

# Discovery of potent inhibitors for interleukin-2-inducible T-cell kinase: structure-based virtual screening and molecular dynamics simulation approaches

Chandrasekaran Meganathan · Sugunadevi Sakkiah · Yuno Lee · Jayavelu Venkat Narayanan · Keun Woo Lee

Received: 6 January 2012 / Accepted: 12 July 2012 / Published online: 27 September 2012  
© Springer-Verlag 2012

**Abstract** In our study, a structure-based virtual screening study was conducted to identify potent ITK inhibitors, as ITK is considered to play an important role in the treatment of inflammatory diseases. We developed a structure-based pharmacophore model using the crystal structure (PDB ID: 3MJ2) of ITK complexed with BMS-50944. The most predictive model, SB-Hypo1, consisted of six features: three hydrogen-bond acceptors (HBA), one hydrogen-bond donor (HBD), one ring aromatic (RA), and one hydrophobic (HY). The statistical significance of SB-Hypo1 was validated using wide range of test set molecules and a decoy set. The resulting well-validated model could then be confidently used as a 3D query to screen for drug-like molecules in a database, in order to retrieve new chemical scaffolds that may be potent ITK inhibitors. The hits retrieved from this search were filtered based on the maximum fit

value, drug-likeness, and ADMET properties, and the hits that were retained were used in a molecular docking study to find the binding mode and molecular interactions with crucial residues at the active site of the protein. These hits were then fed into a molecular dynamics simulation to study the flexibility of the activation loop of ITK upon ligand binding. This combination of methodologies is a valuable tool for identifying structurally diverse molecules with desired biological activities, and for designing new classes of selective ITK inhibitors.

**Keywords** ITK · Inflammatory · LigandScout · Virtual screening · Molecular docking

## Abbreviation

ITK	Interleukin-2-inducible T-cell kinase
SB_Hypo1	Structure-based hypothesis
PTK	Protein tyrosine kinase
PH	Pleckstrin homology
SH3	Src homology3
SH2	Src homology2
DS	Discovery Studio v.2.5
HBA	Hydrogen-bond acceptor
HBD	Hydrogen-bond donor
HY	Hydrophobic
R	Ring aromatic
EF	Enrichment factor
GH	Goodness of hit
ADMET	Absorption, distribution, metabolism, excretion, and toxicity
BBB	Blood–brain barrier
MD	Molecular dynamics
RMSD	Root mean square deviation

**Electronic supplementary material** The online version of this article (doi:10.1007/s00894-012-1536-7) contains supplementary material, which is available to authorized users.

C. Meganathan · S. Sakkiah · Y. Lee · 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

J. V. Narayanan  
Department of Oral and Maxillofacial Surgery,  
Ragas Dental College and Hospital,  
Uthandi,  
Chennai, Tamilnadu 600019, India

## Introduction

Protein tyrosine kinases (PTKs) are vital components of many signal transduction pathways in multicellular organisms. They can be subdivided into two families: receptor and nonreceptor (cytoplasmic) kinases. Src and eight related molecules formed the largest family of cytoplasmic PTKs, while the second largest is the Tec family, which consists of five mammalian members: BTK (Bruton agammaglobulinemia tyrosine kinase), BMX (bone marrow kinase gene on the X chromosome), ITK (interleukin-2-inducible T-cell kinase), TEC (tyrosine kinase expressed in hepatocellular carcinoma), and TXK (T and X cell-expressed kinase) [1]. The Tec family plays an essential role in T-cell activation as well as signaling through antigen receptors such as T-cell and B-cell receptors [2]. Three members of the Tec family—ITK, TXK, and TEC—are activated downstream of antigen receptor engagement in T cells, and transmit signals to downstream effectors, including PLC- $\gamma$  [3]. BTK acts independently of T-cell signaling, and is essential for B-cell development and activation. The biological role of the final member of this family, BMX, has not yet been identified.

ITK is an important member of the Tec family because of its key role in the regulation of various signaling pathways in immune cells, such as actin reorganization, PLC- $\gamma$  activation, calcium mobilization, and NFAT activation in T cells [4]. ITK also plays a vital role in the secretion of Th2 cytokines, such as IL-4, IL-5, and IL-13, and the development of an effective Th2 response during allergic asthma or infections by parasitic worms. Immunological deficiencies were seen in ITK knockout mice under normal conditions, which emphasizes the importance of ITK in T-cell development and function. However, studies of ITK knockout mice in an animal model of allergic asthma have indicated that ITK is important as not only a mediator of the secretion of specific cytokines by Th2 cells but also in the release of cytokines and chemokines from mast cells [5]. Together, these findings have illuminated the vital roles of ITK in various signaling pathways in immune cells. Consequently, ITK is considered a promising drug target for the treatment of numerous inflammatory diseases, including allergies, allergic asthma, and atopic dermatitis. Recently, it was also shown that the inhibition of ITK blocks HIV infection by affecting multiple HIV replication steps [4].

The structure of ITK is very similar those of other Tec family kinase members. It contains a pleckstrin homology (PH) domain at the N-terminus, which allows it to reversibly recruit the membrane and TEC-homology domain [4]. The TEC-homology domain is characterized by the BTK motif, which is found only in the Tec kinases [6–8], and a proline-rich region. ITK also has Src homology3 (SH3), Src homology2 (SH2), and kinase domains [9, 10]. The SH3 domain interacts with the proline-rich regions of other proteins, the

SH2 domain interacts with tyrosine phosphorylated proteins, and the kinase domain has a catalytic function and carries out the tyrosine phosphorylation of substrates. In the case of ITK, any of these domains could potentially be targeted to prevent its activation. However, the PH, SH3, and SH2 domains may be involved in the regulation of ITK as well as the regulation of signaling pathways independent of the kinase domain. Thus, more work to clearly understand this process is needed before this approach is attempted [11]. Hence, in the work described in the present paper, we targeted the kinase domain of ITK on order to prevent the initiation of kinase activity.

Several research works on the inhibition of ITK signaling activity have already been performed, and many of other works are currently underway. While many small molecules have already been tested for ITK activity, none of them have shown selective activity towards ITK [11]. Hence, our aim was to find potent selective small-molecule ITK inhibitors within the Tec family, and inhibitors of other kinases. Thus, in our study, we derived the basic features needed to inhibit ITK activity, which are therefore useful when designing novel potential inhibitors for use as anti-inflammatory therapeutics. Structure-based pharmacophore model approaches were utilized to identify the critical chemical features required to inhibit ITK. The structure-based pharmacophore hypothesis developed here was then thoroughly validated by utilizing test and decoy sets to prove its predictive ability and statistical significance. This well-validated model was used as a query to screen a chemical database for novel leads. A molecular docking study was then performed to identify the appropriate orientation and flexibility of the lead molecule at the protein active site. The final drug-like hits were fed into a molecular dynamics (MD) simulation, as the crystal structures of human ITK complexed with different inhibitors (PDB ID: 1SNU, 1SM2, 3MIY, 3MJ1, 3MJ2, 3QGW, and 3QGY) have already given us a clear view of the binding mode of ITK inhibitors [3, 4]. Among these seven crystal structures, only one (PDB ID:3MJ2; ITC complexed with BMS-509744) was fully resolved, including the activation loop, and this activation loop formed a short alpha helix to block the protein substrate binding site [4]. Thus, we carried out MD simulations using GROMACS v.4.0 to assess the flexibility of the critical amino acids present at the protein active site with and without the inhibitor bound to it. Also, the interactions between the inhibitor and the side chains of the critical amino acids F435, C442, and S499 were used to determine the selectivity of the ITK inhibitors.

