DOI 10.5012/bkcs.2010.31.11.3333

Pharmacophore Design for Anti-inflammatory Agent Targeting Interleukin-2 Inducible Tyrosine Kinase (Itk)

Meganathan Chandrasekaran,[a] Sugunadevi Sakkiah,[a] Sundarapandian Thangapandian, Sundaraganesan Namadevan,[†] Hyong-Ha Kim,[‡] Yongseong Kim,[§] and Keun Woo Lee[*]

Division of Applied Life Science(BK21 Program), Environmental Biotechnology National Core research Center (EB-NCRC), Plant Molecular Biology and Biotechnology Research Center(PMBBRC), Gyeongsang National University, Jinju 660 701, Korea. [*]E-mail: kwlee@bio.gnu.ac.kr

[†]Department of Physics (FEAT), Annamalai University, Annamalai Nagar608 002, Tamilnadu, India

[‡]Division of Quality of Life, Korea Research Institute of Standards and Science, Daejeon 305-340, Korea

[§]Department of Science Education, Kyungnam University, Masan 631-701, Korea

Received May 25, 2010, Accepted September 15, 2010

A three dimensional pharmacophore model was generated for the molecules which are responsible for anti-inflammatory activities targeting Interleukin-2 inducible tyrosine kinase (Itk). 16 structurally diverse molecules were selected as training set to generate the hypotheses using Discovery Studio v2.1. The best hypothesis, *Hypo1*, comprises two hydrogen bond acceptor (*HBA*), one hydrophobic aromatic (*HA*), one ring aromatic (*RA*) and shows high cost difference (63.71), high correlation coefficient (0.97) as well as low RMS deviation (0.81). *Hypo1* has been further validated toward a test set, decoy set and Fischer's randomization method. Furthermore, *Hypo1* was used to screen NCI and Maybridge databases. Finally, 2 hit molecules were identified as potential leads against Itk, which may be useful for future drug development.

Key Words: Itk, Pharmacophore modeling, Virtual screening, Docking, HypoGen

Introduction

Asthma and rheumatoid arthritis are some of the inflammatory diseases caused by Interleukin-2 inducible tyrosine kinase (Itk) which affect 7%[1,2] and 1% of the world population, respectively. Itk belongs to Tec family of non-receptor tyrosine kinases, which are important components of antigen receptor signaling pathways in B-cells, T-cells, and mast cells.[3] It is expressed primarily in T-cells, natural killer, and mast cells.[4] It has been found that it plays an essential role in initiation of regular TCR (T-Cell antigen Receptor) signaling, vital for T-cell activation and proliferation. There are three distinct classes of non-receptor protein tyrosine kinases, namely Src family kinases (Lck and Lyn), Syk family kinases (ZAP-70 and Syk), and Tec family kinases (Itk, Txk, and Tec).[5] Among these three types, Itk is very important one for its critical activity such as phosporylated Itk required for phosporylation of PLC-γ5 and hence calcium mobilization and activation of downstream pathways.[6] A Gene knockout study reveals that $CD4^+$ T- cells mixed with lymphocyte reaction[7] or upon anti CD3 stimulation, which reduces cytokines IL-2, IL-4, IL-5 and IL-13 production, which leads inflammatory diseases such as rheumatoid arthritis and allergic asthma.[8] T-cell modulation could be achieved by selective inhibition of Itk which may be useful in treating inflammatory and autoimmune conditions. Selective inhibitors of Itk could be useful as an immunosuppressive and/or anti-inflammatory agent where T-cell activation contributes to the disease pathophysiology and an attractive modulator of dysregulated allergic pathway.

In this paper, we have reported a predictive pharmacophore based model[9] to find out the potential inhibitors for the Itk. First, we have collected selective and structurally diverse Itk inhibitors from various literatures.[10-12] Based on these collected Itk inhibitors, we have developed quantitative pharmacophore models whose purpose is to identify the critical pharmacophoric features necessary for potent Itk inhibitors as well as to clarify the quantitative structure-activity relationship for the known Itk inhibitors. Correlation between actual and estimated biological activities was calculated to optimize the hypotheses. Further, Pharmacophore model was evaluated as an activity prediction model by estimating the anti-inflammatory activities of a test set consisting of 12 molecules taken from literatures[11-14] and also by a decoy set. The optimized model was then validated by Fischer's randomization test at 95% confidence level. Then the best quantitative model was used as 3D query for searching or screening chemical databases, including NCI and Maybridge to identify new inhibitors of Itk. The hit compounds were subsequently subjected to filtering by Lipinski's rule of five,[15] ADME and molecular docking studies to refine the retrieved hits. It was expected that this established model should be capable of correctly reflecting the structure-activity relationship of Itk kinase inhibitors, as a result, be helpful in the identification of novel Itk inhibitors. However, known to date our work is the first report on quantitative pharmacophore modeling of Itk inhibitors.

Materials and Methods

Hypothesis generation. The entire calculations were performed in Discovery Studio v2.1 (DS). 3D-QSAR pharmacophore

[a]These authors contribute equally as first author.

models[16] were generated within DS from the training set compounds. The 2D structures of all compounds were drawn by using ChemSketch. These 2D structures were converted to 3D form and conformational models of all training set molecules were generated by applying "*Best quality*" conformational search option in DS using a constraint of a 20 kcal/mol energy threshold above the global minimum and CHARMm force field parameters. A maximum number of 255 conformation models were generated for each compound and these conformational models represent the flexibility of each compound and used not only in hypothesis generation but also fitting the compound to a hypothesis and for estimation of the activity of the compound.

