Med Chem Res (2012) 21:4083–4099
DOI 10.1007/s00044-011-9936-2

MEDICINAL CHEMISTRY RESEARCH

ORIGINAL RESEARCH

Combined chemical feature-based assessment and Bayesian model studies to identify potential inhibitors for Factor Xa

Meganathan Chandrasekaran · Sugunadevi Sakkiah · Keun Woo Lee

Received: 19 April 2011 / Accepted: 8 December 2011 / Published online: 30 December 2011

Abstract In our study, we have described chemical feature-based 3D QSAR pharmacophore models with help of known inhibitors of Factor Xa (FXa). The best model, Hypo1, has validated by various techniques to prove its robustness and statistical significance. The well validated Hypo1 was used as 3D query in the virtual screening to retrieve potential leads for FXa inhibition. The hit molecules were sort out by applying drug-like filters and molecular docking. Bayesian model was developed using training set compounds which provides molecular feature that are favoring or not favoring for FXa inhibition.

Keywords Factor Xa · HypoGen · Anticoagulant · Virtual screening · Molecular docking · Bayesian model

Abbreviation

FXa	Factor Xa
DS	Discovery studio v2.5
HBA	Hydrogen bond acceptor
HBD	Hydrogen bond donor
H	Hydrophobic
Haro	Hydrophobic aromatic
Hali	Hydrophobic aliphatic
R	Ring aromatic
PI	Positive ionizable
RMS	Root mean square
EF	Enrichment factor
GH	Goodness of hit
ADMET	Absorption, distribution, metabolism, excretion, and toxicity
BBB	Blood–brain barrier
ECFP	Extended-connectivity fingerprints
ROC	Receiver operating curve

Electronic supplementary material The online version of this article (doi:10.1007/s00044-011-9936-2) contains supplementary material, which is available to authorized users.

M. Chandrasekaran · S. Sakkiah · K. W. Lee (✉)
Division of Applied Life Science (BK21 Program), Systems and Synthetic Agrobiotech Center (SSAC), Plant Molecular Biology and Biotechnology Research Center (PMBBRC), Research Institute of Natural Science (RINS), Gyeongsang National University (GNU), 501 Jinju-daero, Gazha-dong, Jinju 660-701, Republic of Korea
e-mail: kwlee@gnu.ac.kr

Introduction

Blood coagulation cascade is a series of stepwise enzymatic conversions which activated by both intrinsic and extrinsic pathways. Both the pathways are converged at the point of Factor Xa (FXa) activation (Borensztajn and Spek, 2011) leading to conversion of prothrombin to thrombin and the subsequent conversion of fibrinogen to fibrin (Roehrig *et al.*, 2005). Thrombin is a multifunctional serine protease, which has several thrombotic functions, including the conversion of fibrinogen to fibrin, platelets activation, feedback activation of other coagulation factors, resulting in the amplification of its own formation. FXa is a trypsin-like serine protease located at the confluence of intrinsic and extrinsic pathways of coagulation cascade. Inhibition of both the enzymes such as thrombin and FXa provide antithrombotic effect. However, inhibition of FXa has been decreased the amplified generation of thrombin by diminishing thrombin-mediated activation of both coagulation and platelets, without affecting existing thrombin level. Moreover, FXa activity is enhanced while combine with cofactor, Factor Va, and calcium on the platelet

phospholipid membrane surface to form the prothrombinase complex (Adler *et al.*, 2000). Also, this prothrombinase complex, not thrombin, is responsible for the major procoagulant activity on human whole-blood clots. This has supported that FXa could be a more important mediator of thrombus progression than thrombin (Adler *et al.*, 2000). Further, the recent studies revealed that FXa has an essential player of cell biology through activation of protease-activated receptors (PAR)-1 and -2 (Borensztajn and Spek, 2011). Due to its pleiotropic role of coagulation enzyme, FXa has emerged as an attractive target for developing safer anticoagulant drugs as well as in the treatment of proliferative diseases.

There are two types of anticoagulant such as vitamin K antagonist (e.g., warfarin), and heparin have been prescribed for the past five decades. However, these two antagonists share the several limitations. Warfarin has narrow therapeutic window, unpredictable pharmacokinetics profile and its therapeutic level is difficult to control also it activity is affected by food and drug interactions (de Candia *et al.*, 2009). Further, it has required continuous monitoring since the half-life time period of warfarin was approximately 40 h thus the anticoagulant effect retained for several day after completion of the treatment which makes the difficulty in crucial situations (Borensztajn and Spek, 2011). Heparin is acted by the body's own "mopping up" agents for the enzymes thrombin and FXa. Heparin also need of laboratory monitoring for dose adjustment as well as it induced thrombocytopenia. Since several limitations of both drugs need new anticoagulants which bypass these limitation. The crucial role of FXa in the coagulation cascade, direct FXa inhibitors would offer more efficiency for the treatment of thromboembolic diseases (Ansell, 2007). Moreover, some previous works (Alban, 2005; Walenga *et al.*, 2003) have been suggested that FXa inhibitors might be superior to direct thrombin inhibitors. While compared with direct thrombin inhibitors, the direct FXa inhibitors have a less bleeding complication in animal model when given in doses with similar antithrombotic activity (Carreiro and Ansell, 2008). Some of advantages have been explained potential superiority of direct FXa inhibitors: (1) the small amount of anticoagulant drug is enough to inhibit coagulation due to its amplified nature of coagulation cascade, (2) a wide concentration range is sufficient to progressively inhibit the FXa than thrombin, suggesting that FXa inhibitors may have a wider therapeutic window than thrombin inhibitors, (3) one of the main disadvantage of direct thrombin inhibitors is associated with rebound hypercoagulable state but this does not appear to be associated with direct FXa inhibitors. These theoretical considerations suggested that direct FXa inhibitors are more efficient than thrombin inhibitors. This will motivate us to develop the chemical feature based pharmacophore models using FXa inhibitors which may be useful to identify novel potential virtual leads to inhibit the FXa activity.

In our study, we have developed chemical feature-based 3D QSAR pharmacophore model using known inhibitors of FXa. For the purpose to identify the critical chemical feature necessary to FXa activity, we performed extensive survey about FXa inhibitor. Taha *et al.* (2005) have been carried out ligand-based assessment of FXa binding site flexibility, in which they suggested that the features hydrogen bond acceptor (HBA), Hydrophobic aromatic (Haro), ring aromatic (R), and positive ionizable (PI) groups are important to generate the pharmacophore models for FXa inhibitors. This we keep in mind to use these features as one of the input in our pharamacophore generation phase. In addition, Bayesian model was developed to assess the validity of our training set compounds as well as this model can suggest the molecular feature which are favoring to inhibit the FXa activity and not favoring for inhibit FXa activity.

Materials and methods

The entire calculations were carried out by using the Discovery Studio v2.5 (DS) software package (Accelrys, San Diego, USA). The scheme of our study has been given in Fig. 1.

Pharmacophore generation

Data set preparation and molecular modeling

A total of 40 FXa inhibitors were collected from literatures (Roehrig *et al.,* 2005, Yoshikawa *et al.*, 2009, Shi *et al.,* 2009, 2008). The experimental inhibitory activity (IC_{50}) values of these compounds were tested using unique assay procedure (Enzyme inhibition assay). These 40 compounds were divided into two categories: (1) training set to generate the pharmacophore models, (2) test set to validate the best model. The selection of correct choice training set compounds is a fundamental consequence in an automated pharmacophore generation phase. To assure the statistic relevance of the generated model some guidelines for training set selection must be followed, for example, the minimum of 16 molecules to ensure statistical significance of pharmacophore (Ravikumar *et al.*, 2008). The selected molecules should have the activity range at least four orders of magnitude; the most active and inactive molecules must be included and each order of magnitude should be represented by at least three molecules. Each compounds in the trainings set should provide new structural information; the activity data, in this case of IC_{50} value

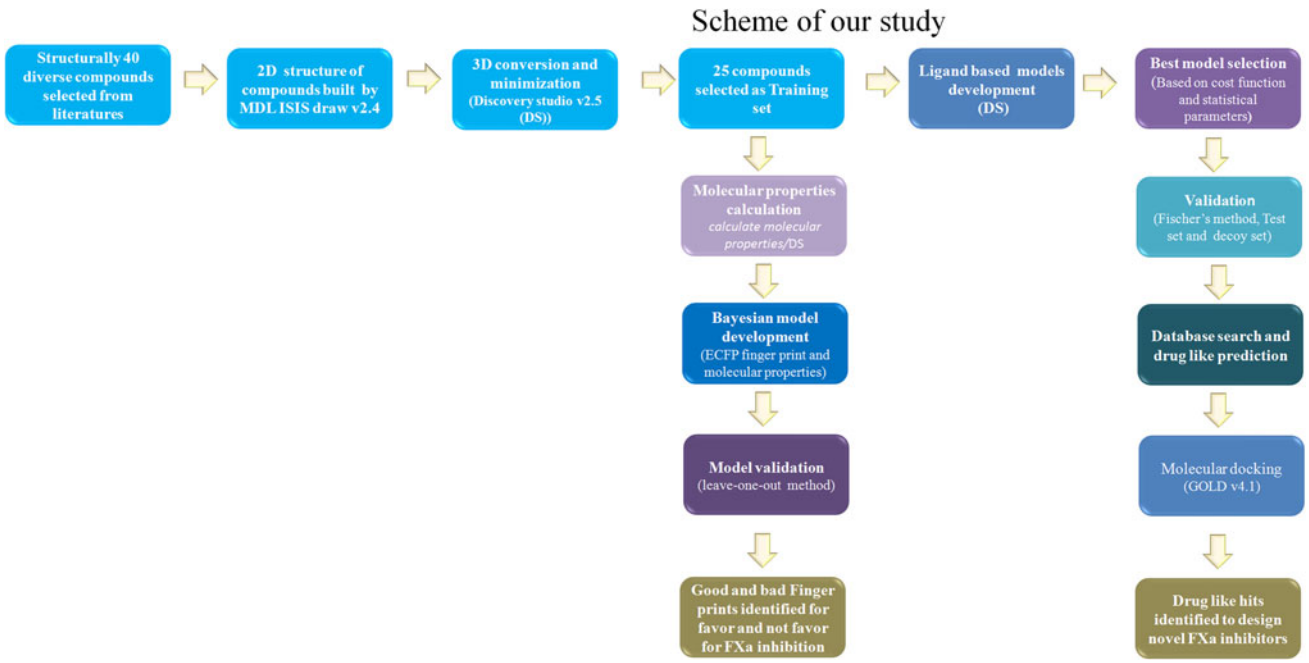

Fig. 1 Basic work flow of our study

(Yang *et al.*, 2009). Based on the above valuable information, 25 structurally diverse compounds were selected as training set (Fig. 2) and the remaining 15 compounds considered as test set (Fig. 3), respectively. Training set molecules have both structural diversity and wide coverage of the activity (IC_{50}) value (0.1–1,200 nM) that spans over four orders of magnitude toward FXa, to achieve a reasonable pharmacophore model in terms of predictive ability and statistical significance.