Studying the crystal structures of ITK–inhibitor complexes proved to be a useful way to identify how to block the substrate binding site and prevent the initiation of kinase activity. We utilized this valuable information in structure-based database screening, molecular docking, and MD simulation studies to select novel potential leads in ITK inhibitor design.

## Materials and methods

### Generation of the structure-based pharmacophore model

Selecting a crystal structure to use to generate a structure-based pharmacophore model is a challenging task. Seven well-resolved X-ray structures of ITK complexed with various inhibitors are available. For each of these seven structures, the binding mode and molecular interactions at the protein's active site are clearly defined. For example, the X-ray structure (PDB ID: 3MIY) of ITK complexed with sunitinib, a broad-spectrum tyrosine kinase inhibitor [4], shows the binding mode and molecular interactions at the protein's active site in detail. Sunitinib was observed to form a hydrogen-bond network with the critical amino acids M438 and E436, and to participate in an edge-to-face interaction with the gatekeeper residue F435. However, helix C in the ITK–sunitinib structure adopts a different conformation than in previously reported structures, and this conformation is considered the “active conformation” [12–14]. The other X-ray structures (PDB ID: 1SM2, 1SNU, 3MJ1, 3QGW, and 3QGY), of ITK complexed with staurosporine, R05191614, and other inhibitors, show hydrogen-bond networks and binding modes that are quite similar to those of sunitinib. However, none of these five crystal structures have an activation loop in the alpha-helix conformation or incorporate the activating trans-phosphorylation site Y512. In general, the regulatory mechanisms of various protein kinases have been described based on their activation loops. In the inactive state, the nonphosphorylated activation loop binds and blocks access to the protein, whereas, in the active state, phosphorylation of the activation loop releases this intramolecular interaction and activates the catalytic activity [12–14]. Note that ITK may not exhibit this regulatory mechanism, since its apo (PDB ID: 1SNX), unphosphorylated (PDB ID: 1SNU), and phosphorylated (PDB ID: 1SM2) crystal structures do not have the activation loop. However, the crystal structure of its complex with BMS-509744 (PDB ID: 3MJ2) can exhibit an autoinhibitory confirmation, and it is also the only structure that is fully resolved and has an activation loop. In addition, BMS-509744 is the only inhibitor that shows van der Waals contact with the side chains of the active-site residues F435, C442, and S499, which are crucial to ITK selectivity. Hence, we chose the crystal structure of ITK complexed with BMS-509744 to generate the structure-based pharmacophore model. The crystal structure of 3MJ2 was obtained directly from the Brookhaven Protein Data Bank (PDB). Water-mediated hydrogen bonds are reported to play a crucial role in the stabilization of the ligand at the active site. Therefore, all of the water molecules in the crystal structure were retained. The structure-based pharmacophore model was generated using the LigandScout software package

[15], which uses an algorithm that interprets ligand–receptor interactions such as hydrogen bonds, charge transfer, and hydrophobic regions of their macromolecular environment from PDB files, allowing the automated construction of the pharmacophore model. The generated pharmacophore model included the excluded volume spheres, which represent the inaccessible areas on any potential ligand [16]. Discovery Studio v.2.5 (DS) was used to convert .hypoedit to .chm files, since this is a suitable format for performing the virtual screening of various commercially available databases in order to retrieve novel scaffolds. Several complementary pharmacophore features were generated to describe the binding site. These features were analyzed (except for the excluded volumes) based on the interaction pattern of the ligand at the active site, and the most representative features were selected for inclusion in the best pharmacophore model.

### Validation of the pharmacophore model and virtual screening

Test and decoy sets were employed to validate the predictive ability and statistical significance of the generated model, respectively. The test set contained 30 structurally diverse molecules with a wide range of activities ( $IC_{50}$ ), between 0.003  $\mu\text{M}$  and 50  $\mu\text{M}$ , which were collected from various literature sources [17–23]. All of the test set molecules were grouped into three categories: highly active (+++,  $IC_{50} < 1 \mu\text{M}$ ), moderately active (++,  $1 \leq IC_{50} < 10 \mu\text{M}$ ), and inactive (+,  $IC_{50} \geq 10 \mu\text{M}$ ). The 2D structures of the selected compounds were built using MDL ISIS Draw v.2.4. In addition, these 2D structures were converted to their corresponding 3D forms by importing them into DS. Optimization was carried out for all compounds by minimizing the energy to the closest local minimum with the CHARMM-like force field [24]. The conformational analysis for each compound, which investigated its flexibility, was performed using a poling algorithm [25–27] and CHARMM force field parameters. Two methods are included in DS for conformational analysis: fast and best-quality analysis. Both methods were carried out using a Monte Carlo-like algorithm together with poling [25–27], which considerably reduced the probability that very similar conformers would appear through the use of a penalty function. In this work, a maximum of 255 conformations were allowed for each compound by using the best conformation generation method with a cut-off of 20  $\text{kcal mol}^{-1}$  energy above the global energy minimum to ensure maximum coverage of the conformational space. Conformational variations yielding similar conformers were explicitly suppressed by poling and all are equally treated. The second validation method utilized a decoy set consisting of 1200 molecules, including known inhibitors of ITK (30) and unknown (1170) compounds. The generated hypothesis was employed to screen this decoy set using the

ligand pharmacophore mapping protocol in DS to check how well the hypothesis distinguished the active from the inactive molecules. Parameters such as false positives, false negatives, the enrichment factor (EF), and the goodness of hit (GH) were calculated to determine the robustness of the hypothesis.

Database screening was used to retrieve new molecules from databases [28] for biological testing. The main purpose of virtual screening is to find novel scaffolds that can inhibit the activities of various targets, as well as to find potential leads that are suitable for further development. The ligand pharmacophore mapping protocol in DS was employed for database screening. In this, there are two conformation options: best/flexible and fast/flexible. In our study, to attain the best results, we used the best/flexible conformation option during database screening. The well-validated six-feature structure-based pharmacophore model was then used as 3D query to retrieve the hits. Two databases, namely Maybridge and Chembridge, containing 0.1 million structurally diverse small molecules, were screened. In addition, the geometric fit of each database compound was calculated based on how well its chemical substructure fitted with the pharmacophoric feature location constraints and their distances from the centers of the features. Hit compounds that scored a geometric fit value of  $\geq 3$  (pharmacophore fit value) were selected and checked for drug-like properties using methods such as Lipinski's rule of five and ADMET (absorption, distribution, metabolism, excretion, and toxicity) descriptors. Hit compounds that passed through the drug-like property filters (Lipinski's rule of five and ADMET) were taken into account in our molecular docking study.