In our study the energetically reasonable four features, which are hydrogen bond acceptor (*HBA*), hydrogen bond donor (*HBD*), hydrophobic (*HA*) and ring aromatic (*RA*) were used to generate pharmacophore hypotheses. The minimum and maximum counts for all the five features were kept 0 and 5, respectively. Uncertainty value, minimum points and minimum subset points were kept at default values. Three important cost calculations were carried out by *HypoGen* module in the unit of bits which determine the success of any pharmacophore hypothesis. First, fixed cost, which represents the simplest model, fits all data perfectly. Second, the null cost, that represents the highest cost of a pharmacophore with no features and estimates the activity to be the mean of the activity data of the training set compounds. Since the null pharmacophore has no features, there is no contribution from the weight or configuration component. A meaningful pharmacophore hypothesis may result when the difference between these two values (null cost − fixed cost) is more than at least 20. Third, total cost is the sum of three factors, a weight, an error and a configuration cost. The weight cost increases if the weight factor for the chemical features deviates from default value of 2. Error cost term is the most important one and it is dependent upon the root mean square (RMS) differences between estimated and measured activities of the training set compounds. The RMS value represents the quality of correlation between actual and estimated activity data. The entropy or configuration cost describes as log2P, where P is the number of initial hypotheses created in the constructive phase and that survived in the subtractive phase and it should not be more than 17.

Pharmacophore validation. The generated pharmacophore model is whether able to predict the activity of molecules accurately and also to discover the active molecules from databases. Hence the derived pharmacophore model should be validated by using (1) Test set (2) Decoy set and (3) Fischer's method.

Test and decoy sets: Two statistical methods were employed to find out the predictive ability of the selected hypothesis. In the first method, the predictive ability of *Hypo1* was determined from a set of 12 compounds that were used in test set which are not included in the training set. All the test set molecules were built, minimized and conformational models were generated using same protocols like training set molecules. Based on the activity range of test set molecules were also classified into three categories highly active, moderately active and inactive molecule. The second statistical method was decoy set which includes 1100 molecules. The parameters such as false positives, false negatives, enrichment factor (EF), and goodness of hit (GH) were calculated to determine the robustness of hypothesis.

Fischer's method: Fischer's randomization method was used to measure the statistical significance of our model. Fischer method uses the experimental activity data associated with the training set is scrambled randomly. Then the randomized data sets were used to generate hypothesis by using the same feature and parameters which was used in original hypothesis generation. The randomized data sets should generate hypotheses without statistical relevance, otherwise the original model has been considered to generate by a chance. The following formula is used to calculate the statistical significance:

$$\text{Significance} = 100(1 - 1 + x/y)$$

Where 'x' is the total number of hypotheses having total cost lower than the most-significant hypothesis and 'y' is the number of initial *HypoGen* runs plus random runs. In our study, 19 random spreadsheets were generated to achieve the 95% of confidence level.

Virtual screening. Our aim is to identify the potent inhibitors for Itk from the databases. For conducting virtual screening, we used two publically avail databases, namely, NCI[17] and Maybridge,[18] which comprise a large collection of synthetic compounds, including more than 2,00,000 and 50,000 molecules, respectively, along with their conformation to discover the novel inhibitors of Itk. There are two methods available in virtual screening technique, one is ligand based virtual screening and the other one is structure based virtual screening.[19] Pharmacophore based database screening is a ligand based virtual screening method, which can efficiently identify the novel potential leads from databases for further development. The best pharmacophore hypothesis *Hypo1* was used as query in the virtual screening to retrieve the potential leads from databases, using *Fast Flexible* search method with in DS. The hit compounds from each database were selected based on Lipinski's rule of five and ADME properties.

Molecular docking protocol. Molecular docking study was performed using *LigandFit* within DS. A systematic search of positions, orientations and conformations of the ligand in the protein-binding pocket was carried out through a series of hierarchical filters by shape based ligand docking. Preparation of protein for molecular docking includes the removal of ligand and water molecules coupled with addition of hydrogen atoms. An initial ligand conformation was generated by using the monte-carlo procedure, in which nearby torsional minima is examined, and the orientation of peripheral groups of the ligand is refined. The crystal structure of Itk-staurosporine complex was obtained from Protein Data Bank (www.rcsb.org, PDB ID: 1SM2). The total conformational search was carried out for all molecules using CHARMm force field, and imposing a cutoff of allowed value of the total conformational energy compared to the lowest-energy state. The RMSD threshold, energy threshold and grid resolution were set to 2.0Å, 20 kcal/mol and 0.5Å, respectively. The scoring functions including LigScore1, LigScore2,[20] PLP1, PLP2,[21] Jain,[22] PMF[23] and Ludi[24,25] were used to evaluate each of the saved conformations.

Results and Discussion

Generation of pharmacophore. A good set of diverse inhibitors of a particular target with a wide range of activity data is needed to generate quantitative pharmacophore models. A set of 28 Itk inhibitors have been taken from literatures, from which, 16 compounds were carefully chosen as training set, which was based on the principles of structural diversity and wide coverage of activity (IC_{50}) that spans over 4 orders of magnitude i.e. 9 nM to 17000 nM. The structure and biological activity of training set molecules have shown in Fig. 1. The remaining compounds were used as a test set which was used to validate the selected pharmacophore. The training set was roughly classified into three catagories such as highly active, moderately active and inactive compounds to attain important information on pharmacophoric necessities for Itk inhibition.