The 2D form of training set compounds were built using MDL-ISIS draw v2.4 and these 2D structures were converted to 3D form by importing in DS. All compounds were optimized to the closest local minimum using CHARMm-like force field (Mackerell *et al.*, 1998). The conformational analysis for each compound was generated by poling algorithm (Smellie *et al.*, 1995a, b, c) and CHARMM force field parameters. These conformations are used not only in hypothesis generation but also for fitting the compound to a hypothesis and for estimating the activity of the compound. There are two methods available in DS to generate the conformation: Fast and Best quality analysis. Both methods were carried out using a Monte Carlo-like algorithm together with poling (Smellie *et al.*, 1995a, b, c), which reduces considerably the probability of reappearance of nearly similar conformers by usage of penalty function. In this study, a maximum of 255 conformations was set for each compound by using "*Best conformation generation*" method with a constraint of 20 kcal/mol energy cut off above the global energy minimum to ensure maximum coverage of the conformational space. The conformational variations of similar conformer have explicitly promoted away from each other by poling and all are equally treated. Same procedures and methods followed to prepare the test set compounds.

Hypothesis generation and evaluation

The *HypoGen* algorithm allows developing pharmacophore models (hypothesis) which determined the 3D arrangement of a collection of features important for the biological activity of the ligands. Before generation of the model, we first analyzed important feature necessary to inhibit the FXa activity. Hence, we carried out the feature selection for the pharmacophore generation using *Feature mapping* algorithm in which all the features such as HBA, hydrogen bond donor (HBD), HBA lipid, H, Haro, hydrophobic aliphatic (Hali), ring aromatic (R), PI, negative ionizable (NI) were used to map the training set compounds. This method has revealed that HBA, HBD, Hali, R features could effectively map all training set molecules. Further, similar kind of work has been carried out by Taha *et al.* (2005) suggested that HBA, Haro, R, and PI features important for FXa active site. Together these results, we have utilized HBA, HBD, R, Haro, Hali, and PI features to generate the pharmacophore models. The uncertainty value of 2.1 for each compound was defined instead of a default value of 3, to represent the ratio range of uncertainty in the activity value based on the expected statistical straggling of

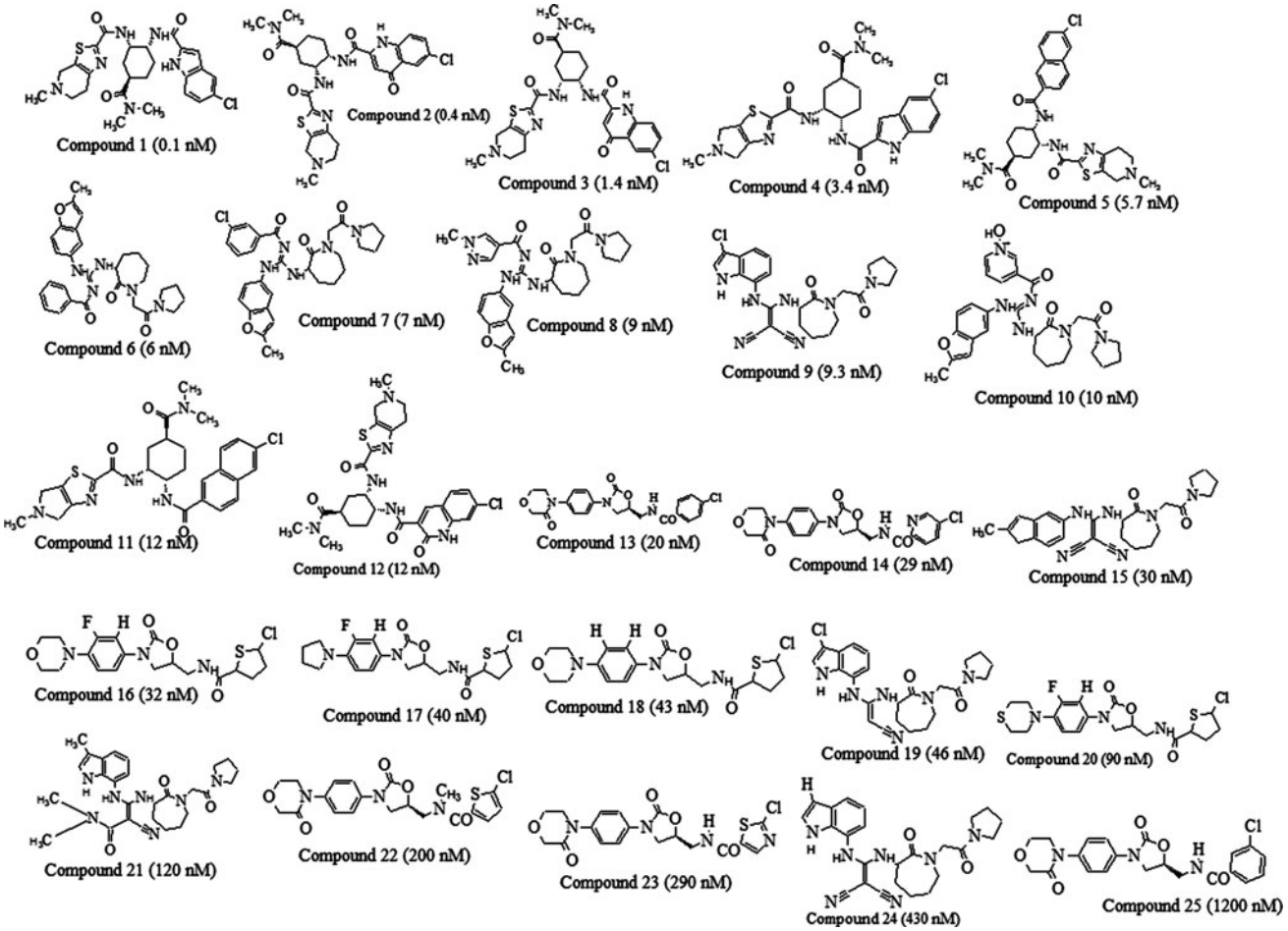

Fig. 2 2D representation of training set molecules along with IC_{50} values

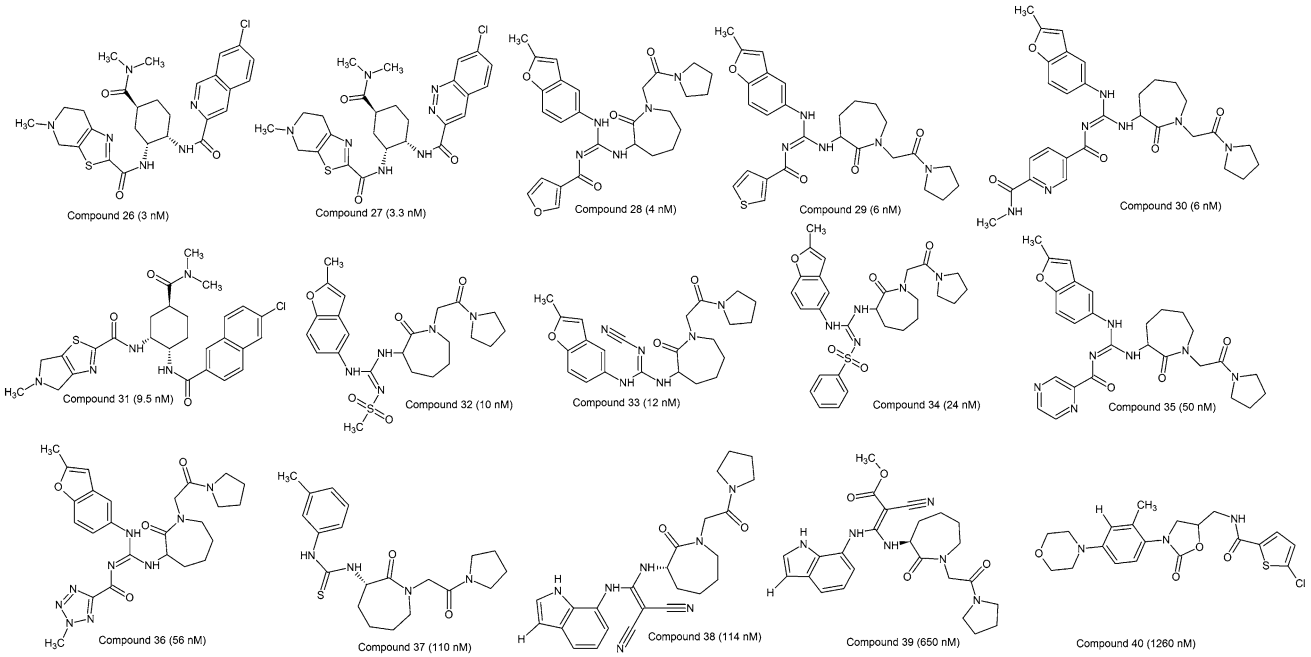

Fig. 3 2D representation of test set molecules along with IC_{50} values

biological data collection (Yang *et al.*, 2009). The minimum and maximum count for all the features in the hypothesis run was set of 0 and 5, respectively. Default values were set for all other parameters such as maximum points and minimum subset points.

