRF), South Africa Innovation post-doctoral fellowship scheme. Authors are thankful to the CHPC (www.chpc.ac.za) for providing computational resources and tools.

**Conflict of Interest**

The authors declares that they have no conflicts of interest with the contents of this article.

**References**

Al-Balas, Q. A., Amawi, H. A., Hassan, M. A., Qandil, A. M., Almaaytah, A. M., & Mhaidat, N. M. (2013). Virtual lead identification of farnesyltransferase inhibitors based on ligand and structure-based pharmacophore techniques. *Pharmaceuticals (Basel), 6*(6), 700-715. doi: 10.3390/ph6060700

Aleksandrowicz, P., Marzi, A., Biedenkopf, N., Beimforde, N., Becker, S., Hoenen, T., . . . Schnittler, H. J. (2011). Ebola virus enters host cells by macropinocytosis and clathrin-mediated endocytosis. *The journal of infectious diseases, 204 Suppl 3*, S957-967. doi: 10.1093/infdis/jir326

Amaravadhi, H., Baek, K., & Yoon, H. S. (2014). Revisiting de novo drug design: receptor based pharmacophore screening. *Current topics in medicinal chemistry, 14*(16), 1890-1898

Briand, S., Bertherat, E., Cox, P., Formenty, P., Kieny, M. P., Myhre, J. K., . . . Dye, C. (2014). The international Ebola emergency. *The new england journal of medicine, 371*(13), 1180-1183. doi: 10.1056/NEJMp1409858

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry, 4*(2), 187–217

Carette, J. E., Raaben, M., Wong, A. C., Herbert, A. S., Obernosterer, G., Mulherkar, N., . . . Brummelkamp, T. R. (2011). Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature, 477*(7364), 340-343. doi: 10.1038/nature10348

Chandran, K., Sullivan, N. J., Felbor, U., Whelan, S. P., & Cunningham, J. M. (2005). Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. *Science, 308*(5728), 1643-1645. doi: 10.1126/science.1110656

Chhabria, M. T., Brahmkshatriya, P. S., Mahajan, B. M., Darji, U. B., & Shah, G. B. (2012). Discovery of novel acyl coenzyme a: cholesterol acyltransferase inhibitors: pharmacophore-based virtual screening, synthesis and pharmacology. *Chemical biology & drug design, 80*(1), 106-113. doi: 10.1111/j.1747-0285.2012.01384.x

D.A. Case, T.A. Darden, T.E. Cheatham, I., C.L. Simmerling, J. Wang, R.E. Duke, . . . Kollman, P. A. (2012). AMBER (Version 12). University of California, San Francisco.

De Vivo, M., Masetti, M., Bottegoni, G., & Cavalli, A. (2016). Role of Molecular Dynamics and Related Methods in Drug Discovery. *Journal of medicinal chemistry, 59*(9), 4035-4061. doi: 10.1021/acs.jmedchem.5b01684

Debnath, A. K. (2002). Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of Mycobacterium avium complex dihydrofolate reductase. *Journal of medicinal chemistry, 45*(1), 41-53

Debnath, A. K. (2003). Generation of predictive pharmacophore models for CCR5 antagonists: study with piperidine- and piperazine-based compounds as a new class of HIV-1 entry inhibitors. *J Med Chem, 46*(21), 4501-4515. doi: 10.1021/jm030265z

Feldmann, H., & Geisbert, T. W. (2011). Ebola haemorrhagic fever. *Lancet, 377*(9768), 849-862. doi: 10.1016/S0140-6736(10)60667-8

Fu, Y., Sun, Y. N., Yi, K. H., Li, M. Q., Cao, H. F., Li, J. Z., & Ye, F. (2017). 3D Pharmacophore-Based Virtual Screening and Docking Approaches toward the Discovery of Novel HPPD Inhibitors. *Molecules, 22*(6). doi: 10.3390/molecules22060959

Geisbert, T. W., Hensley, L. E., Jahrling, P. B., Larsen, T., Geisbert, J. B., Paragas, J., . . . Vlasuk, G. P. (2003). Treatment of Ebola virus infection with a recombinant inhibitor of factor VIIa/tissue factor: a study in rhesus monkeys. *Lancet, 362*(9400), 1953-1958. doi: 10.1016/S0140-6736(03)15012-X