This training set was used to generate pharmacophore models using *HypoGen* algorithm avail in DS. During the hypothesis generation phase, pharmacophore features such as *HBA*, *HBD*, *HA*, and *RA* were considered. The top ten hypotheses composed of above mentioned features were generated. The generated hypotheses were analyzed using Debnath[26] in terms of cost function and statistical parameters which were calculated by *HypoGen* module during hypothesis generation. Statistical parameters of top ten hypotheses are summarized in Table 1, among all ten hypotheses the best hypothesis *Hypo1*, shown in Fig. 2, generated with four features: two *HBA*, one *HA*, one *RA*. *Hypo1* is then characterized by highest cost difference between null and total cost value of 63.21, correlation coefficient of 0.97 and the lowest RMS deviation of 0.81 as well as the good configuration cost value of 16.17. Furthermore, the null cost and fixed cost value of 133.69, and 70.48, supports that this pharmacophore is of statistical significance and could be considered as a reliable pharmacophore model for further studies. The *Hypo1* was subsequently used to calculate the inhibitory activities of the 16 training set compounds. The error value was calculated as difference between experimental and predicted activity values. This difference may have negative sign when the experimental value is higher than the predicted value and positive sign when the experimental value is lower than the predicted value. All compounds in the current study were grouped into highly active (+++), moderately active (++) and inactive compounds (+). Table 2 shows all the compounds that were predicted accurately except one, compound14, which is experimentally inactive but predicted as moderately active compound. Fig. 3a, b represents the *Hypo1* aligned with the most-active compound1 (IC_{50}: 9 nM) and least active compound15 (IC_{50}: 4500 nM) in the training set, respectively.

Pharmacophore validations. A good pharmacophore model should have the capability to predict the activity of training set

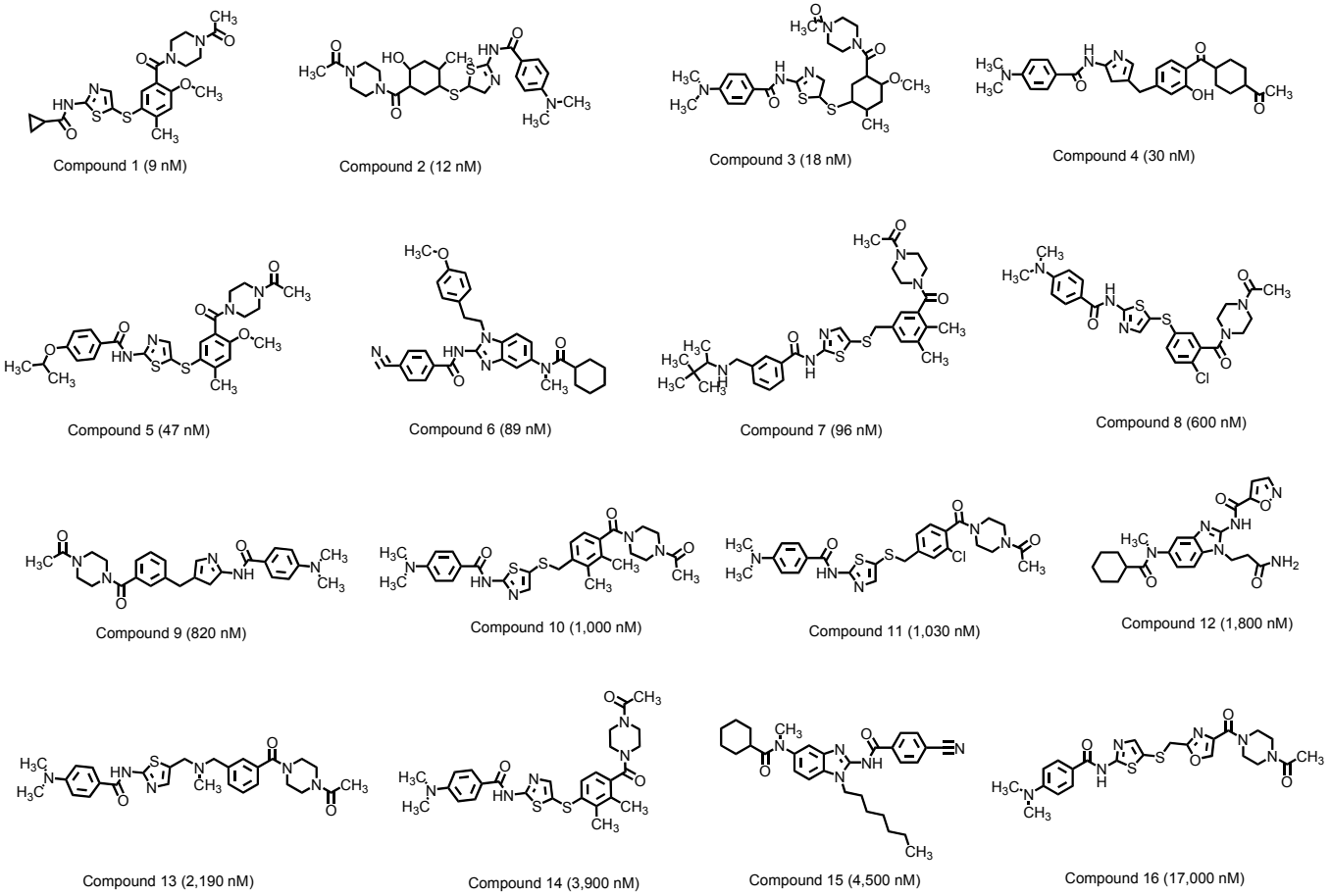

Figure1. Training set of Itk inhibitors together with their biological activity data (IC_{50} values, nM) for *HypoGen* run.