There are three steps followed in hypothesis generation phase, First, all the active compounds from the trainings set were identified in the Constructive phase. Second, the inactive compounds were identified in the subtractive phase and finally optimization phase: DS attempts to minimize a cost function consisting of two terms. One penalizes the deviation between the estimated activities of the training set molecules and their experimental values and other penalizes the complexity of the hypothesis. The generation process stops when optimization no longer improves the score. Finally, HypoGen was generated top ten hypotheses using the training set.

There are three cost calculation performed in the HypoGen module such as fixed cost, null cost, and total cost to determine the success of the pharmacophore. The fixed cost represents the simplest model that fits the data perfectly. The null cost represents the highest cost of a hypothesis with no features that estimates every activity to be the average activity. For the simplicity, the large difference between fixed cost and null cost value above 60 bits could entail a 90% probability for correlating the experimental and predicted activity data. The total cost for each hypothesis is the summation of the three cost components [error (E), weight (W), and configuration cost (C)] multiplied by a coefficient (default coefficient is 1.0 for each). The error cost is solely dependent on the root mean square (RMS) deviations. The RMS value represents the quality of the correlation between the experimental and the estimated activity data. The weight cost increases, the weight factor for the chemical features deviate from the default value of 2. The configuration cost is represented as Log_2P, where P is the number of initial hypotheses created in the constructive phase and that survived in the subtractive phase. The total cost of any hypothesis should be close to the fixed cost for a good model. Among top 10 hypotheses, one hypothesis was selected as best hypothesis based on the cost functions and statistical parameters.

The predictability of the best model was initially analyzed by using training set compounds. Thus, the training set activity could be estimated using regression parameters which are computed by the regression analysis using the relationship of predicted and experimental activities. The error value is calculated by the difference between predicted and experimental activity. A positive error indicates that the predicted IC_{50} value is higher than the experimental IC_{50} value, while a negative error value indicates that the predicted IC_{50} value is lower than the experimental IC_{50} value.

Validation of pharmacophore model

The generated pharmacophore model should have the ability to predict the activity of molecules accurately and also to discover the potent leads from databases. Hence, the reliability and predictability of derived pharmacophore model was validated by using (1) Fischer's method, (2) Test set, and (3) Decoy set.

Fischer's method

In the Fischer's method (Fischer, 1966), the experimental activities (IC_{50} value) of the training set was arbitrarily scrambled with use of CatScramble program. The randomized training set was used to generate the hypothesis using the same protocols in original hypothesis generation phase. These randomized training set should yield hypotheses without statistical significance; otherwise, the original model is considered to generate by chance. The following formula is used to calculate the statistical significance:

$$\text{Significance} = 100\left(1 - 1 + \frac{x}{y}\right)$$

where "x" is the total number of hypothesis having a total cost lower than the best-significant hypothesis and "y" is the number of initial HypoGen runs plus random runs.

Test and decoy sets

The predictability of generated model was not only validated by training set compound but also two external methods such as test and decoy sets. In the first method, the model was validated using test set of 15 known inhibitors of FXa. These 15 compounds were not included in the training set. Decoy set, second method, which contains known and unknown inhibitors of FXa. The best model used to screen the decoy set and check how well it retrieves the active candidate from inactive molecules. The parameters such as false positives, false negatives, enrichment factor (EF) and goodness of hit (GH) were calculated to determine the robustness and statistical significance of hypothesis.

Virtual screening

The main aim of virtual screening is to retrieve the potent molecules based on the chemical feature and shape (Bajorath, 2002). Since the biological activity of molecules mainly depends on chemical features for binding in the active site and shape to fit into the active site. Virtual screening is an inexpensive and fast alternative tool to identify potent inhibitors of target protein (Hahn, 1997,

Putta *et al.*, 2002). In our study, the best predictive model was used as query to search potential leads against FXa from Chembridge and Maybridge databases. *Ligand Pharmacophore Mapping/DS* protocol was used to screen the database with maximum Omitted Features value of "0." The *Fast Flexible* database search option was employed to perform the virtual screening. The hit molecules from database were subjected to filter by applying Lipinski's rule-of-five (Lipinski *et al.*, 1997) and absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties.

Molecular docking

Molecular docking was used to refine hit the molecules through find the affinity between ligand and critical amino acid present in the active site of the protein. Several algorithms have been described for molecular docking that could be able to find the ligand protein affinity through scoring function and also interaction between ligand and critical amino acid in the active site. In this study, we used GOLD v4.1 (Genetic Optimization for Ligand Docking) from Cambridge Crystallographic Data Center, UK (Jones *et al.*, 1997) uses genetic algorithm for docking flexible ligands into protein binding sites to explore the full range of ligand conformation flexibility with partial flexibility of the protein. The X-ray crystal structure of FXa (PDB ID: 2XBV) for molecular docking study was directly obtained from the Protein Data Bank (PDB) (www.rcsb.org). Preparation of protein for docking by removing all the water molecules and hydrogen atoms are added to calculate the bond orders for the protein and ligand using GOLD. The binding site of protein was defined for all the atoms within 10 Å on the basis of co-crystallized ligand in the X-ray structure. During the docking study, ten poses were generated for each ligand and the best poses were selected based on the GOLD fitness score. Default fitness function (VDW 4.0, H-bonding 2.5) and evolutionary parameters were performed in the GOLD docking experiments: population size 100; selection pressure 1.1; operation 100,000; islands 5; niche size 2; migration 10; crossover 95. Gold fitness score, binding mode, and molecular interactions with catalytically important residues in protein active site were used to sort the molecules.

Bayesian model

Bayesian model is mainly depends on Bayes's rule of conditional probability and it has been applied to many real-world problems such as pattern recognition, machine learning, medicine, public health, bioinformatics, and cheminformatics (Klon *et al.*, 2006). Herein, we applied Bayesian model to determine the efficiency of training set molecules by leave-one-out cross-validation method and identify the important finger prints necessary to inhibit the FXa activity. Hence, we have developed the Bayesian model for same training set molecules used in pharmacophore generation phase. Some of the reported result suggested that the 2D information to extract useful information and compete favorably against traditional 3D approaches (Rogers *et al.,* 2005). The following 2D descriptors: AlogP, molecular weight, number of rotatable bonds, number of rings, number of aromatic rings, number of HBAs and number of HBDs, and molecular fractional polar surface area were used as independent variables during model generation. In addition, we have used extended-connectivity fingerprints at maximum diameter 6 (ECFP_6) which means that it contains all circular substructures, around each atom, up to a maximum width of six bonds. Since ECFPs are a class of 2D fingerprint for molecular characterization. They are based on a process derived using a variant of the Morgan algorithm (Rogers *et al.,* 2005). The protocols such as "*calculate molecular properties*" used for descriptor calculation and "*Create Bayesian Model*" for Bayesian model generation.

Result and discussion

Hypothesis generation and validation

All ten pharmacophore models have the combination of two features such as, HBA, Hali. This can be suggested that these two features are very important to inhibit the FXa activity. Further, based on chemical features and validation through external molecules, this obtained pharmacophore models have shown much difference while compared with already reported result (Taha *et al.*, 2005). The biological activities are mainly depends on the chemical features of molecules but chemical features alone is not a sufficient condition to inhibit a target protein it should have the correct shape to fit into the active site of protein (Adane *et al.*, 2010). With this fact on our mind, we have carefully analyzed all ten hypotheses by cost function, correction coefficient, and RMS deviation.

In our study, the null cost and fixed cost values of all ten hypotheses were 169.9 and 89.6, respectively, and the difference between them is 80.3 bits. The difference of more than 60 bits between null cost and fixed cost represents the 90% chance of predictive hypotheses. The correlation coefficient range of all ten pharmacophore was 0.97–0.87 which is calculated between experimental and predicted values of all molecules present in the training set. The cost difference between null cost and total cost was observed more than 60 bits which indicates that the all the ten hypotheses having 75–90% (Zhang *et al.*, 2009; Kansal

Table 1 Statistical parameters of top ten pharmacophore hypotheses

Hypotheses	Total cost	ΔCost (null cost − total cost)	RMS	Correlation	Features
1	94.8	75.1	0.63	0.97	HBA HBA Hali Hali
2	103.2	66.7	1.00	0.93	HBA HBA Hali Hali Hali
3	105.5	64.4	1.12	0.91	HBA HBA Hali Hali
4	107.9	62.0	1.20	0.89	HBA HBA Hali Hali
5	110.7	59.2	1.29	0.88	HBA HBA Hali Hali
6	111.0	58.9	1.23	0.89	HBA HBA Hali Hali Hali
7	111.7	58.2	1.27	0.88	HBA HBA Hali Hali Hali
8	112.6	57.3	1.35	0.87	HBA HBA Hali Hali
9	113.5	56.4	1.34	0.87	HBA HBA Hali Hali
10	114.1	55.8	1.36	0.87	HBA Hali Hali Hali

RMS root mean square, *HBA* hydrogen bond acceptor, *Hali* hydrophobic aliphatic

Null cost = 169.9; Fixed cost = 89.6; Configuration cost = 14.2. All cost units are in bits

et al., 2009) of true correlation. The cost functions and statistical parameters of entire pharmacophore details are listed in Table 1. After thorough investigation about all ten hypotheses, the first hypothesis, Hypo1, was chosen as the best hypothesis. Since, it was characterized by highest cost difference (75.1), the highest correlation coefficient (0.97), and lowest RMSD (0.63) as well as good configuration cost of 14.2 (Table 1). The spatial arrangement of Hypo1 consists of four chemical features: two HBA and two Hali. The 3D space and geometric constraints of Hypo1 has shown in Fig. 4.