Geisbert, T. W., Lee, A. C., Robbins, M., Geisbert, J. B., Honko, A. N., Sood, V., . . . Maclachlan, I. (2010). Postexposure protection of non-human primates against a lethal Ebola virus challenge with RNA interference: a proof-of-concept study. *Lancet, 375*(9729), 1896-1905. doi: 10.1016/S0140-6736(10)60357-1

Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery, 10*(5), 449-461. doi: 10.1517/17460441.2015.1032936

Golbraikh, A., & Tropsha, A. (2002). Beware of q2! *J Mol Graph Model, 20*(4), 269-276

Gotz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., & Walker, R. C. (2012). Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of chemical theory and computation, 8*(5), 1542-1555. doi: 10.1021/ct200909j

Harty, R. N., Brown, M. E., Wang, G., Huibregtse, J., & Hayes, F. P. (2000). A PPxY motif within the VP40 protein of Ebola virus interacts physically and functionally with a ubiquitin ligase: implications for filovirus budding. *Proceedings of the national academy of sciences U S A, 97*(25), 13871-13876. doi: 10.1073/pnas.250277297

Harvey, M. J., & De Fabritiis, G. (2009). An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware. *Journal of chemical theory and computation, 5*(9), 2371-2377. doi: 10.1021/ct900275y

Horton, R. (2014). Ebola: protection of health workers on the front line. *Lancet, 384*(9942), 470. doi: 10.1016/S0140-6736(14)61319-2

Huang, D., Zhu, X., Tang, C., Mei, Y., Chen, W., Yang, B., . . . Huang, W. (2012). 3D QSAR pharmacophore modeling for c-Met kinase inhibitors. *Medicinal chemistry, 8*(6), 1117-1125

Kandakatla, N., & Ramakrishnan, G. (2014). Ligand Based Pharmacophore Modeling and Virtual Screening Studies to Design Novel HDAC2 Inhibitors. *Advanced bioinformatics, 2014*, 812148. doi: 10.1155/2014/812148

Kim, K. H., Kim, N. D., & Seong, B. L. (2010). Pharmacophore-based virtual screening: a review of recent applications. *Expert opinion on drug discovery, 5*(3), 205-222. doi: 10.1517/17460441003592072

Kristam, R., Gillet, V. J., Lewis, R. A., & Thorner, D. (2005). Comparison of conformational analysis techniques to generate pharmacophore hypotheses using catalyst. *Journal of chemical information and modeling, 45*(2), 461-476. doi: 10.1021/ci049731z

Kubinyi, H., Hamprecht, F. A., & Mietzner, T. (1998). Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. *Journal of medicinal chemistry, 41*(14), 2553-2564. doi: 10.1021/jm970732a

Lee, J. E., & Saphire, E. O. (2009). Ebolavirus glycoprotein structure and mechanism of entry. *Future virology, 4*(6), 621-635. doi: 10.2217/fvl.09.56

Leroy, E. M., Epelboin, A., Mondonge, V., Pourrut, X., Gonzalez, J. P., Muyembe-Tamfum, J. J., & Formenty, P. (2009). Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector Borne Zoonotic Disease, 9*(6), 723-728. doi: 10.1089/vbz.2008.0167

Li, H., Sutter, J., & Hoffman, R. (1999). An automated system for generating 3D predic-tive pharmacophore models. In O. F. Guner (Ed.), *Pharmacophore Perception, Development, and Use in Drug Design* (pp. 173–189). La Jolla,CA: International University Line.

Li, H., Sutter, J., & Hoffman, R. (2000). *Pharmacophore Perception, Development, and Use in Drug Design*. California: International University Line.