Table 1. Statistical parameters of the top10 Hypotheses

Hypothesis No.	Total Cost	Cost Difference[a]	RMS[b]	Correlation	Fit Value	Features[b]
Hypo1	70.48	63.21	0.81	0.97	10.10	*2HBA ,HA RA*
Hypo2	77.14	56.54	1.25	0.92	9.49	*2HBA ,HA RA*
Hypo3	85.73	47.95	1.65	0.86	8.65	*2HBA ,HA RA*
Hypo4	85.81	47.87	1.66	0.86	8.04	*3HBA,HA*
Hypo5	88.61	45.08	1.75	0.84	8.77	*3HBA,HA*
Hypo6	89.62	44.07	1.79	0.83	7.97	*2HBA,HBD,HA*
Hypo7	90.19	43.50	1.78	0.84	9.69	*2HBA,HBD,HA*
Hypo8	90.77	42.91	1.83	0.83	8.33	*3HBA,HA*
Hypo9	90.84	42.84	1.84	0.83	7.71	*2HBA ,HA RA*
Hypo10	91.35	42.34	1.83	0.83	6.49	*2HBA,HBD,HA*

[a]Cost difference between the null and the total cost = 63.21, null cost = 133.694, fixed cost = 70.486, and configuration cost = 16.17. [b]Abbreviation used for features: RMS = root mean square deviation; *HBA* = hydrogen bond acceptor, *HA* = hydrophobic aromatic, *RA* = Ring aromatic, *HBD* = hydrogen bond donor.

Table 2. Experimental and predicted IC_{50} data of training set molecules against *Hypo1* model

Compound No.	Fit Value[a]	Experimental IC_{50} nM	Predicted IC_{50} nM	Error[b]	Experimental scale[c]	Predicted scale[c]
1	9.00	9	10	+ 1.1	+++	+++
2	8.47	12	33	+ 2.8	+++	+++
3	8.79	18	16	− 1.1	+++	+++
4	8.37	30	42	+ 1.4	+++	+++
5	8.39	47	40	− 1.2	+++	+++
6	7.61	89	240	+ 2.7	+++	+++
7	8.11	96	77	− 1.2	+++	+++
8	7.00	600	990	+ 1.7	++	++
9	7.16	820	680	− 1.2	++	++
10	7.26	1000	540	− 1.9	++	++
11	7.25	1000	550	−1.9	++	++
12	6.66	1800	2200	+1.2	++	++
13	6.82	2200	1500	−1.5	++	++
14	6.93	3900	1200	−3.3	+	++
15	6.20	4500	6300	+1.4	+	+
16	5.75	17000	18000	+1	+	+

[a]Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule. Fit=weight x [max(0,1 –SSE)] where SSE = (D/T)2, D = displacement of the feature from the center of the location constraints and T = the radius of the location constraint sphere for the feature (tolerance). [b]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 indicates that the predicted IC_{50} is equal to the experimental IC_{50}. [c]Activity scale : IC_{50} <300 nM = +++ (highly active); 300 nM ≤ IC_{50} < 3000 nM = ++ (moderately active); IC_{50} ≥ 3000 nM = + (low active).

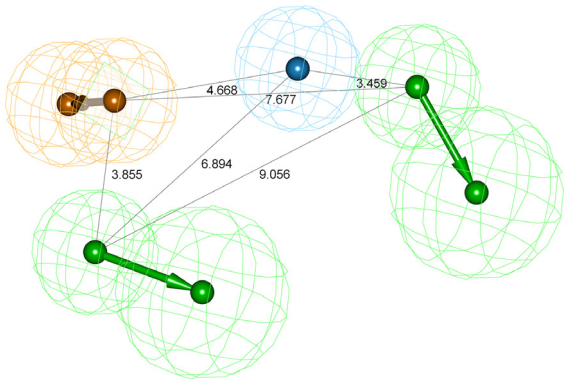

Figure 2. Geometrical parameters of pharmacophore model of Itk inhibitors by *HypoGen* (Green: hydrogen bond acceptor (*HBA*), Orange: ring aromatic (*RA*), Blue: hydrophobic aromatic (*HA*)).