The predictability of Hypo1 was initially validated using training set compounds (Internal validation). For this purpose, all the training set molecules were grouped into three categories such as highly active (+++, $IC_{50} < 20$ nM) moderately active (++, $20 \leq IC_{50} < 100$ nM) and inactive (+, $IC_{50} \geq 100$ nM). The validation result such as the experimental together with estimated inhibitory values for training set compound are summarized in Table 2. Out of 25 compounds, 23 were predicted their own activity range also the estimated IC_{50} values was very close to the corresponding experimental IC_{50} value. The estimated activities of other two compounds: one highly active (+++) compounds underestimated as moderately active, one inactive (+) overestimated as moderately active (Table 2). This validation method has clearly determined the good predictability of Hypo1. Furthermore, Hypo1 was again analyzed by mapping the chemical feature on the training set compound. Figure 5a, b represents the Hypo1 aligned with most active compound **1** (IC_{50}: 0.1 nM) and least active compound **25** (IC_{50}: 1,200 nM) in the training set. In that, all the features were perfectly mapped well with

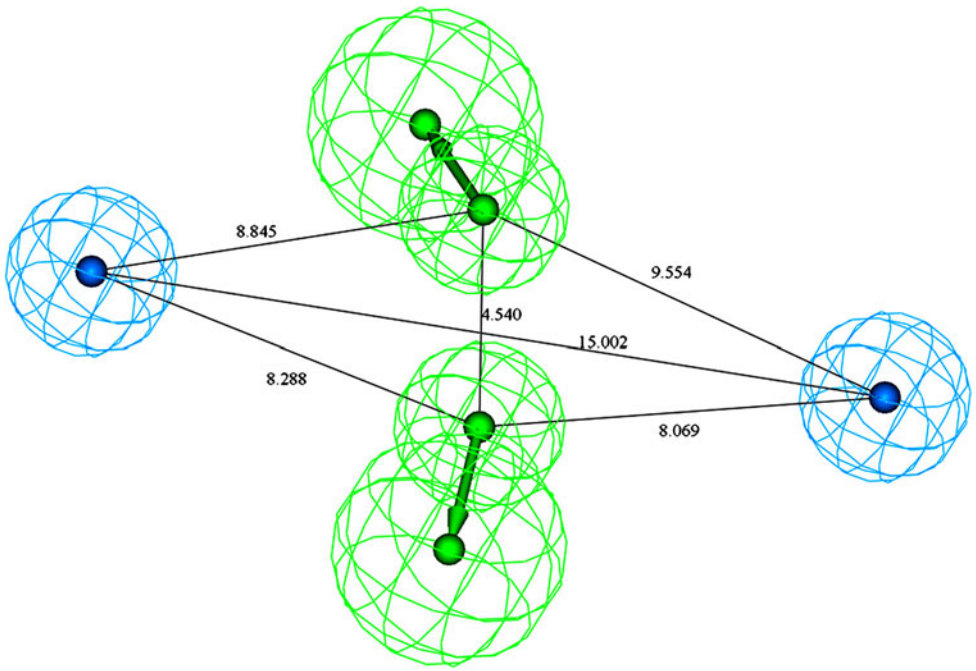

Fig. 4 3D spatial relationship and geometrical parameters of the best quantitative model Hypo1 color coded for features: *green* HBA, *blue* Hali (Color figure online)

Table 2 Experimental and predicted activities of training set compounds measured on the basis of Hypo1

Compound nos.	Fit value	Experimental IC_{50} (nM)	Predicted IC_{50} (nM)	Error[a]	Experimental scale[b]	Predicted scale[b]
1	8.41	0.1	0.05	−1.7	+++	+++
2	7.43	0.4	2.5	+1.3	+++	+++
3	6.78	1.4	4.5	+1.8	+++	+++
4	6.52	3.4	4.3	+1.3	+++	+++
5	6.53	5.7	9.3	−1.3	+++	+++
6	6.20	6.0	9.1	+1.6	+++	+++
7	6.21	7.0	8.6	+1.3	+++	+++
8	6.24	9.0	6.4	−1.1	+++	+++
9	6.37	9.3	15	−1.5	+++	+++
10	5.98	10	15	+1.5	+++	+++
11	5.65	12	33	+2.7	+++	++
12	6.33	12	6.9	−1.7	+++	+++
13	5.54	20	42	+2.1	++	++
14	5.50	29	46	+1.6	++	++
15	5.86	30	21	−1.5	++	++
16	5.66	32	32	+1	++	++
17	5.40	40	59	+1.5	++	++
18	5.65	43	34	−1.3	++	++
19	5.49	46	48	+1	++	++
20	5.53	90	44	−2	++	++
21	5.37	120	64	−1.9	+	++
22	4.95	200	170	−1.2	+	+
23	4.97	290	160	−1.8	+	+
24	4.48	430	500	+1.2	+	+
25	4.35	1200	660	−1.8	+	+

Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule. Fit = weight × [max (0,1,SSE)] where SSE = $(D/T)^2$, D is displacement of the feature from the center of the location constraints and T is the radius of the location constraint sphere for the feature (tolerance)

[a] Difference between the predicted and experimental values. "+ve" indicates that the predicted IC_{50} is higher than the experimental IC_{50}, "−ve" indicates that the predicted IC_{50} is lower than the experimental IC_{50}, a value of 1 indicated that the predicted IC_{50} is equal to the experimental IC_{50}

[b] Activity scale: $IC_{50} < 20$ nM = +++ (highly active); $20 \leq IC_{50} < 100$ nM = ++ (moderately active); $IC_{50} \geq 100$ nM = + (inactive)

the compound **1**, whereas compound **25** fails to fit one of Hali feature. Both validations suggested that the validity and fit value of Hypo1 was reliable.

Pharmacophore validation

Fischer's randomization method

Fischer's method was applied to validate the statistical relevance of Hypo1; there is any strong correlation between the chemical structures and biological activities as well as cost function. In cross-validation method, the generation of spreadsheets depends on confidence level which avail in Fischer's model 90, 95, 98% and then 19, 49, and 99 spread sheets have to be generated, respectively (Zhang *et al.*, 2009). According to the confidence level, activities of the training set molecules scrambled in the random manner and produced different order of training set. The modified training set was used to generate the spread sheets using the same features and parameters used in the original pharmacophore generation. In our study, we set 95% confidence level hence 19 spread sheets were generated. The statistical parameters of the 19 spread sheets and original hypothesis (Hypo1) have been described in Table 3. When compared with Hypo1, none of the generated hypotheses have gained the good correlation coefficient, RMS, and cost difference. As well as, the total cost of generated hypotheses are considerably above when compared to Hypo1 (Fig. 6). Based on the Fischer's randomization result, Hypo1 is not generated by chance and its values far more superior than those of the 19 randomly produces hypotheses.

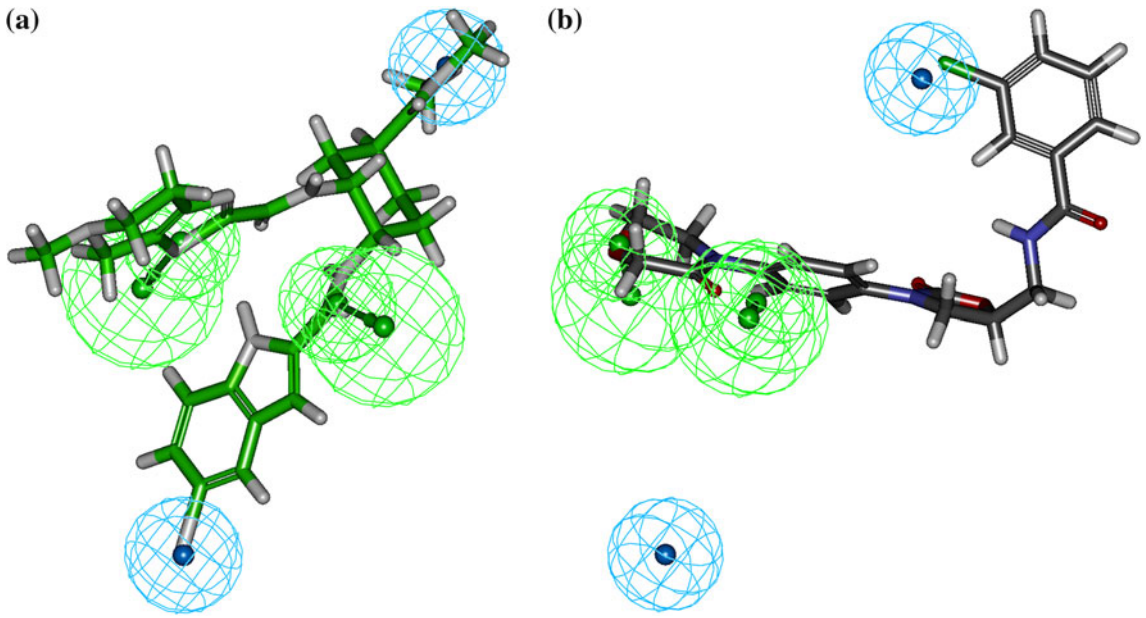

Fig. 5 Hypo1 aligned with the **a** most active Compound **1** (IC_{50}: 0.1 nM), **b** least active Compound **25** (IC_{50}: 1,200 nM). Color coded for features: *green* HBA, *blue* Hali (Color figure online)

Test and decoy sets

In order to assert the predictability of Hypo1 in the external compound, we prepared a test set of 15 FXa inhibitors with different structural information and various activity classes. The test set compounds are classified like as training set: highly active (+++, $IC_{50} < 20$ nM), moderately active (++, $20 \leq IC_{50} < 100$ nM), and inactive (+, $IC_{50} \geq 100$ nM). Hypo1 was used to screen the test set using *Ligand pharmacophore mapping* protocol with Maximum Omitted Feature value of "0". The experimental versus predicted activities and error values of test set have shown in Table 4. Hypo1 was regressed against the test set molecules and gave a correlation coefficient value of 0.90 this was quite equal to the training set correlation value of 0.97 (Fig. 7). Hypo1 was predicted six out of eight highly active molecules correctly as their own activity range and the rest two molecules underestimated as moderately active. There was no discrepancy found in the moderately active and inactive molecules. Overall prediction, 13 out of 15 (86%) test set molecules were predicted in their own activity range with the error value less than 5.