### Molecular docking

Molecular docking was performed to refine the hit molecules derived from database screening in order to determine the affinity between the ligand and the critical amino acids present at the active site of the protein. Several algorithms that perform molecular docking in order to find the ligand–protein affinity (using the scoring function) and the interactions between the ligand and critical amino acids at the active site have been reported. In this work, GOLD (genetic optimization for ligand docking) v.4.1 from the Cambridge Crystallographic Data Center, UK [29], was employed. This uses a genetic algorithm to dock flexible ligands into protein-binding sites in order to explore the full range of ligand conformation flexibility while allowing partial protein flexibility. The X-ray crystal structure of ITK (PDB ID: 3 MJ2) that was used in the molecular docking study was obtained directly from the Protein Data Bank (<http://www.rcsb.org>). All of the water molecules associated

with the protein were also included in the docking study. Hydrogen atoms were also added using GOLD. The binding site on the protein was defined using all of the atoms within 12 Å of it based on the co-crystallized ligand included 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. The default fitness function (VDW 4.0, H-bonding 2.5) and evolutionary parameters were used in the GOLD docking experiments: population size 100; selection pressure 1.1; operation 100,000; islands 5; niche size 2; migration 10; crossover 95. The GOLD fitness score, the binding mode, and molecular interactions with catalytically important residues at the protein's active site were used to sort the molecules.

### Molecular dynamics simulation

We performed MD simulations (using GROMACS v.4.0 [30, 31] with the GROMOS93a1 force field) of the apo form of ITK, the complex of ITK with BMS-509744, and the complexes of ITK with three drug-like leads found in the database. The GROMACS molecular topology files for BMS-509744 and database hits were generated using the PRODRG server (<http://davapc1.bioch.dundee.ac.uk/programs/prodrg>) [32]. Hydrogen atoms were added, and the protonation states of the ionizable groups on the protein were set to pH 7.0. Each starting structure was placed in a cubic box with 1 nm as the minimum distance between the protein and the edge of the box [33]. The simple point charge (SPC) water model was used. In order to neutralize the system, ten Na<sup>+</sup> counterions were added by replacing water molecules. Long-range electrostatics were treated using the particle mesh Ewald method [34]. Energy minimization was performed by applying the steepest descent algorithm to the initial structure of the protein to remove bad contacts until energy convergence reached 2,000 kJ mol<sup>-1</sup>. Each component of the system was subjected to equilibration at 300 K and under normal pressure (1 bar) for 100 ps, applying positional restraints to heavy atoms. The LINCS [35] and SETTLE [36] algorithms were used to constrain the bond lengths within the protein and the geometries of the water molecules, respectively. In our study, for all of the systems, we carried out 5 ns production run in the isothermal-isobaric (NPT) ensemble under periodic boundary conditions. The Lennard–Jones and electrostatic interaction cut-offs were set 0.9 nm and 1.4 nm, respectively. The time step of the simulation was 2 fs without any positional restraint, and conformations were saved during every analysis. All analyses were performed using PyMOL, DS, and simulation trajectory analysis using tools included in the GROMACS package, including the DSSP program [37].

## Results and discussion

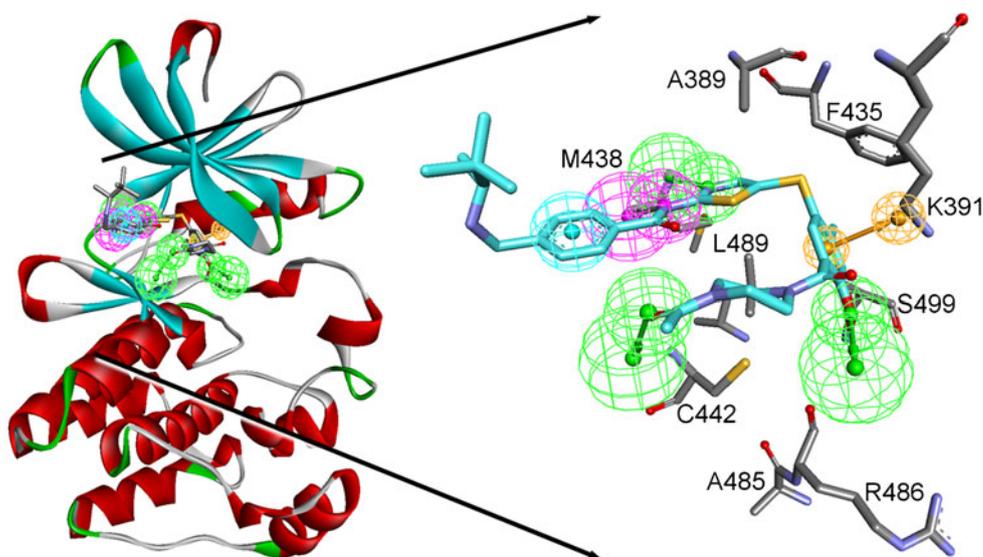
### Structure-based pharmacophore generation

ITK is the most effective drug target for the treatment of Th2-mediated inflammatory diseases. Recent reports also suggest that selective blockage of ITK affects HIV replication by interrupting T-cell activation [38]. So far, 16 different chemical structure classes of ITK inhibitors have been discovered, but none of these chemical entities have been reported in clinical trials [38]. Among these 16 different chemical structures, only BMS-509744 has been shown to exhibit an activation loop when binding with the protein. In addition, only the X-ray crystal structure of ITK complexed with BMS-509744 (PDB ID: 3 MJ2) has shown van der Waals interactions with residues that are important to ITK selectivity, such as F435, C442, and S499. Thus, we used this structural data as the starting point to generate a structure-based pharmacophore for a BMS-509744 analog that could aid in the identification of new chemical scaffolds for use as potential ITK inhibitors. The LigandScout software package was used to carry out an extensive investigation of the chemical features that are required for the ITK–BMS-509744 interaction. The generated pharmacophore contained nine features: four hydrogen bond acceptors (HBA), three hydrophobic (HY) features, one hydrogen bond donor (HBD), and one ring aromatic (RA)—along with eighteen excluded volumes—all of which determine the interaction between protein and ligand (see Fig. S1 of the “Electronic supplementary material,” ESM).

The generated pharmacophore features were then analyzed by mapping them to the active-site residues of the protein, and the most representative features were included in the pharmacophore. The final pharmacophore (SB\_Hypo1) consisted of six features: three HBAs, one HBD, one HY, and one RA; the

excluded volumes were deleted, since they are areas that are inaccessible to any potential ligand (Fig. 1). HBD and HBA features from the aminothiazole core of the ligand form essential hydrogen-bond interactions with the crucial amino acid M438. The other two HBA features correspond to the methoxy group and the terminal carbonyl-linked acetyl piperazine moiety of the ligand. In addition, these two features are situated near to the residues C442 and S499, which are influential in ITK selectivity; hence, these residues can form water-mediated hydrogen bonds with the ligand, and such interactions stabilize the ligand in the active site. The HY feature is positioned on the aromatic ring of the phenyl methylamine of the ligand. The RA feature corresponds to the phenylsulfonyl moiety of the ligand and enables aromatic stacking between the benzyl ring of the ligand and F435. Further, comparison of this automated SB\_Hypo1 (Fig. S2a of the ESM) with a previously developed ligand-based pharmacophore model [39] showed 1.5 Å of RMS displacement, and highlighted some differences in chemical features (Fig. S2b of the ESM). In particular, the SB\_Hypo1 model has additional HBA and HBD features compared to the ligand-based pharmacophore model. This mismatch occurs because the SB\_Hypo1 model describes the predominant interactions of the BMS-509744 ligand with the protein, whereas the ligand-based pharmacophore model describes the interactions that are commonly seen for several highly active compounds belonging to different chemical classes of inhibitors. Among the ITK competitive inhibitors, only BMS-509744 has been shown to exhibit an activation loop while binding with the protein, and to trigger the autoinhibitory mechanism of the protein. This pharmacophore model has a clearly defined active site and ligand binding, making it more effective in database queries that are performed to search for compounds similar to known ITK competitive inhibitors.