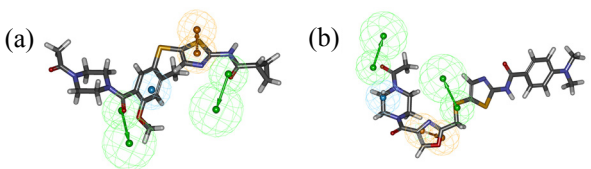

Figure 3. (a) *Hypo1* aligned with most active compound 1(IC_{50} : 9 nM), (b) *Hypo1* aligned with least active compound 15 (IC_{50}: 4500 nM) (Green: hydrogen bond acceptor (*HBA*), Orange: ring aromatic (*RA*), Blue: hydrophobic aromatic (*HA*)).

compounds and also external test set.[27,28] In our study we used three methods to ascertain the predictive ability of *Hypo1*. In the first method, an independent test set was used to analyze the established model (*Hypo1*), which contains 12 structurally different compounds from the training set molecules. *Hypo1*

Table 3. Experimental and predicted IC_{50} data of 12 test set molecules against *Hypo1* model

Compound No.	Fit Value[a]	Experimental IC_{50} nM	Predicted IC_{50} nM	Error[b]	Experimental scale[c]	Predicted scale[c]
T1	9.00	9	9.68	+1.08	+++	+++
T2	9.00	9	9.68	+1.08	+++	+++
T3	7.27	12	29.99	+2.50	+++	+++
T4	7.80	1000	758.69	+1.32	+++	++
T5	7.39	260	151.89	-1.71	+++	+++
T6	8.84	600	524.39	-1.14	++	++
T7	7.15	700	188.38	-3.72	++	+++
T8	8.51	820	683.68	-1.20	++	++
T9	7.19	900	2403.63	+2.67	++	++
T10	7.29	1000	497.76	-2.01	++	++
T11	7.71	1440	2356.74	+1.64	++	++
T12	7.73	2200	1364.97	-1.61	++	++

[a]Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule. Fit = weight x [max(0,1 -SSE)] where SSE = (D/T)2, D = displacement of the feature from the center of the location constraints and T = the radius of the location constraint sphere for the feature (tolerance). [b]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 indicates that the predicted IC_{50} is equal to the experimental IC_{50}. [c]Activity scale : $IC_{50} < 300$ nM = +++ (highly active); 300 nM $\leq IC_{50} < 3000$ nM = ++ (moderately active); $IC_{50} \geq 3000$ nM = + (low active).

Table 4. Statistical parameters from screening of *Hypo1*

No.	Parameter	Values
1	Total number of molecules in database (D)	1100
2	Total number of actives in database (A)	19
3	Total number of hit molecules from the database (Ht)	18
4	Total number of active molecules in hit list (Ha)	15
5	% Yield of actives [(Ha/Ht) X 100]	83.33
6	% Ratio of actives [(Ha/A) X 100]	78.94
7	Enrichment Factor (EF)	48.24
8	False negatives [A-Ha]	4
9	False Positives [Ht – Ha]	3
10	Goodness of Hit score[a] (GH)	0.82

[a]$[(H_a/4H_tA)(3A + H_t) \times (1- ((H_t - H_a)/(D-A))]$, GH above 0.7 indicate very good model

regressed against the 12 test compounds and gave a correlation coefficient of 0.94 between experimental and predicted activities values. The experimental, predicted activities and error values of test set are summarized in Table 3. From this table, the experimental and predicted inhibitory activity values for the test set compounds show quite good predictive ability of *Hypo1* for all compounds except compounds T4 & T7. Compound T4 is experimentally highly active but predicted as moderately active and compound T7 is experimentally moderately active but predicted as highly active. These results suggest that *Hypo1* reflects a reliable model to accurately predict the activities of both the training and test set compounds. The second method is calculating various statistical parameters to validate *Hypo1* hypothesis by screening a decoy set spiked with some known inhibitors. This step is to determine how many active molecules are picked in the screening process, and the efficiency in reducing the false positives or false negatives. Totally, 1100 molecules, comprised of 1081 unknown molecules and 19 known inhibitors of Itk, were used in decoy set to screen the *Hypo1* by using *Best flexible* screening technique. The parameters include total hit molecules (Ht), percentage yields of molecules, percentage ratio of actives in the hit list, enrichment factor (E), false negatives, false positives and goodness of hit score (GH) are calculated and given in Table 4. *Hypo1* was successfully retrieved 18 molecules from the database in that 15 (83%) molecules were known inhibitors for Itk and then it also recall 3 inactive compounds (false positive) and predicted 4 active compounds as inactive (false negative). An enrichment factor (EF) and goodness of hit (GH) were calculated using below mentioned formula

$$\%A = (H_a/A) \times 100$$

$$\%Y = (H_a/H_t) \times 100$$

$$E = (H_a/H_t)/(A/D)$$

$$GH = ((H_a\ (3A + H_t))/(4H_tA))\ (1- ((H_t - H_a) / (D - A)))$$

Where H_t = the number of hits retrieved, H_a = the number of active molecules in the hit list, A = the number of active molecules present in the database, and D = the total number of molecules in the database and the value is 48.24, 0.82. As a result, the *Hypo1* has a superior ability to identify false positive and could differentiate to high structural similarity in active and inactive Itk inhibitors.