The reliability of Hypo1 was again validated by decoy set which contains known and unknown ligands of FXa activity. Decoy molecules have possessed with similar composition to the known ligands, but with a different topology that are assumed not to bind to the target protein. A total of 1,500 molecules comprised of 1,475 unknown molecules and 25 known inhibitors of FXa were used in decoy set. This method used to calculate necessary 1D property such as false positive, false negative, EF, and GH, to determine the validity and robustness of Hypo1. The GH method also called as Güner–Henry (GH) scoring method (Ravikumar *et al.*, 2008) which is used to assess the quality of our model by quantifying the precision of hits and the recall of actives from inactives. The GH score ranges from 0, which indicates the null model, to 1, which indicates the ideal model (Ravikumar *et al.*, 2008). The below mentioned formulae are used to calculate EF and GH values, respectively

$$\mathrm{EF} = \left(\frac{H_a}{H_t}\right) \div \left(\frac{A}{D}\right)$$

$$\mathrm{GH} = \left(\frac{H_\mathrm{a}(3A + H_\mathrm{t})}{4H_\mathrm{t}A}\right)\left(1 - \frac{H_\mathrm{t} - H_\mathrm{a}}{D - A}\right)$$

where H_t is the number of hits retrieved, H_a is the number of active molecules in the hit list, A is the number of active molecules present in the database, and D is the total number of molecules in the database. Hypo1 was performed database mining using decoy set by *Best Flexible* searching technique and the results were summarized in Table 5. Hypo1 was successfully predicted 22 compounds as active compounds from database. Of these 19 compounds (86%) were known inhibitors of FXa activity, the rest 3 inactive compounds predicted as active (false positive), and predicted 6 active compounds as inactive (false negative). EFs of 5.18 indicate that Hypo1 has five times more probable to pick an active compound from the database than the inactive one (Gopalakrishnan *et al.*, 2005). Hypo1 has yielded the GH score value of 0.83 which represents the quality of our model was acceptable since

Table 3 Results from cross-validation using CatScramble implemented in DS

Validation nos.	Total cost	RMS	Correlation	Cost difference
Hypo1	94.8	0.63	0.97	75.1
Results for scrambled				
Random1	132.4	1.78	0.76	37.5
Random2	162.2	2.44	0.46	7.7
Random3	149.4	2.22	0.59	20.5
Random4	118.6	1.51	0.83	51.3
Random5	157.8	2.38	0.50	12.1
Random6	132.7	1.82	0.75	37.2
Random7	148.4	2.18	0.61	21.5
Random8	132.5	1.89	0.72	37.4
Random9	145.6	2.17	0.61	24.3
Random10	120.4	1.58	0.81	49.5
Random11	136.1	1.93	0.71	15.7
Random12	134.4	1.88	0.73	14
Random13	134.0	1.87	0.73	35.9
Random14	114.3	1.38	0.86	55.6
Random15	134.1	1.90	0.72	35.8
Random16	130.6	1.84	0.74	39.3
Random17	130.8	1.78	0.76	39.1
Random18	152.3	2.25	0.57	17.6
Random19	146.3	2.13	0.63	23.6

RMS root mean square

Null cost = 169.9; Fixed cost = 89.6

GH score value above 0.7 can be very good model (Wolber and Langer, 2004). According to the both the validation method Hypo1 clearly proved its ability to distinguish the active from inactive molecules. Thus, Hypo1 has taken into the further process such as database screening to retrieve the novel potential candidate for FXa.

Virtual screening

Identification of small molecule modulator for a target protein is grand challenge in the drug discovery process. However, nowadays, the technological advances have made some possible methods such as virtual screening and high-throughput screening. These methods provide an opportunity to collect the candidate molecules with desired target in a more automatic and systematic manner. In our study, virtual screening approach was used to find the potent inhibitors for FXa from the database. Thus, we used publically avail databases, Maybridge and Chembridge which contains 50,000 and 60,000 synthetic compounds along with their conformation to discover the novel inhibitors of FXa, respectively. The well validated Hypo1 was used as 3D query to retrieve the potential new hit molecules from the databases. A total of 10,155 and 6,711 compounds were retrieved from both databases (Maybridge and Chembridge) in the initial screening. The initial hits from databases were further filtered by applying maximum fit value, ADMET descriptors, and Lipinski's rule-of-five to make them more drug-likeness. A maximum fit value of 8.41 for most active compound in the training set was applied to filter the hits. The molecules scored fit value ≥ 8 were retained. A total of 993 molecules (934 & 59) from both databases were gained fit value more than 8. In addition, drug-like predictions such as ADMET and Lipinski's rule-of-five were carried out for the hits with maximum fit value molecules. In case of ADMET, we mainly focused on Solubility, human intestinal absorption (HIA) and blood–brain barrier (BBB) since these three parameters are very important to take forward the drug candidate for next process. The levels of 3, 0, 3, for solubility, HIA and BBB, were applied to sort out the molecules and 16 and 15 molecules obtained from both databases. Furthermore, Lipinski's rule-of-five (Lipinski *et al.*, 1997) was applied to screen the molecules to make them more drug-likeness. The molecular weight > 500, HBD (OH + NH groups) > 5 and HBAs (O's +N's atoms) > 10 were removed in the hit compounds by applying rule of five. Finally, a total of **16** and **15** compounds were passed out and these were taken into the further analysis such as molecular docking. Virtual screening using well validated pharmacophore and steps involved in the screening along with drug-like hits are represented in Fig. 8.

Molecular docking

Molecular docking is another putative method to reduce the false positive of hit compound by checking the interaction between the protein and ligand. Hence, all the database hit compounds (retrieved from virtual screening) were further refined by using GOLD v4.1. The crystal structure of FXa (PDB ID: 2XBV) was downloaded from the PDB for the docking study. After thorough investigation on FXa crystal structure, it consists of two pockets in the active site S_1 and S_4 (Taha *et al.*, 2005) and also a molecule which possess the chlorinated phenyl or heteroaromatic rings could be favorable neutral motif for binding to the S_1 pocket. S_4 pocket called as hydrophobic box which formed by the aromatic side chains of Y99, F174, and W215. Inhibitors mainly get a significant amount of their affinity by binding into this pocket (Taha *et al.*, 2005). The reproducing ability of GOLD was tested against more than 300 protein–ligand complex which extracted from PDB and it succeeded 71% cases (Jones *et al.*, 1997). In addition, comparative study has been carried out by Srivastava *et al.* (2011) on docking protocols such as GOLD, GLIDE, CDOCKER, and AUTODOCK in 57 DNA minor grooves binders obtained

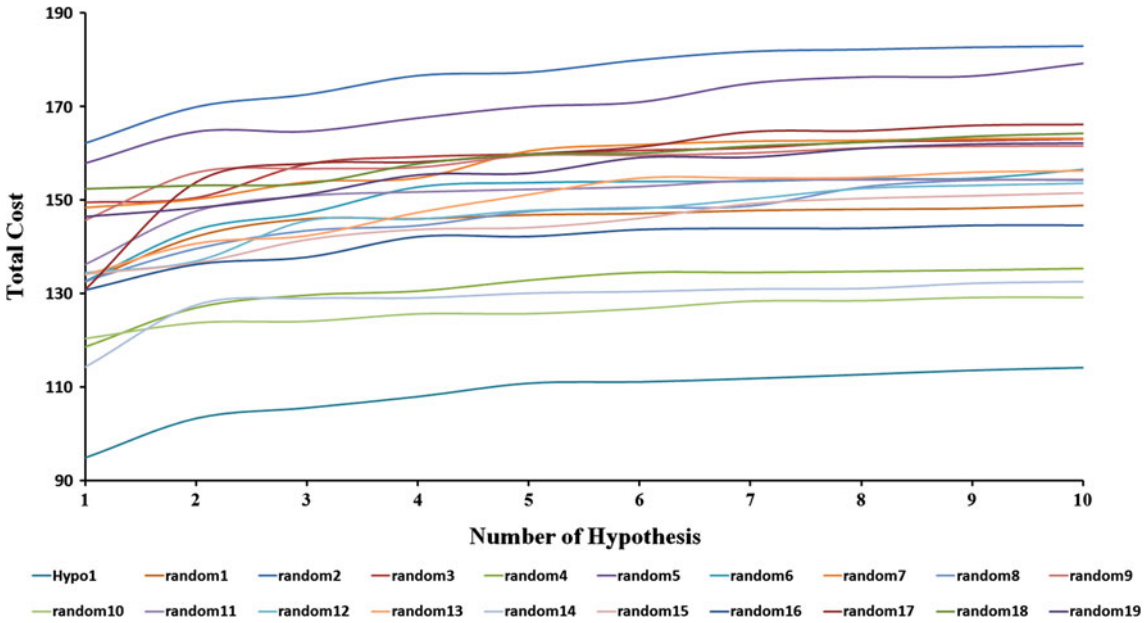

Fig. 6 The difference in correlations of hypotheses between the Hypo1 and 19 random spreadsheets after randomization run

Table 4 Experimental and predicted activities of test set compounds calculated by Hypo1

Compound nos.	Fit value	Experimental IC_{50} (μM)	Predicted IC_{50} (μM)	Error[a]	Experimental scale[b]	Predicted scale[b]
26	7.2	3.0	0.8	−3.3	+++	+++
27	7.1	3.3	1.1	−2.8	+++	+++
28	7.0	4.0	1.3	−3.0	+++	+++
29	6.6	6.0	3.3	−1.8	+++	+++
30	5.7	6.0	23.7	+3.9	+++	++
31	6.7	9.5	2.6	−3.5	+++	+++
32	5.6	10	31.7	+3.1	+++	++
33	6.0	12	12.1	+1.0	++	++
34	5.3	24	73.1	+3.0	++	++
35	6.1	50	11.6	−4.3	++	++
36	5.2	56	92.3	+1.6	++	++
37	4.8	110	232.6	+2.1	+	+
38	5.1	114	111.5	−1.0	+	+
39	4.1	650	1145.6	+1.7	+	+
40	4.4	1260	529.7	−2.3	+	+

Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule. Fit = weight × [max (0,1,SSE)] where SSE = $(D/T)^2$, D is displacement of the feature from the center of the location constraints and T is the radius of the location constraint sphere for the feature (tolerance)

[a] Difference between the predicted and experimental values. "+ve" indicates that the predicted IC_{50} is higher than the experimental IC_{50}, "−ve" indicates that the predicted IC_{50} is lower than the experimental IC_{50}, a value of 1 indicated that the predicted IC_{50} is equal to the experimental IC_{50}

[b] Activity scale: $IC_{50} < 20$ nM = +++ (highly active); $20 \leq IC_{50} < 100$ nM = ++ (moderately active); $IC_{50} \geq 100$ nM = + (inactive)

from PDB in which they suggested that GOLD and GLIDE docking protocols seems very reliable and modeling nucleic acid ligand complex. However, some of cases GOLD failed to reproduce the experimental bound conformation. Hence, we first check reproducibility of GOLD for FXa by docking a co-crystal molecule. The crystal structure 2XBV bound with an inhibitor was selected as receptor and the active site was defined around the bound inhibitor with a radius of 10 Å. The co-crystal molecule was docked into the protein active site. The docked

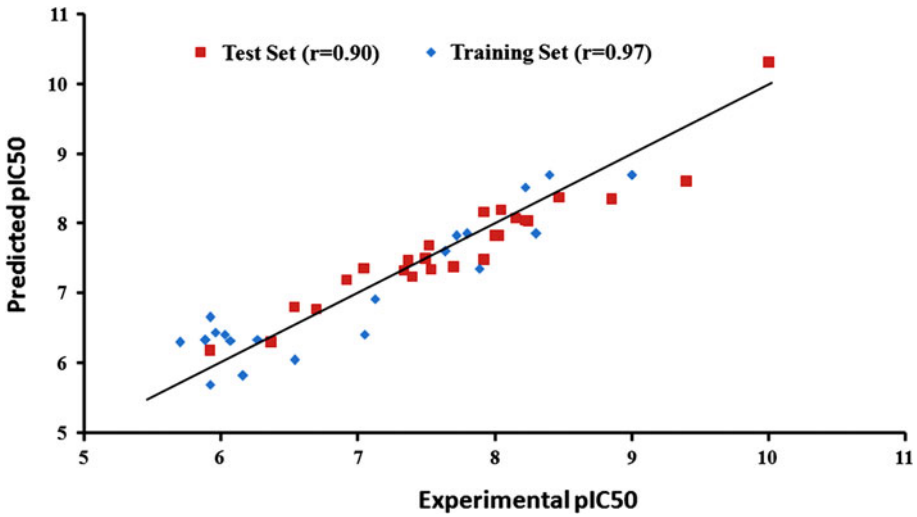

Fig. 7 Correlation (r) graph between the experimental activity and the predicted activity by Hypo1 for the test set molecules

Table 5 Statistical parameters of Hypo1 from screening the Decoy set

No.	Parameter	Values
1	Total number of molecules in database (D)	1,500
2	Total number of actives in database (A)	25
3	Total number of hit molecules from the database (H_t)	22
4	Total number of active molecules in hit list (H_a)	19
5	% Yield of actives [(H_a/H_t) × 100]	86.36
6	% Ratio of actives [(H_a/A) × 100]	76
7	Enrichment factor (EF)	5.181
8	False negatives [$A - H_a$]	6
9	False Positives [$H_t - H_a$]	3
10	Goodness of hit score[a] (GH)	0.83

[a] $[(H_a/4H_tA)(3A + H_t) \times (1 - ((H_t - H_a)/(D - A))]$; GH score of 0.7–0.8 indicated a very good model (Wolber and Langer, 2004)

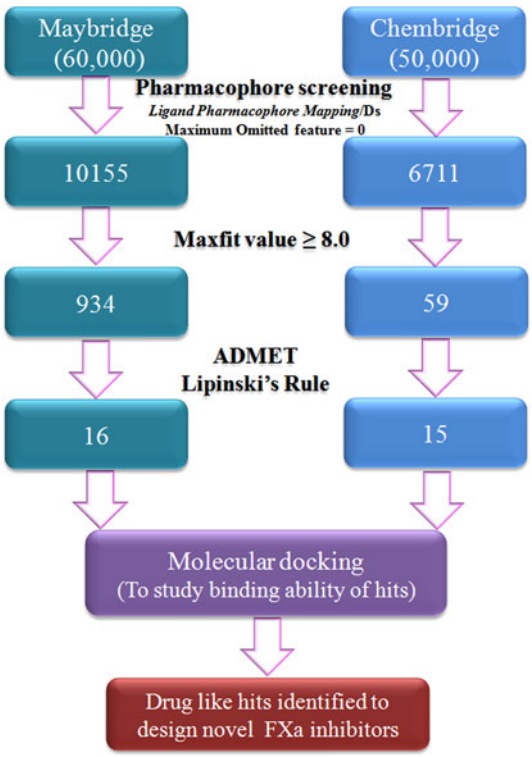

Fig. 8 Scheme of virtual screening

conformation was compared with inhibitor bound known crystal structure conformation which has shown very close to all atom with RMSD of 0.9 Å (Fig. 9). This has determined the reliability of GOLD program. The GOLD fitness score, binding mode, and molecular interaction of docked conformation of database hits were compared with rivaroxaban, the most active inhibitor of FXa (Borensztajn and Spek, 2011). Hence, rivaroxaban was first docked into the protein active site and calculated the GOLD fitness score of 78.84. It has shown very strong hydrogen bond interactions with S1 pocket crucial amino acids Gly216 (2.6 Å) and G218 (2.7 Å) as well as these two hydrogen bonds supported the ligand bind L-shape which is needed for FXa inhibitors (Fig. 10a). The S4 pocket (hydrophobic box) is characterized with aromatic ring of Y99, F174, and W215. The aryl ring of rivaroxaban showed face to face interaction with W215 and the morpholinone moiety is sandwiched between Y99 and F174. The chlorothiophene moiety of rivaroxaban interacts with the aromatic ring of Y228. This is key interaction of ligand and FXa since this chlorine–Y228 interaction enabled the combination of high potency and good oral bioavailability toward FXa (Roehrig *et al.*, 2005). The final hits from databases were docked into protein active site with same protocol which followed for rivaroxaban. The docked hits were selected based on the GOLD fitness score >78 and checked their binding mode and molecular interaction with catalytically important amino acids at protein active site. Two compounds from both databases, compound 33475 (Chembridge) and

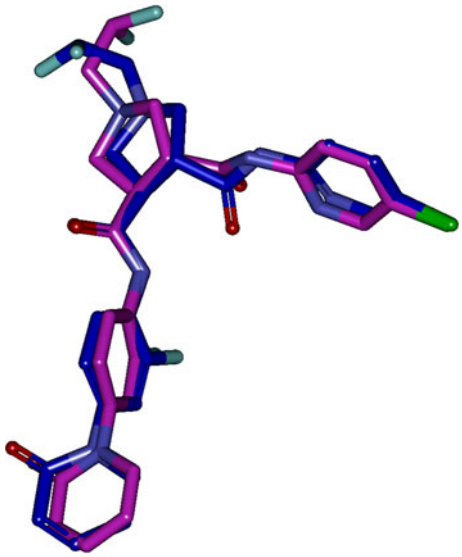

Fig. 9 Superposition of inhibitor compound with its crystal structure (2XBV), color coded GOLD prediction *pink* and bound *violet* (Color figure online)

HTS05096 (Maybridge), have scored GOLD fitness score of 93.98 and 87.39, respectively. The binding mode of both lead compounds has shown quite similar (L-shape) that of rivaroxaban (Fig. 10d). Moreover, the binding mode of these two compounds were once again checked by overlaid on the Hypo1 during database screening which showed very much close to the binding mode of rivaroxaban (Fig. 11). Some of drug-like database hit compounds with highest GOLD fit score have been rejected since not matching binding mode with rivaroxaban. The molecular interaction of hit compound, compound 33475, from Chembridge database has shown very good hydrogen bond with G216 (3.1 Å) and G218 (1.6 Å) in S1 pocket which is quite similar that of rivaroxaban (Fig. 10b). The piperazine moiety of compound 33475 stacking against phenyl moiety of F174 and indole derivative of W215 as well as this piperazine moiety is sandwiched between Y99 and F174 in S4 pocket. In addition, O–CH_3 group of phenoxy moiety involves key interaction with aromatic ring of Y228 in the bottom of S1 pocket. Maybridge compound, HTS05096, has shown the similar type of hydrogen bond interaction with G216 (2.9 Å) and G218 (2.8 Å) in S1 pocket (Fig. 10c). Further, the fluorine group of methyl pyrimidine derivative showed hydrogen bond with NH of indole derivative of W215 in S4 pocket this interaction was not present in the rivaroxaban. In addition, the methyl pyrimidine group was sandwiched between Y99 and F174 and also O–CH_3 group of phenoxy moiety involves key interaction with aromatic ring of Y228 in the bottom of S1 pocket. A similar kind of interaction has been observed in morpholinone group and chlorothiophene moiety of rivaroxaban. Based on the GOLD fitness score, binding mode, and molecular interaction, the database hit compounds have well competed with the already reported compound like rivaroxaban. The 2D representation of these database drug-like leads have depicted in Fig. 12. Hence, we concluded that these two compounds are quite good to design the novel potential new classes of FXa inhibitors. Further, the novelty of these compounds was checked through *Scifinder Scholar* (Wagner, 2006) and *Pubchem structure search* (Wang *et al.*, 2010) tools.