**Fig. 1** SB\_Hypo1 overlaid onto the active site of ITK



## Validation

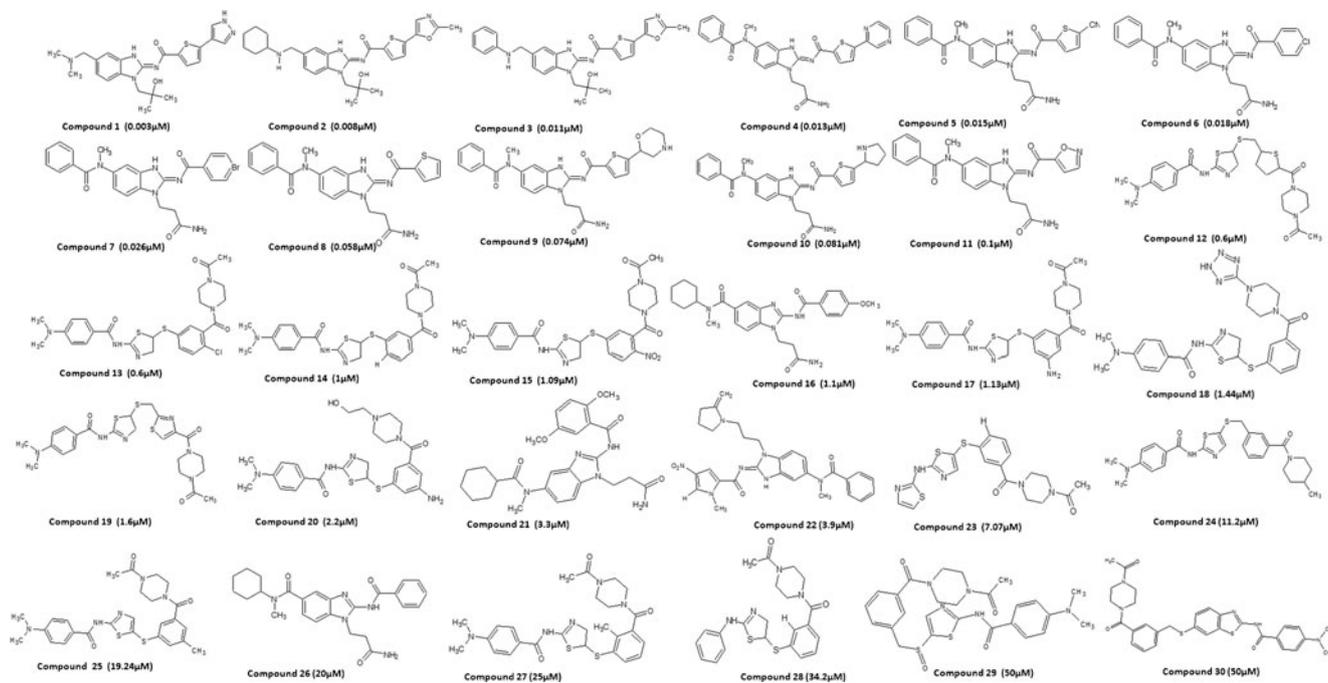
### Test set

The predictive power of the generated pharmacophore model SB\_Hypo1 was tested using a set of 30 ITK inhibitors (Fig. 2), which included the most active, moderately active, and inactive molecules; conformation studies were done as described earlier. When using a structure-based pharmacophore model, molecules can be separated into active, moderately active, and inactive using the fit value based on the geometric scoring function. On the other hand, a ligand-based pharmacophore model uses estimated activity values, since it is derived from a set of active compounds. Thus, we classified the compounds in the test set as follows: compounds with fit values of  $>3$  were considered the most active compounds; those with fit values of between 2 and 3 were considered moderately active, and those with fit values of  $<2$  were considered inactive compounds. Hypo1 was used to screen the test set molecules using the ligand pharmacophore mapping protocol in DS, with the maximum omitted features value set to zero. Of these, 13 of the 13 most active compounds, 4 of the 10 moderately active compounds, and 7 of the 7 inactive compounds were predicted correctly (Table 1). The activities of six moderately active compounds were underestimated (they were considered to be inactive compounds). This discrepancy only occurred for the moderately active compounds,

and may have happened because the value of maximum omitted features was set to zero. These results clearly demonstrated that the features present in SB\_Hypo1 are critical to the inhibition of ITK. The alignments of Hypo1 with the most active compound, 1 ( $IC_{50}$ : 0.003  $\mu$ M), and with the least active compound, 30 ( $IC_{50}$ : 50  $\mu$ M), in the test set are shown in Fig. 3. All of the features of Hypo1 mapped perfectly to compound 1, while compound 30 could not map to one of the HBA features of Hypo1. This again showed that small-molecule inhibitors of ITK are clearly highly dependent on the interaction between ITK and its ligand. This further verifies the predictive ability of the pharmacophore model Hypo1.

### Decoy set

Hypo1 was once again validated by using it to search for the active molecules in a large decoy set of active and inactive molecules. The database we used contained 1200 molecules, including 30 known inhibitors of ITK. This “spiked” database was screened by Hypo1 using the ligand pharmacophore mapping protocol with a maximum omitted features value of zero. The calculated statistical parameters are summarized in Table S1 of the ESM. As seen in Table S1, Hypo1 successfully screened the decoy set and retrieved 36 molecules. Among these 36 molecules, 30 molecules (83.33 %) were potent inhibitors of ITK, while 6 molecules were actually inactive (i.e., false positives). The EF and GH score



**Fig. 2** 2D structures of the test set molecules that were used to validate SB\_Hypo1

**Table 1** Experimental and predicted activities of the test set compounds, as calculated utilizing SB\_Hypo1

Compound no.	Experimental IC <sub>50</sub> (μM)	Fit value <sup>a</sup>	Experimental scale <sup>b</sup>	Predicted scale <sup>b</sup>
1	0.003	4.25	+++	+++
2	0.008	4.18	+++	+++
3	0.011	4.09	+++	+++
4	0.013	3.87	+++	+++
5	0.015	3.85	+++	+++
6	0.018	3.85	+++	+++
7	0.026	3.85	+++	+++
8	0.058	3.83	+++	+++
9	0.074	3.78	+++	+++
10	0.081	3.76	+++	+++
11	0.1	3.73	+++	+++
12	0.6	3.21	+++	+++
13	0.6	3.14	+++	+++
14	1	2.99	++	++
15	1.09	2.84	++	++
16	1.10	2.38	++	++
17	1.13	2.31	++	++
18	1.44	1.93	+	+
19	1.60	1.78	+	+
20	2.20	1.68	+	+
21	3.30	1.64	+	+
22	3.90	1.63	+	+
23	7.0	1.60	+	+
24	11.2	1.59	+	+
25	19.2	1.57	+	+
26	20	1.57	+	+
27	25	1.35	+	+
28	34.2	1.24	+	+
29	50	1.23	+	+
30	50	0.94	+	+