In the third method, Fischer's randomization test[29] was used to further cross-validate the consistency and statistical significance of *Hypo1* using DS. It checks whether there is a strong correlation between the structures and activity values by mixing up activity values of all training set molecules. The purpose of this cross-validation is to randomize the activity data among the training set compounds and generate random pharmacophore models using the same features and parameters as used in developing the original *Hypo1* hypothesis. In that way the confidence level was set to 95% (significance = $[1 - (1 + 0)/ (X + 1)] - 100\% = 95\%$), 19 random spreadsheets were generated to construct the hypotheses and the results are summarized in Table 5. One can see that after randomization, all 19 random

Table 5. Results from cross validation using CatScramble implemented in DS[a]

Validation No.	Total cost	RMS	Correlation	Cost difference
Hypo1	70.48	0.810	0.970	63.51
Results for Scrambled				
Random1	91.68	1.85	0.827	42.01
Random2	104.48	2.16	0.760	29.21
Random3	102.74	2.21	0.742	29.21
Random4	75.66	1.01	0.931	58.03
Random5	98.71	2.01	0.794	34.98
Random6	90.76	1.68	0.864	43.69
Random7	102.04	2.06	0.787	31.65
Random8	95.97	2.04	0.785	37.72
Random9	80.92	1.40	0.899	52.77
Random10	84.09	1.48	0.895	49.60
Random11	82.09	1.46	0.898	51.60
Random12	101.87	2.14	0.763	31.82
Random13	74.48	1.14	0.930	59.21
Random14	105.03	2.18	0.754	28.66
Random15	88.65	1.67	0.866	45.04
Random16	89.46	1.64	0.875	44.23
Random17	95.66	1.84	0.836	38.03
Random18	89.18	1.76	0.844	44.51
Random19	90.74	1.80	0.830	42.95

[a]Null cost = 133.694

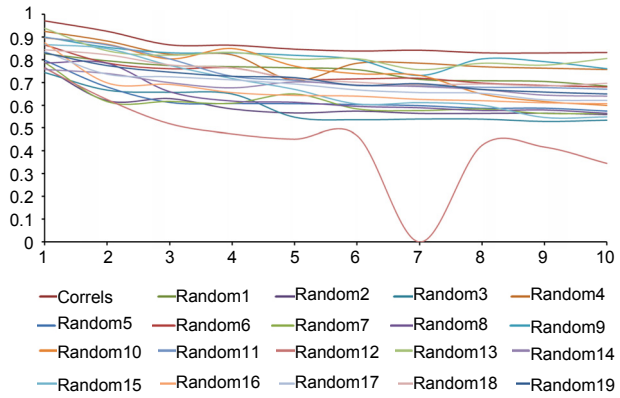

Figure 4. The difference in correlations of hypotheses between the initial spreadsheet and 19 random spreadsheets after CatScramble run.

spread sheets are much higher than the total costs and lower correlation coefficient of corresponding original pharmacophore model, *Hypo1*. Only two random hypotheses, 4 and 13 had correlation values more than 0.9, but the RMS deviation was high and the cost difference was below 60 and not desirable to be good hypothesis. The correlation graph between initial spread sheet and 19 random spread sheets was plotted and depicted in Fig. 4. The results of Fischer's randomization test clearly express that the *Hypo1* is not generated by chance and its values are far more superior than those of the 19 randomly produced hypotheses, which provide confidence on our pharmacophore hypothesis. Accordingly, all above validation methods demonstrate the robustness of the selected *Hypo1* hypothesis and conclude that *Hypo1* is far more superior than other

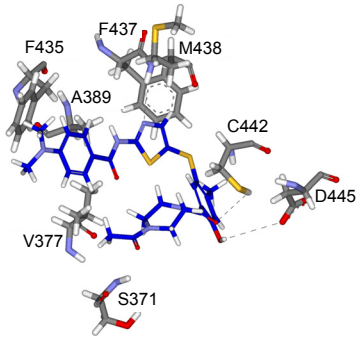

Figure 5. The molecular docking result: The docked compound 1 of training set shown good interaction with the catalytic residue (M438).

hypotheses.

Virtual screening. Virtual screening has been investigated as an *in silico* tool for drug discovery because of this study is not only to construct the pharmacophore hypothesis to predict compound activities, hence the pharmacophore model was employed in virtual screening to search for potential leads to be used in designing of Itk inhibitors. *Hypo1* was used as a 3D query to retrieve potential leads from chemical databases such as NCI and Maybridge by using *Fast Flexible* search in DS. A result, total of 31,890 and 2,460 molecules from NCI and Maybridge databases showed very good agreement with the *Hypo1*. Furthermore, these compounds were screened by using Lipinski's rule of five,[15,30] and ADME properties which are a simple model to forecast the absorption and intestinal permeability of a compound. 36 and 4 molecules passed this filtration from NCI and Maybridge databases, respectively.

Molecular docking. The X-ray crystal structure of Itk (PDB ID: 1SM2) was used in our molecular docking studies.[31] The binding pocket of Itk has a glycine-rich loop which contains the consensus kinase sequence Gly-X-Gly-X-X-Gly. Previous docking study on Itk states that the binding of the ligand in Itk kinase domain at the interface of the N- and C-terminal lobes mainly shows hydrophobic interactions with I369, V377 and A389 (N-terminal), F435 and Y437 (hinge region) and L489 and C442 (C-terminal lobe) among these residue F435 acts as a gate keeper, makes beneficial edge face interaction with conjugated system of ligand and additional key residues in the active site such as R486, E436, M438. These three residues form hydrogen bond (H-bond) interactions with inhibitor.[31] All the compounds including training set of 16 molecules and 40 new hit molecules retrieved from the NCI and Maybridge databases were docked into the active site of Itk. The bound conformations inside the Itk active site were visually examined. Albeit a significant correlation between dock score and inhibitory activities of the compounds was not routinely observed, in general, a trend was seen that among all the scoring functions, the Ligscore performed better, in which many of the active compounds scored high. Thus in our work Ligscore was used to sort the molecules. Ten distinct poses of each ligand in the active site of Itk were generated. The docked conformation of one of the highly active compounds in the training set has shown the dock score value of (90.01) and formed hydrophobic interactions with essential active site residues as well as Hydrogen bond interactions with critically important M438,