Li, H., Sutter, J., & Hoffmann, R. (2000). *Pharmacophore Perception, Development, and Use in Drug Design*. La Jolla, CA: International University Line.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins, 78*(8), 1950-1958. doi: 10.1002/prot.22711

Madan, K., Verma, A., Paliwal, S., Yadav, D., Sharma, S., & Sharma, M. (2018). Pharmacophore Modeling and Database Mining to Identify Novel Lead Compounds Active Against the Disease Stage of Trypanosomiasis in the Central Nervous System. *International journal of nutrition, pharmacology, neurological diseases, 8*(1), 16-31. doi: 10.4103/ijnpnd.ijnpnd_53_17

Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design, 7*(2), 146-157

Mitra, I., Saha, A., & Roy, K. (2010). Pharmacophore mapping of arylamino-substituted benzo[b]thiophenes as free radical scavengers. *Journal of molecular modeling, 16*(10), 1585-1596. doi: 10.1007/s00894-010-0661-4

Momany, F. A., & Rone, R. (1992). Validation of the general purpose QUANTA ®3.2/CHARMm® force field. *Journal of computational chemistry, 13*(7), 888–900

Nanbo, A., Imai, M., Watanabe, S., Noda, T., Takahashi, K., Neumann, G., . . . Kawaoka, Y. (2010). Ebolavirus is internalized into host cells via macropinocytosis in a viral glycoprotein-dependent manner. *PLoS pathogens, 6*(9), e1001121. doi: 10.1371/journal.ppat.1001121

Noda, T., Sagara, H., Suzuki, E., Takada, A., Kida, H., & Kawaoka, Y. (2002). Ebola virus VP40 drives the formation of virus-like filamentous particles along with GP. *Journal of virology, 76*(10), 4855-4865

Ojha, P. K., Mitra, I., Das, R. N., & Roy, K. (2011). Further exploring rm2 metrics for validation of QSPR models. *Chemometrics and intelligent laboratory systems, 107*(1), 194–205

Olinger, G. G., Jr., Pettitt, J., Kim, D., Working, C., Bohorov, O., Bratcher, B., . . . Zeitlin, L. (2012). Delayed treatment of Ebola virus infection with plant-derived monoclonal antibodies provides protection in rhesus macaques. *Proceedings of the national academy of sciences U S A, 109*(44), 18030-18035. doi: 10.1073/pnas.1213709109

Olsson, M. H., Sondergaard, C. R., Rostkowski, M., & Jensen, J. H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of chemical theory and computation, 7*(2), 525-537. doi: 10.1021/ct100578z

Qiu, X., Audet, J., Wong, G., Pillet, S., Bello, A., Cabral, T., . . . Kobinger, G. P. (2012). Successful treatment of ebola virus-infected cynomolgus macaques with monoclonal antibodies. *Science translational medicine, 4*(138), 138ra181. doi: 10.1126/scitranslmed.3003876

Qiu, X., Wong, G., Fernando, L., Ennis, J., Turner, J. D., Alimonti, J. B., . . . Kobinger, G. P. (2013). Monoclonal antibodies combined with adenovirus-vectored interferon significantly extend the treatment window in Ebola virus-infected guinea pigs. *Journal virology, 87*(13), 7754-7757. doi: 10.1128/JVI.00173-13

Ray, R. B., Basu, A., Steele, R., Beyene, A., McHowat, J., Meyer, K., . . . Ray, R. (2004). Ebola virus glycoprotein-mediated anoikis of primary human cardiac microvascular endothelial cells. *Virology, 321*(2), 181-188. doi: 10.1016/j.virol.2003.12.014

Roe, D. R., & Cheatham, T. E., 3rd. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of chemical theory and computation, 9*(7), 3084-3095. doi: 10.1021/ct400341p

Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N., & Kabir, H. (2012). Comparative studies on some metrics for external validation of QSPR models. *Journal of chemical information and modeling, 52*(2), 396-408. doi: 10.1021/ci200520g

Roy, P. P., Paul, S., Mitra, I., & Roy, K. (2009). On two novel parameters for validation of predictive QSAR models. *Molecules, 14*, 1660-1701

Roy, P. P., & Roy, K. (2008). On Some Aspects of Variable Selection for Partial Least Squares Regression Models. *QSAR & combinatorial science*, *27*(3), 302–313

Sadler, B. R., Cho, S. J., Ishaq, K. S., Chae, K., & Korach, K. S. (1998). Three-dimensional quantitative structure-activity relationship study of nonsteroidal estrogen receptor ligands using the comparative molecular field analysis/cross-validated r2-guided region selection approach. *Journal of medicinal chemistry, 41*(13), 2261-2267. doi: 10.1021/jm9705521