<sup>a</sup> Fit value indicates how well the features in the pharmacophore overlap with the chemical features in the molecule. Fit = weight × [max (0, 1, SSE)] where SSE = (D/T)<sup>2</sup>, 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)

<sup>b</sup> Activity scale: IC<sub>50</sub> < 1 μM = +++ (highly active); 1 ≤ IC<sub>50</sub> < 10 μM = ++ (moderately active); IC<sub>50</sub> ≥ 10 μM = + (inactive)

values for Hypo1 were calculated using the following formulae:

$$EF = \left( \frac{H_a}{H_t} \right) \div \left( \frac{A}{D} \right) \quad (1)$$

$$GH = \left( \frac{H_a(3A + H_t)}{4H_tA} \right) \left( 1 - \frac{H_t - H_a}{D - A} \right) \quad (2)$$

Here,  $H_t$  = 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. GH scoring was used to assess the quality of the pharmacophore. GH ranges from 0, indicating a null model, to 1, indicating the ideal model [40]; a score value of  $\geq 0.60$  indicates a very good model [41]. In our case, Hypo1 gained a GH value of 0.87, which indicates that it is highly able to identify false positives, and can even differentiate between active and inactive ITK inhibitors that are very similar structurally. Hypo1 also has an EF score of 33.32, which indicates that it is 33 times more likely to pick an active compound from the database than an inactive one [42]. According to both validation methods, Hypo1 is very able to distinguish active from inactive molecules, so it was then used in database screening to retrieve novel potential ITK inhibitors.

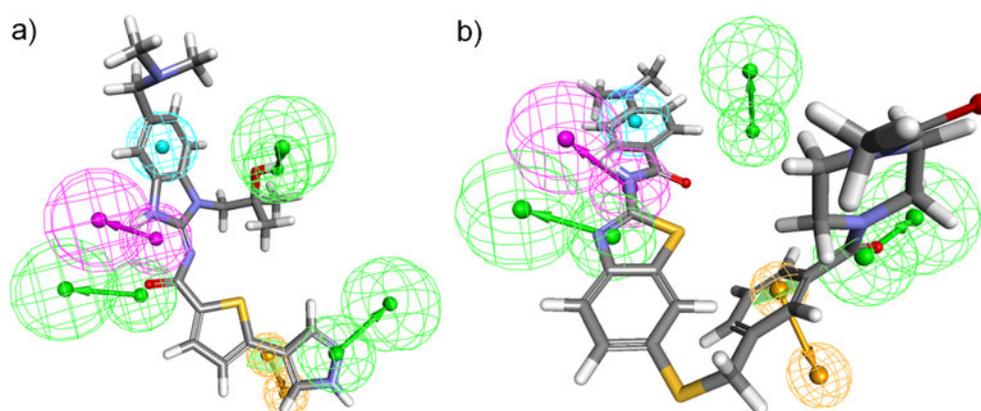
#### Database screening

In order to identify novel potential ITK inhibitors, the well-validated pharmacophore model Hypo1 was used as a 3D query to screen two databases: Maybridge (60,000 molecules) and Chembridge (50,000 molecules). The screening process was carried out using the ligand pharmacophore mapping protocol in DS, with the best/flexible algorithm and a maximum omitted features value of 1. In the initial screening, the compounds which mapped well to five or six features of SB\_Hypo1 were considered hits. These initial hit molecules were then screened again (based on the maximum fit value) and a third time (through a drug-like filter incorporating Lipinski's rule of five and ADMET descriptors). Finally, a total of 24 potential candidates were obtained after all three screening processes, and these molecules were analyzed in molecular docking studies.

#### Molecular docking

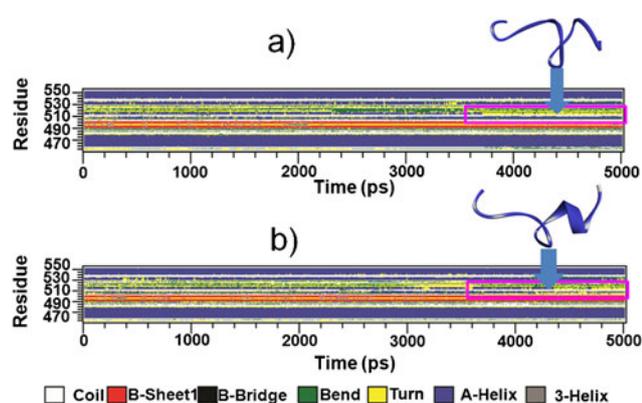
A molecular docking study of the candidate compounds obtained from the database screening process was performed to identify their affinities for the active site of the protein. GOLD v4.1 was employed as a docking tool to analyze the binding modes and molecular interactions of the candidate compounds. Details about the receptor and its preparation for the docking study were described earlier. We tested the reliability of the docking method as a way to predict the bioactive conformations and GOLD fitness scores, which distinguish the ITK-active from the ITK-inactive molecules. Thus we evaluated by co-crystal structure (BMS-509744), which was docked into the binding sites of ITK. The re-docked conformation of BMS-509744 that was predicted using GOLD is shown in Fig. S3a. The root mean-square deviation between the actual

**Fig. 3** Mapping the pharmacophore to test set molecules: **a** compound 1; **b** compound 30

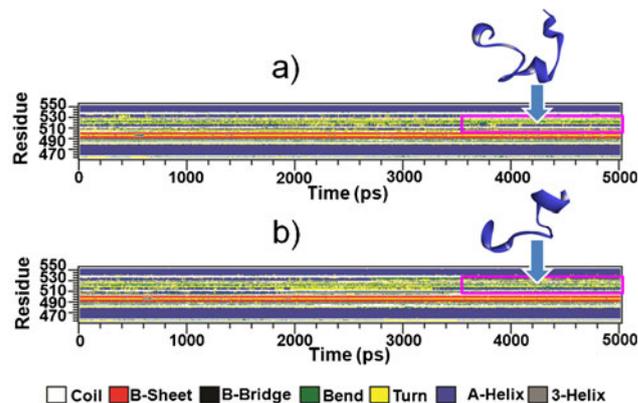


and the predicted conformations was 0.98 Å, suggesting that GOLD reliably reproduces the experimentally observed binding modes for ITK inhibitors. Moreover, the GOLD fitness score for BMS-509744 was 95, and it indicated very strong hydrogen-bond interactions with M438 and C442, as well as a water-mediated hydrogen bond with the water molecule X158 (Fig. S3b). Hence, we used BMS-509744 as our reference molecule for the candidate compounds we obtained from database screening. Molecules that gave GOLD fitness scores of >90 were selected as lead compounds, and their binding modes and molecular interactions were checked. The final hit compounds were docked into the protein's active site using the same protocol utilized for the reference molecule. Nine out of the 24 candidate molecules showed GOLD fitness scores of >90, and their molecular interactions are summarized in Table S2 of the ESM. As seen in Table S2, three compounds—CD\_01889 and JFD\_01598 (Maybridge) as well as compound 11513 (Chembridge)—yielded GOLD fitness scores of 94, 93, and 94, respectively. Also, these three compounds showed very strong hydrogen-bond interactions with the critical residue M438 in the hinge region and a water-mediated

hydrogen bond with a water molecule (X158). Further, these compounds exhibited hydrogen bonds with residues C442 and S499, whereas our reference compound (BMS-509744) did not show a hydrogen-bond interaction with S499. According to previous studies, F435, C442, and S499 are the residues that have the most influence on ITK selectivity, which obviously makes us more confident that these three compounds do indeed show ITK selectivity. Figures S4 and S5 of the ESM show these reference and hit compounds overlaid onto the pharmacophore, as well as their molecular interactions with the crucial amino acids and their binding modes. The results of all of the above investigations of fit value, drug-likeness, pharmacophore overlay, binding mode, and molecular interactions with residues that are crucial to ITK selectivity suggest that the retrieved hits could represent good leads in the design of novel inhibitors of ITK. However, to further confirm the suitability of these compounds as leads, we investigated the flexibility of the residues in the protein upon the binding of the inhibitor using molecular dynamics simulation.