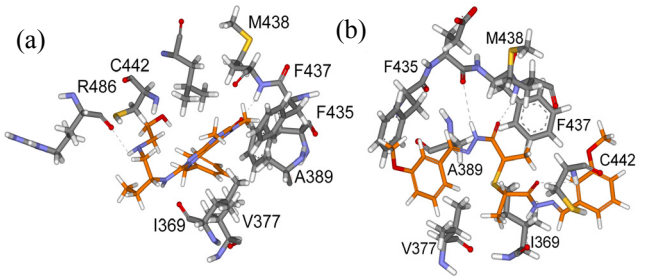

Figure 6. The molecular docking results (a) NCI compound NCI-0056737, it shows the good interaction with the catalytic residues (M438, Glu436), (b) Maybridge compound RJC03502, and it also shows good interaction with the catalytic residues (M438, R486).

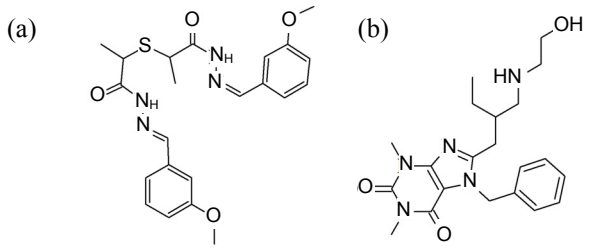

Figure 7. Molecular structure of (a) NCI0056737 (NCI), (b) RJC03502 (Maybridge).

which is more similar to recent docking studies on Itk.[31] Fig. 5 represents the H-bond interaction of the most active compound in the active site of Itk. Some of the hits retrieved from databases have also shown the hydrophobic and H-bond interaction with critically important residues in the active site of Itk. 25 out of 40 compounds, which were selected based on dock score ≥ 90, considered as final hits for further evaluation. Compound NCI0056737 retrieved from NCI database, showed good *HypoGen* estimation activity of (6.3 nM) as well as dock score value of (101.69) and formed two conserved bond with M438, E436. RJC03502 which is retrieved from Maybridge database showed dock score of (99.75), estimated activity of (4.6 nM) formed H-bond interaction with M438, R486 (Fig. 6a, b). Both the compounds (Fig. 7a, b), which are retrieved from NCI and Maybridge databases, showed good interaction with essential residues of Itk and better dock score values than the best active compound from training set. All other properties like estimated activity, binding affinity, calculated drug-like properties and thus these compounds can be treated as potential leads in the designing of potent inhibitors of Itk.

Conclusion

In our study, we have reported pharmacophore modeling of Itk inhibitors, using *HypoGen* algorithm available in DS software package. Ten hypotheses were generated using training set of 16 structurally diverse compounds. The best pharmacophore hypothesis *Hypo1*, which was characterized with the lowest RMS deviation, and highest correlation co-efficient, comprising of two *HBA*, one *HA*, and one *RA* features. Further this pharmacophore model was validated by test set, statistical method (decoy set) and Fisher randomization test to confirm the statistical confidence. All of these validation results suggested that *Hypo1* has the ability to accurately predict Itk inhibitors. Then *Hypo1* was used as a 3D query to screen NCI and Maybridge databases. Subsequently the hit compounds were filtered using Lipinski's rule of five, ADME properties and molecular docking study to refine the retrieved hits. Finally, 2 hits were identified, and then our pharmacophoric determination, docking experiments for Itk inhibitors can provide researchers valuable insight in rational drug design.

Acknowledgments. This research was supported by Basic Science Research Program (2009-0073267), Pioneer Research Center Program (2009-0081539), and Environmental Biotechnology National Core Research Center program (20090091489) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (MEST). And all students were recipient of fellowship from the BK21 Program of MEST.