Bayesian model

The Bayesian model was developed using nine descriptors including ECFP_6 and eight interpretable descriptors. Training set molecules were divided as two categories: most active and inactive. The most active molecules (<12 nM) were designated as active with a value of "1" and remaining molecules designated as inactive with a value of "0". The developed Bayesian model was validated using leave-one-out cross-validation by which each compound of training set left out one at a time, a model built using rest of the compounds, and that model used assess the value of left out molecule. This process was repeated until all molecules had a prediction, and a receiver operating curve (ROC) plot was generated which used to estimated the predictiveness of the modeling process. The ROC curve demonstrates the model sensitivity such as the ability to identify true positives, and specificity, the ability to avoid false negatives. The area under the ROC curve is a quantitative measure of model's performance. A value of 1.0 represents the ability of perfectly discriminate between true positives and true negatives, while a value of 0.5 and below indicates that the model has no predictive ability (Klon *et al.*, 2006). In our study, the value of ROC curve was 0.929 which revealed that the model presents excellent prediction accuracy for the training set (Fig. 13). In addition, leave-one-out cross-validation and ROC curve was not only validated the developed model but also determined the efficiency of the training set molecules. Based on the validation, each and every molecules present in the training set are important for model generation. A contingency table is constructed using Best split, which is calculated by picking the split that minimized the sum of the percent misclassified for inhibitors and for non inhibitors. The cross-validated score for each compound, as well as number of true positive (TP), false negative (FN), false positive (FP), and true negative (TN) were also calculated and summarized in Table 6. Among 11 most active compounds, 10 were predicted as true positive and rest one was predicted as false negative. 13 out of 14 inactive compounds were predicted true positives and remaining one predicted as true positive. It was clearly determined the predictability of our model. Further, this model was

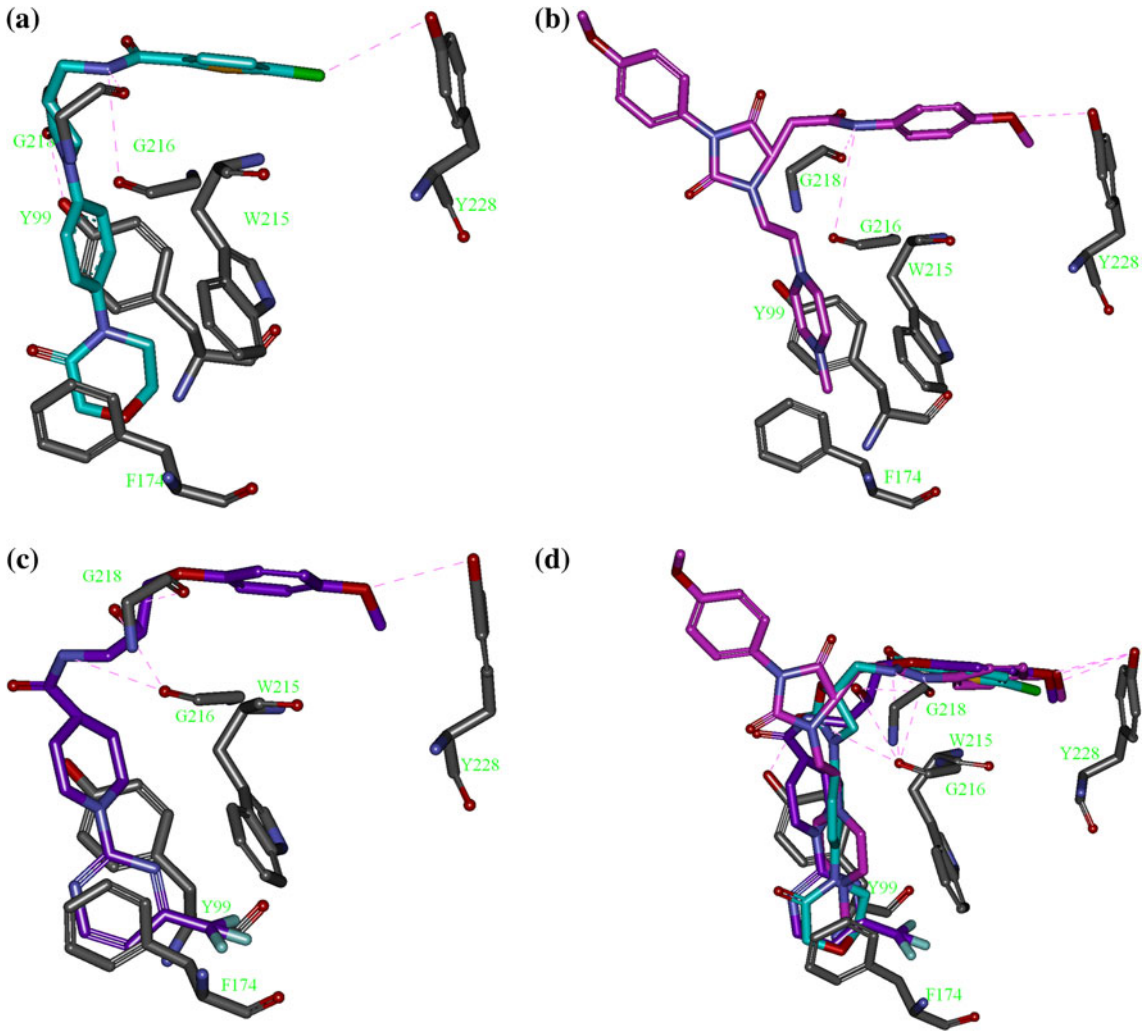

Fig. 10 **a** Molecular interaction of rivaroxaban with catalytically important residues, **b** drug-like lead Compound 33475 from Chembridge database with active site residues, **c** drug-like lead HTS05096 from Maybridge database with active site residues, **d** Binding mode overlay of rivaroxaban and drug-like leads Compound 33475 and HTS05096. Color code: residues are in elemental color, rivaroxaban *cyan*, Compound 33475 *pink*, HTS05096 *violet*, *dotted lines* (*pink*) denoted as hydrogen bond (Color figure online)

developed using ECFP_6 fingerprint descriptors and it has helped to identify the molecular feature that are favoring to inhibit FXa activity and not favoring to inhibit FXa activity. There were each 20 fingerprints that positively (favoring) and negatively (not favoring) contributed to inhibit function of FXa (Supplementary Figs. S1, S2).

Conclusion

FXa enzyme has proven to be exciting and promising target for thromembolic and proliferative diseases. Hence, to inhibit the FXa activity, we have developed chemical feature based pharmacophore model (Hypo1) using known inhibitors of FXa. The reliability of the Hypo1 was analyzed by various potential methods such as Fischer' method, test, and decoy set. The well-validated model was used as 3D query in Chembridge and Maybridge databases to retrieve potential virtual leads. The retrieved hits from databases were sort out by applying Maximum fit value of Hypo1 as well as drug-like prediction such as ADMET descriptors and Lipinski's rule-of-five. The hit molecules passed out above filtrations were once again validated by binding mode and molecular interaction in the protein active site in molecular docking study (GOLD v4.1). Finally, two drugs like hits were shown good Gold fit score, suitable binding mode and molecular interactions with catalytically important residues when compared with the most active compound of FXa (rivaroxaban). Furthermore, the novelties of these two drug leads (Compound 33475 & HTS05096) were confirmed by Scifinder scholar and Pubchem structure search tools. Together these results, two drug-like leads were used to design novel classes of FXa inhibitor. However, the success of this work could be

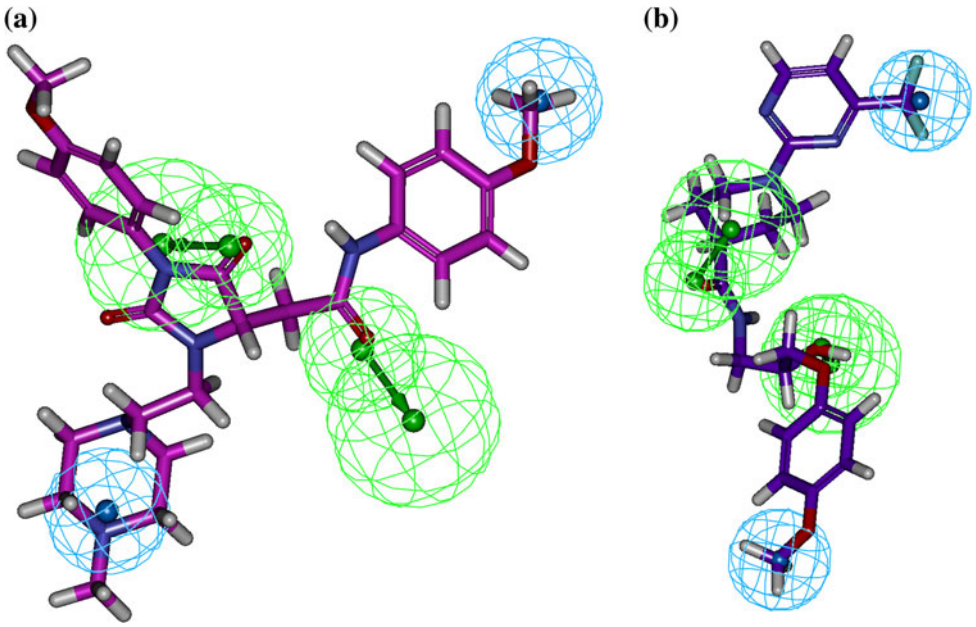

Fig. 11 Pharmacophore overlay of drug-like database lead compounds **a** compound 33475, **b** HTS05096