**Fig. 4** Evolution of the secondary structure of the activation loop of ITK, computed using the DSSP package, for ITK complexed with **a** apo and **b** BMS-509744. The *x*-axis represents the time elapsed during the simulation, while the *y*-axis represents the numerical label of the residue (450–550). Each *arrow* indicates the simulation time at which the shown snapshot of the structure was taken



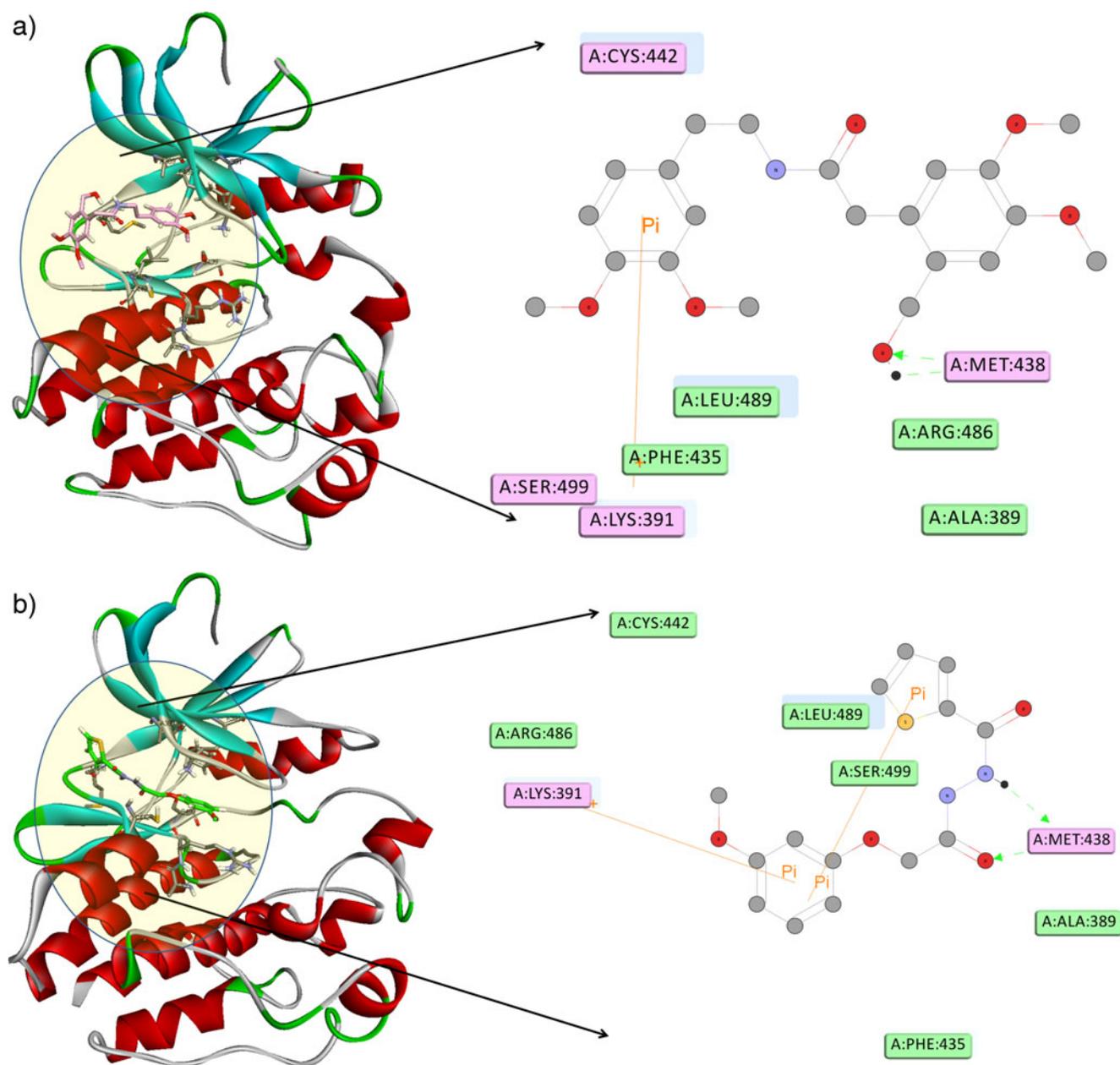
**Fig. 5** Evolution of the secondary structure of the activation loop of ITK, computed using the DSSP package, for ITK complexed with **a** CD\_01889 and **b** compound 11513. The *x*-axis represents the time elapsed during the simulation, while the *y*-axis represents the numerical label of the residue (450–550). Each *arrow* indicates the simulation time at which the shown snapshot of the structure was taken

## Molecular dynamics simulation

The activities of several kinases are known to be regulated by a region known as the activation loop. When the kinase is inactive, the (nonphosphorylated) activation loop binds and blocks access to the protein substrate binding site. However, in the activated state, phosphorylation of the activation loop releases these intramolecular interactions, which in turn “switches on” the catalytic activity of the kinase. That said, ITK may not exhibit this regulatory mechanism, as the electron density of the activation loop was not found in the crystal structures of the complexes of

ITK with Apo (ISNX), staurosporine (1SNU and 1SM2), sunitinib (3MIY), R05191614 (3 MJ1), 3QGW, and 3QGY. Only the crystal structure of the complex of ITK with BMS-509744 exhibits an activation loop, and shows evidence of an autoinhibitory mechanism.