Saeed, M. F., Kolokoltsov, A. A., Albrecht, T., & Davey, R. A. (2010). Cellular entry of ebola virus involves uptake by a macropinocytosis-like mechanism and subsequent trafficking through early and late endosomes. *PLoS pathogens, 6*(9), e1001110. doi: 10.1371/journal.ppat.1001110

Sakkiah, S., Thangapandian, S., John, S., Kwon, Y. J., & Lee, K. W. (2010). 3D QSAR pharmacophore based virtual screening and molecular docking for identification of potential HSP90 inhibitors. *European journal of medicinal chemistry, 45*(6), 2132-2140. doi: 10.1016/j.ejmech.2010.01.016

Schornberg, K., Matsuyama, S., Kabsch, K., Delos, S., Bouton, A., & White, J. (2006). Role of endosomal cathepsins in entry mediated by the Ebola virus glycoprotein. *Journal of Virology, 80*(8), 4174-4178. doi: 10.1128/JVI.80.8.4174-4178.2006

Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W., Jr. (2014). Computational methods in drug discovery. *Pharmacological reviews, 66*(1), 334-395. doi: 10.1124/pr.112.007336

Smellie, A., Teig, S. L., & Towbin, P. (1995). Poling: Promoting conformational variation. *Journal of computational chemistry, 16*(2), 171-187

Smith, L. M., Hensley, L. E., Geisbert, T. W., Johnson, J., Stossel, A., Honko, A., . . . Karp, C. L. (2013). Interferon-beta therapy prolongs survival in rhesus macaque models of Ebola and Marburg hemorrhagic fever. *Journal of infectious disease, 208*(2), 310-318. doi: 10.1093/infdis/jis921

Sondergaard, C. R., Olsson, M. H., Rostkowski, M., & Jensen, J. H. (2011). Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *Journal of chemical theory and computation, 7*(7), 2284-2295. doi: 10.1021/ct200133y

Sullivan, N. J., Sanchez, A., Rollin, P. E., Yang, Z. Y., & Nabel, G. J. (2000). Development of a preventive vaccine for Ebola virus infection in primates. *Nature, 408*(6812), 605-609. doi: 10.1038/35046108

Taha, M. O., Habash, M., Al-Hadidi, Z., Al-Bakri, A., Younis, K., & Sisan, S. (2011). Docking-based comparative intermolecular contacts analysis as new 3-D QSAR concept for validating docking studies and in silico screening: NMT and GP inhibitors as case studies. *Journal of chemical information and modeling, 51*(3), 647-669. doi: 10.1021/ci100368t

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., . . . Mackerell, A. D., Jr. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computional chemistry, 31*(4), 671-690. doi: 10.1002/jcc.21367

Vetter, P., Fischer, W. A., 2nd, Schibler, M., Jacobs, M., Bausch, D. G., & Kaiser, L. (2016). Ebola Virus Shedding and Transmission: Review of Current Evidence. *Journal of infectious disease, 214*(suppl 3), S177-S184. doi: 10.1093/infdis/jiw254

Wang, F., & Chen, Y. (2013). Pharmacophore models generation by catalyst and phase consensus-based virtual screening protocol against PI3Kα inhibitors. *Molecular simulation, 39*(7), 529-544. doi: 10.1080/08927022.2012.751592

Wang, J., Deng, Y., & Roux, B. (2006). Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys journal, 91*(8), 2798-2814. doi: 10.1529/biophysj.106.084301

Wang, J., Wang, W., Kollman, P. A., & Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling, 25*(2), 247-260. doi: 10.1016/j.jmgm.2005.12.005

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of computational chemistry, 25*(9), 1157-1174. doi: 10.1002/jcc.20035

*World Health Orginization (WHO)*. (2016). Genevahttp://apps.who.int/ebola/ebola-situation-reports.

Yermolina, M. V., Wang, J., Caffrey, M., Rong, L. L., & Wardrop, D. J. (2011). Discovery, synthesis, and biological evaluation of a novel group of selective inhibitors of filoviral entry. *Journal of medicinal chemistry, 54*(3), 765-781. doi: 10.1021/jm1008715

Zhao, Y., Ren, J., Harlos, K., Jones, D. M., Zeltina, A., Bowden, T. A., . . . Stuart, D. I. (2016). Toremifene interacts with and destabilizes the Ebola virus glycoprotein. *Nature, 535*(7610), 169-172. doi: 10.1038/nature18615