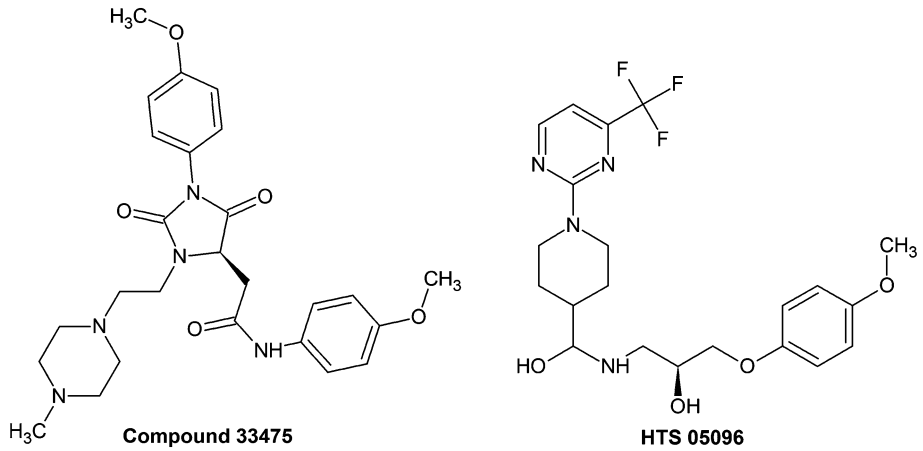

Fig. 12 2D representation of drug-like database hit compounds

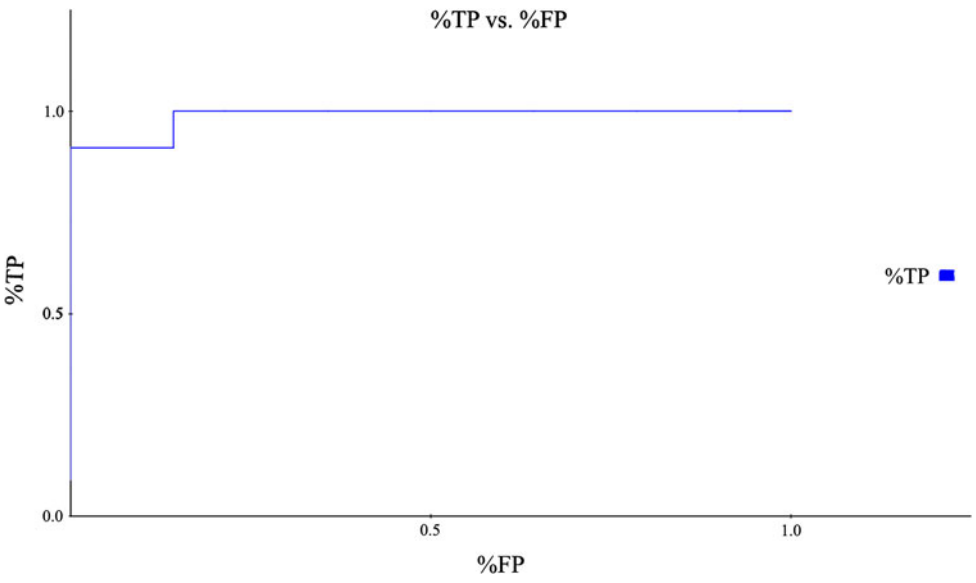

Fig. 13 Receiver operating curve (ROC) generated by leave-one-out cross-validation

Table 6 Leave-one-out cross-validation results

Output	XV ROC AUC	Best split	TP/FN FP/TN	#in Category
Bayesian temp model	0.929	16.08	10/1 1/13	11

determined through biological testing as well as optimize the hits subsequently. For the purpose, we have developed Bayesian model using finger print ECFP_6 and various molecular features as descriptors. Bayesian model was used to validate the efficiency of the training set molecules by leave-one-out cross validation method. In addition, the developed Bayesian model has disclosed molecular feature favoring and not favoring to inhibit FXa activity and it could be helpful to design new classes of inhibitors.

Acknowledgments This research was supported by Basic Science Research Program (2009-0073267), Pioneer Research Center Program (2009-0081539), and Management of Climate Change Program (2010-0029084) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (MEST) of Republic of Korea. And this study was also supported by the Next-Generation BioGreen21 Program (PJ008038) from Rural Development Administration (RDA) of Republic of Korea. Author thanks to Mr. Sundarapandian Thangapandian, Ph.D Scholar, Department of Biochemistry, and Gyeongsang National University, for his constant support in Bayesian Model part.

References

Adane L, Patel DS, Bharatam PV (2010) Shape- and chemical feature-based 3D-pharmacophore model generation and virtual screening: identification of potential leads for *P. falciparum* DHFR enzyme inhibition. Chem Biol Drug Des 75:115–126

Adler M, Davey DD, Phillips GB, Kim S-H, Jancarik J, Rumennik G et al (2000) Preparation, characterization, and the crystal structure of the inhibitor ZK-807834 (CI-1031) complexed with factor Xa. Biochemistry 39:12534–12542

Alban S (2005) From heparins to factor Xa inhibitors and beyond. Eur J Clin Invest 35:12–20

Ansell J (2007) Factor Xa or thrombin: is factor Xa a better target? J Thromb Haemost 5(suppl 1):65–67

Bajorath J (2002) Integration of virtual and high-throughput screening. Nat Rev Drug Discov 1:882–894

Borensztajn K, Spek CA (2011) Blood coagulation factor Xa as an emerging drug target. Exp Opin Ther Targets 15:341–349

Carreiro J, Ansell J (2008) Apixaban, an oral direct Factor Xa inhibitor: awaiting the verdict. Exp Opin Investig Drugs 17:1937–1945

de Candia M, Lopopolo G, Altomare C (2009) Novel factor Xa inhibitors: a patent review. Exp Opin Ther Patents 19:1535–1580

Fischer R (1966) The principle of experimentation illustrated by a psycho-physical experiment, Chap II. Hafner Publishing, New York

Gopalakrishnan B, Aparna V, Jeevan J, Ravi M, Desiraju GR (2005) A virtual screening approach for thymidine monophosphate kinase inhibitors as antitubercular agents based on docking and pharmacophore models. J Chem Inf Model 45(4):1101–1108

Hahn M (1997) Three-dimensional shape-based searching of conformationally flexible compounds. J Chem Inf Comput Sci 37: 80–86

Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748

Kansal N, Silakari O, Ravikumar M (2009) A three dimensional pharmacophore modeling for KDR and Tie-2 receptor tyrosine kinase inhibitors and virtual screening for new multikinase inhibitors. QSAR Comb Sci 28:1130–1147

Klon AE, Lowrie JF, Diller DJ (2006) Improved naïve Bayesian modeling of numerical data for absorption, distribution, metabolism and excretion (ADME) property prediction. J Chem Inf Model 46:1945–1956

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25

MacKerell AD, Bashford D, Bellott, Dunbrack RL, Evanseck JD, Field MJ et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102: 3586–3616

Putta S, Lemmen C, Beroza P, Greene J (2002) A novel shape-feature based approach to virtual library screening. J Chem Inf Comput Sci 42:1230–1240

Ravikumar M, Pavan S, Bairy S, Pramod AB, Sumakanth M, Kishore M et al (2008) Virtual screening of cathepsin K inhibitors using docking and pharmacophore models. Chem Biol Drug Des 72: 79–90

Roehrig S, Straub A, Pohlmann J, Lampe T, Pernerstorfer J, Schlemmer K-H et al (2005) Discovery of the novel antithrombotic agent 5-chloro-*N*-({(5S)-2-oxo-3-[4-(3-oxomorpholin-4-yl)phenyl]-1,3-oxazolidin-5-yl}methyl)thiophene-2-carboxamide (BAY 59–7939): an oral, direct factor Xa inhibitor. J Med Chem 48:5900–5908

Rogers D, Brown RD, Hahn M (2005) Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up. J Biomol Screen 10:682–686

Shi Y, Sitkoff D, Zhang J, Klei HE, Kish K, Liu ECK et al (2008) Design, structure–activity relationships, X-ray crystal structure, and energetic contributions of a critical P1 Pharmacophore: 3-chloroindole-7-yl-based factor Xa inhibitors. J Med Chem 51: 7541–7551

Shi Y, Li C, O'Connor SP, Zhang J, Shi M, Bisaha SN et al (2009) Aroylguanidine-based factor Xa inhibitors: the discovery of BMS-344577. Bioorg Med Chem Lett 19:6882–6889

Smellie A, Kahn SD, Teig SL (1995a) Analysis of conformational coverage. 1. Validation and estimation of coverage. J Chem Inf Comput Sci 35:285a–294a

Smellie A, Kahn SD, Teig SL (1995b) Analysis of conformational coverage. 2. Applications of conformational models. J Chem Inf Comput Sci 35:295b–304b

Smellie A, Teig SL, Towbin P (1995c) Poling: promoting conformational variation. J Comput Chem 16:171c–187c

Srivastava HM, Chourasia M, Kumar D, Narahari Sastry G (2011) Comparison of computational methods to model DNA minor groove binders. J Chem Inf Model 51:558–571

Taha MO, Qandil AM, Zaki DD, AlDamen MA (2005) Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm-based QSAR modeling. Eur J Med Chem 40:701–727

Wagner AB (2006) SciFinder Scholar 2006: an empirical analysis of research topic query processing. J Chem Inf Model 46:767–774

Walenga JM, Jeske WP, Hoppensteadt D, Fareed J (2003) Factor Xa inhibitors: today and beyond. Curr Opin Investig Drugs 4: 272–281

Wang Y, Bolton E, Dracheva S, Karapetyan K, Shoemaker BA, Suzek TO et al (2010) An overview of the PubChem BioAssay resource. Nucl Acids Res 38:D255–D266

Wolber G, Langer T (2004) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model 45:160–169

Yang Q, Du L, Tsai K-C, Wang X, Li M, You Q (2009) Pharmacophore mapping for Kv1.5 potassium channel blockers. QSAR Comb Sci 28:59–71

Yoshikawa K, Kobayashi S, Nakamoto Y, Haginoya N, Komoriya S, Yoshino T et al (2009) Design, synthesis, and SAR of cis-1,2-diaminocyclohexane derivatives as potent factor Xa inhibitors. Part II: Exploration of 6–6 fused rings as alternative S1 moieties. Bioorg Med Chem 17:8221–8233

Zhang J, Liu G, Tang Y (2009) Chemical function-based pharmacophore generation of selective κ-opioid receptor agonists by catalyst and phase. J Mol Model 15:1027–1041