References

1. Merck Manual Home Edition.
2. Merck Manual Professional Edition.
3. Smith, C. I. E.; Islam, T. C.; Mattsson, P. T.; Mohamed, A. J.; Nore, B. F.; Vihinen, M. *BioEssays* **2009**, *23*, 436.
4. Das, J.; Liu, C.; Moquin, R. V.; Lin, J.; Furch, J. A.; Spergel, S. H.; McIntyre, K. W.; Shuster, D. J.; O'Day, K. D.; Penhallow, B.; Chen-Yi Hung.; Kanner, S. B.; Lin, T. A.; Dodd, J. H.; Barrish, J. C.; Wityak, J. *Bioorg. Med. Chem. Lett*. **2006**, *16*, 2411.
5. Snow, R. J.; Abeywardane, A.; Campbell, S.; Lord, J.; Kashem, M. A.; Khine, H. H.; King, J.; Kowalski, J. A.; Pullen, S. S.; Roma, T.; Roth, G. P.; Sarko, C. R.; Wilson, N. S.; Winters, M. P.; Wolak, J. P.; Cywin, C. L. *Bioorg. Med. Chem. Lett*. **2007**, *17*, 3660.
6. Kawakami, Y.; Kitaura, J.; Hata, D.; Yao, L.; Kawakami, T. *J. Leukoc. Biol*. **1999**, *65*, 286.
7. Liao, X. C.; Littman, D. R. *Immunity* **1995**, *3*,757.
8. Forssell, J.; Sideras, P.; Eriksson, C.; Malm-Erjefalt, M.; Rydell-Tormanen, K.; Ericsson, P. O.; Erjefalt, J. S. *Am. J. Respir. Cell Mol. Biol*. **2005**, *32*, 511.
9. Acharya, B. N.; Kaushik, M. P. *Med. Chem. Res*. **2007**, *16*, 213.
10. Cook, B. N.; Bentzien, J.; White, A.; Nemoto, P. A.; Wang, J.; Man, C. C.; Soleymanzadeh, F.; Khine, H. H.; Kashem, M. A.; Kugler S. Z., Jr.; Wolak, J. P.; Roth, G. P.; De Lombaert, S.; Pullen, S. S.; Takahashi, H. *Bioorg. Med. Chem. Lett*. **2009**, *19*, 773.
11. Moriarty, K. J.; Takahashi, H.; Pullen, S. S.; Khine, H. H.; Sallati, R. H.; Raymond, E. L.; Woska, J. R., Jr.; Jeanfavre, D. D.; Roth, G. P.; Winters, M. P.; Qiao, L.; Ryan, D.; Jarlais, R. D.; Robinson, D.; Wilson, M.; Bobko, M.; Cook, B. N.; Yin Lo, H.; Nemoto, P. A.; Kashem, M. A.; Wolak, J. P.; White, A.; Magolda, R. L.; Tomczuk, B. *Bioorg. Med. Chem. Lett*. **2008**, *18*, 5545.
12. Yin Lo, H.; Bentzien, J.; White, A.; Man, C. C.; Fleck, R. W.; Pullen, S. S.; Khine, H. H.; King, J.; Woska, J. R., Jr.; Wolak, J. P.; Kashem, M. A.; Roth, G. P.; Takahashi, H. *Tetrahedron Lett*. **2008**, *49*, 7337.
13. Das, J.; Furch, J. A.; Liu, C.; Moquin, R. V.; Lin, J.; Spergel, S. H.; McIntyre, K. W.; Shuster, D. J.; O'Day, K. D.; Penhallow, B.; Chen-Yi Hung.; Doweyko, A. M.; Kamath, A.; Zhang, H.; Marathe, P.; Kanner, S. B.; Tai-An Lin.; Dodd, J. H.; Barrish, J. C.; Wityak, J. *Bioorg. Med. Chem. Lett*. **2006**, *16*, 3706.
14. Yin Lo, H.; Bentzien, J; Fleck, R. W.; Pullen, S. S.; Khine, H. H.; Woska, J. R., Jr.; Kugler, S. Z.; Kashem, M. A.; Takahashi, H. *Bioorg. Med. Chem. Lett*. **2008**, *18*, 6218.
15. Lipinski, C. A. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235.

16. Lu, I. L.; Tsai, K. C.; Chiang, Y. K.; Jiang, W. T.; Wu, S. H.; Mahindroo, N.; Chien, C. H.; Lee, S. J.; Chen, X.; Chao, Y. S.; Wu, S. Y. *Eur. J. Med. Chem.* **2008**, *43*, 1603.
17. Developmental Therapeutics Program, NCI/NIH, 2000, http://dtp.nci.nih.gov.
18. Maybridge Chemical company (England); http://www.chem.ac.ru/Chemistry/Databases/MAYBRIDGE.en.html.
19. Li, H. F.; Lu, T.; Zhu, T.; Jiang, Y. J.; Rao, S. S.; Hu, L. Y.; Xin, B. T. *E. J. Med. Chem*. **2009**, *44*, 1240.
20. Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. *J. Mol. Graph. Model*. **2003**, *21*, 289.
21. Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. R.; Fogel, L. J.; Freer, S. T. *Chem. & Biol.* **1995**, *2*, 317.
22. Jain, A. N. *J. Comput. Aid. Mol. Des*. **1996**, *10*, 427.
23. Muegge, I.; Martin, Y. C. *J. Med. Chem*. **1999**, *42*, 791.
24. Bohm, H. J. *J. Comput. Aid. Mol. Des*. **1994**, *8*, 243.
25. Bohm, H. J. *J. Comput. Aid. Mol. Des*. **1998**, *12*, 309.
26. Debnath, A. K. *J. Med. Chem*. **2002**, *45*, 41.
27. Sakkiah, S.; krishnamoorthy, N.; Gajendraro, P.; Thangapandian, S.; Lee, Y.; Suh, J.; Kim, H. H.; Lee, K. W. *Bull. Korean Chem. Soc.* **2009**, *30*, 1152.
28. Lee, Y.; Bharatham, N.; Bharatham, K.; Lee, K. W. *Bull. Korean. Chem. Soc*. **2007**, *28*, 561.
29. Fischer, R. A. *The Design of Experiments*; Hafner Publishing: New York, 1966 (Chapter 2).
30. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **2001**, *46*, 3.
31. Brown, K.; Long, J. M.; Vial, S. C. M.; Dedi, N.; Dunster, N. J.; Renwick, S. B.; Tanner, A. J.; Dan Frantz, J.; Fleming, M. A.; Cheetham, G. M. T. *J. Biol. Chem.* **2004**, *279*, 18727.