In our study, we evaluated our candidate leads using MD simulation. Five 5 ns MD simulations were successfully performed and analyzed. Details of the MD simulation environments and the size of each systems are summarized in Table S3 of the ESM. We selected the representative structure of each system, which was the closest conformation to the average structure during the last 2 ns of simulation, in order



**Fig. 6** Molecular interactions of the final lead compounds with the active site of ITK observed during MD simulation: **a** CD\_01889, **b** compound 11513

to compare the protein structures. The calculated average root-mean-square deviations (RMSDs) of  $C_{\alpha}$  for the systems (apo, BMS-509744, CD\_01889, JFD\_01598, compound 11513) during the last 2 ns of simulation were 4257, 4062, 4186, 4188, and 4251, respectively (Fig. S6 of the ESM). There were no significant structural differences compared to the N-terminus and the C-terminus of the apo form (Fig. S7a of the ESM). However, we observed a major change in the activation loop (residues 502–521) when comparing entire protein structures. Upon binding to BMS-509744, an alpha helix formed in the activation loop (Fig. S7b of the ESM); this alpha helix was not present in the apo form. This helix binds and blocks access to the substrate [4], hence this helix conformation strongly supports unphosphorylated form of kinases such as autoinhibitory mechanism. Interestingly, two of the candidate compounds that were retrieved during the database screening process (CD\_01889 and compound 11513) also showed this conformational change in the activation loop (Fig. S7c and d of the ESM). To prove this, we evaluated the secondary structure of each compound using the DSSP program within GRO-MACS. Secondary structure prediction clearly demonstrated the formation of an alpha helix in each system during the last 2 ns of simulation (Figs. 4 and 5). In addition, molecular interactions of the lead compounds were observed during the simulation, and these are depicted in Fig. 6. As shown in the Fig. 6, both of the lead compounds mentioned above (CD\_01889 and compound 11513) present hydrogen-bond interactions with the critical amino acid M438 and hydrophobic interactions with ITK-selective residues F435, C442, and S499. However, an alpha-helix conformation of the activation loop was not observed for the candidate compound JFD\_01598 (Fig. S8 of the ESM). Therefore, although this compound appears to show all of the critical interactions with the active site of the protein needed for ITK activity, it may not be an ITK inhibitor. On the other hand, according to all of the validation data, the other two lead candidates (CD\_01889 and compound 11513) showed similar interactions with the active site to those seen in the complex of ITK with BMS-509744. Hence, we suggest that compounds may represent good leads in the design of novel potential inhibitors of ITK.

## Conclusions

In this study, we generated a structure-based pharmacophore model in order to identify the critical chemical features of and discover novel potent inhibitors of ITK, as ITK has been shown to be promising therapeutic target for Th2-mediated inflammatory diseases as well as HIV infection (ITK influences multiple HIV replication steps). We used a well-resolved and fully solvated crystal structure (PDB ID: 3 MJ2) to generate the structure-based pharmacophore model, which initially had 9 features and 18 excluded volumes. These

features were further filtered by superimposing the model on the active site of the protein, such that only the most representative features were included in the pharmacophore model. The final model consisted of six features: three HBAs, one HBD, one HY, and one RA, and this model was validated by using it to screen test and decoy sets in order to determine its predictive ability. The resulting well-validated model was used as a 3D query to screen two large databases for candidate ITK-inhibiting compounds. The hits retrieved from this screening process were then further filtered based on their maximum fit values with respect to the pharmacophore, and using drug-like filters such as ADMET and Lipinski's rule of five. Next, the hits that passed through these filters were investigated in a molecular docking study to check their affinities for crucial active-site residues. Molecular dynamics simulations were also carried out to evaluate the flexibility of the activation loop in the ITK protein upon ligand binding. Finally, two compounds were identified as potential leads based on all of the above validation, and these compounds represent useful potential leads in the design of novel ITK inhibitors. The pharmacophore model developed here could also be used as a query to search other databases for more novel candidate ITK inhibitors.

**Acknowledgments** This research was supported by the Basic Science Research Program (2009-0073267), the Pioneer Research Center Program (2009-0081539), and the 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 the Republic of Korea. This work was also supported by the Next-Generation BioGreen21 Program (PJ008038) from the Rural Development Administration (RDA) of the Republic of Korea.

## References

1. Smith CIE, Islam TC, Mattsson PT, Mohamed AJ, Nore BF, Vihinen M (2001) The Tec family of cytoplasmic tyrosine kinases: mammalian Btk, Bmx, Itk, Tec, Txk and homologs in other species. *BioEssays* 23(5):436–446. doi:10.1002/bies.1062
2. Miller AT, Berg LJ (2002) New insights into the regulation and functions of Tec family tyrosine kinases in the immune system. *Curr Opin Immunol* 3:331–340
3. Brown K, Long JM, Vial SCM, Dedi N, Dunster NJ, Renwick SB, Tanner AJ, Frantz JD, Fleming MA, Cheetham GMT (2004) Crystal structures of interleukin-2 tyrosine kinase and their implications for the design of selective inhibitors. *J Biol Chem* 279(18):18727–18732. doi:10.1074/jbc.M400031200
4. Kutach AK, Villaseñor AG, Lam D, Belunis C, Janson C, Lok S, Hong LN, Liu CM, Deval J, Novak TJ, Barnett JW, Chu W, Shaw D, Kuglstatter A (2010) Crystal structures of IL-2-inducible T cell kinase complexed with inhibitors: insights into rational drug design and activity regulation. *Chem Biol Drug Des* 76(2):154–163. doi:10.1111/j.1747-0285.2010.00993.x
5. von Bonin A, Rausch A, Mengel A, Hitchcock M, Krüger M, von Ahsen O, Merz C, Röse L, Stock C, Martin SF, Leder G, Döcke W-D, Asadullah K, Zügel U (2011) Inhibition of the IL-2-inducible tyrosine kinase (Itk) activity: a new concept for the therapy of

- inflammatory skin diseases. *Exp Dermatol* 20(1):41–47. doi:10.1111/j.1600-0625.2010.01198.x
6. Andreotti AH, Bunnell SC, Feng S, Berg LJ, Schreiber SL (1997) Regulatory intramolecular association in a tyrosine kinase of the Tec family. *Nature* 385(6611):93–97. doi:10.1038/385093a0
  7. Bunnell SC, Diehn M, Yaffe MB, Findell PR, Cantley LC, Berg LJ (2000) Biochemical interactions integrating Itk with the T cell receptor-initiated signaling cascade. *J Biol Chem* 275(3):2219–2230. doi:10.1074/jbc.275.3.2219
  8. Chamorro M, Czar M, Debnath J, Cheng G, Lenardo M, Varmus H, Schwartzberg P (2001) Requirements for activation and RAFT localization of the T-lymphocyte kinase Rlk/Txk. *BMC Immunol* 2(1):3. doi:10.1186/1471-2172-2-3
  9. Brazin KN, Mallis RJ, Fulton DB, Andreotti AH (2002) Regulation of the tyrosine kinase Itk by the peptidyl-prolyl isomerase cyclophilin A. *Proc Natl Acad Sci USA* 99(4):1899–1904. doi:10.1073/pnas.042529199
  10. Berg LJ, Finkelstein LD, Lucas JA, Schwartzberg PL (2005) Tec family kinases in T lymphocyte development and function. *Annu Rev Immunol* 23(1):549–600. doi:10.1146/annurev.immunol.22.012703.104743
  11. Sahu N, August A (2009) ITK inhibitors in inflammation and immune-mediated disorders. *Curr Top Med Chem* 9(8):690–703
  12. Cowan-Jacob SW, Möbitz H, Fabbro D (2009) Structural biology contributions to tyrosine kinase drug discovery. *Curr Opin Cell Biol* 21(2):280–287. doi:10.1016/j.cob.2009.01.012
  13. Huse M, Kuriyan J (2002) The conformational plasticity of protein kinases. *Cell* 109(3):275–282. doi:10.1016/S0092-8674(02)00741-9
  14. Jacobs MD, Caron PR, Hare BJ (2008) Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: structure of lck/Imatinib complex. *Prot Struct Funct Bioinf* 70(4):1451–1460. doi:10.1002/prot.21633
  15. 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(1):160–169. doi:10.1021/ci049885e
  16. Rella M, Rushworth CA, Guy JL, Turner AJ, Langer T, Jackson RM (2006) Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors. *J Chem Inf Model* 46(2):708–716. doi:10.1021/ci0503614
  17. Snow RJ, Abeywardane A, Campbell S, Lord J, Kashem MA, Khine HH, King J, Kowalski JA, Pullen SS, Roma T, Roth GP, Sarko CR, Wilson NS, Winters MP, Wolak JP, Cywin CL (2007) Hit-to-lead studies on benzimidazole inhibitors of ITK: discovery of a novel class of kinase inhibitors. *Bioorg Med Chem Lett* 17(13):3660–3665. doi:10.1016/j.bmcl.2007.04.045
  18. Riether D, Zindell R, Kowalski JA, Cook BN, Bentzien J, Lombaert SD, Thomson D, Kugler SZ Jr, Skow D, Martin LS, Raymond EL, Khine HH, O'Shea K, Woska JR Jr, Jeanfavre D, Sellati R, Ralph KLM, Ahlberg J, Labissiere G, Kashem MA, Pullen SS, Takahashi H (2009) 5-Aminomethylbenzimidazoles as potent ITK antagonists. *Bioorg Med Chem Lett* 19(6):1588–1591. doi:10.1016/j.bmcl.2009.02.012
  19. Cook BN, Bentzien J, White A, Nemoto PA, Wang J, Man CC, Soleymanzadeh F, Khine HH, Kashem MA, Kugler SZ Jr, Wolak JP, Roth GP, De Lombaert S, Pullen SS, Takahashi H (2009) Discovery of potent inhibitors of interleukin-2 inducible T-cell kinase (ITK) through structure-based drug design. *Bioorg Med Chem Lett* 19(3):773–777. doi:10.1016/j.bmcl.2008.12.028
  20. Lo HY, Bentzien J, White A, Man CC, Fleck RW, Pullen SS, Khine HH, King J, Woska JR Jr, Wolak JP, Kashem MA, Roth GP, Takahashi H (2008) 2-Aminobenzimidazoles as potent ITK antagonists: de novo design of a pyrrole system targeting additional hydrogen bonding interaction. *Tetrahedron Lett* 49(51):7337–7340. doi:10.1016/j.tetlet.2008.10.057
  21. Moriarty KJ, Takahashi H, Pullen SS, Khine HH, Sallati RH, Raymond EL, Woska JR Jr, Jeanfavre DD, Roth GP, Winters MP, Qiao L, Ryan D, DesJarlais R, Robinson D, Wilson M, Bobko M, Cook BN, Lo HY, Nemoto PA, Kashem MA, Wolak JP, White A, Magolda RL, Tomczuk B (2008) Discovery, SAR and X-ray structure of 1H-benzimidazole-5-carboxylic acid cyclohexyl-methyl-amides as inhibitors of inducible T-cell kinase (Itk). *Bioorg Med Chem Lett* 18(20):5545–5549. doi:10.1016/j.bmcl.2008.09.015
  22. Das J, Furch JA, Liu C, Moquin RV, Lin J, Spergel SH, McIntyre KW, Shuster DJ, O'Day KD, Penhallow B, Hung CY, Doweiko AM, Kamath A, Zhang H, Marathe P, Kanner SB, Lin T-A, Dodd JH, Barrish JC, Wityak J (2006) Discovery and SAR of 2-amino-5-(thioaryl)thiazoles as potent and selective Itk inhibitors. *Bioorg Med Chem Lett* 16(14):3706–3712. doi:10.1016/j.bmcl.2006.04.060
  23. Das J, Liu C, Moquin RV, Lin J, Furch JA, Spergel SH, McIntyre KW, Shuster DJ, O'Day KD, Penhallow B, Hung CY, Kanner SB, Lin TA, Dodd JH, Barrish JC, Wityak J (2006) Discovery and SAR of 2-amino-5-[(thiomethyl)aryl]thiazoles as potent and selective Itk inhibitors. *Bioorg Med Chem Lett* 16(9):2411–2415. doi:10.1016/j.bmcl.2006.01.115
  24. MacKerell AD, Bashford D, Bellott DRL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczman K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102(18):3586–3616. doi:10.1021/jp973084f
  25. Smellie A, Teig SL, Towbin P (1995) Poling: promoting conformational variation. *J Comput Chem* 16(2):171–187. doi:10.1002/jcc.540160205
  26. Smellie A, Kahn SD, Teig SL (1995) Analysis of conformational coverage. 2. Applications of conformational models. *J Chem Inf Comput Sci* 35(2):295–304. doi:10.1021/ci00024a019
  27. Smellie A, Kahn SD, Teig SL (1995) Analysis of conformational coverage. 1. Validation and estimation of coverage. *J Chem Inf Comput Sci* 35(2):285–294. doi:10.1021/ci00024a018
  28. Langer T, Krovat EM (2003) Chemical feature-based pharmacophores and virtual library screening for discovery of new leads. *Curr Opin Drug Discov Devel* 6:370–376
  29. 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(3):727–748. doi:10.1006/jmbi.1996.0897
  30. Berendsen HJC, van der Spoel D, van Drunen R (1995) GRO-MACS: a message-passing parallel molecular dynamics implementation. *Comput Phys Commun* 91(1–3):43–56. doi:10.1016/0010-4655(95)00042-E
  31. van der Spoel D, Lindahl E, Hess B, van Buuren AR, Apol E (2005) GROMACS user manual version 4.0. <http://www.gromacs.org>
  32. Schüttelkopf AW, van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr D* 60(8):1355–1363. doi:10.1107/S0907444904011679
  33. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Intermolecular forces. In: Pullman B (ed) *Interaction models for water in relation to protein hydration*. D. Reidel, Dordrecht, pp 331–342
  34. Darden T, York D, Pedersen L (1993) Particle mesh ewald: N-log(N) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089–10092. doi:10.1063/1.464397
  35. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINC: a linear constraint solver for molecular simulations. *J Comput Chem* 18(12):1463–1472. doi:10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h

36. Miyamoto S, Kollman PA (1992) Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* 13(8):952–962. doi:[10.1002/jcc.540130805](https://doi.org/10.1002/jcc.540130805)
37. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12):2577–2637. doi:[10.1002/bip.360221211](https://doi.org/10.1002/bip.360221211)
38. Lo HY (2010) Itk inhibitors: a patent review. *Expert Opin Ther Pat* 20(4):459–469. doi:[10.1517/13543771003674409](https://doi.org/10.1517/13543771003674409)
39. Chandrasekaran M, Sakkiah S, Thangapandian S, Namadevan S, Kim HH, Kim Y, Lee KW (2010) Pharmacophore design for anti-inflammatory agent targeting interleukin-2 inducible tyrosine kinase (Itk) B. *Korean Chem Soc* 31(11):3333–3340. doi:[10.5012/bkcs.2010.31.11.3333](https://doi.org/10.5012/bkcs.2010.31.11.3333)
40. Ravikumar M, Pavan S, Bairy S, Pramod AB, Sumakanth M, Kishore M, Sumithra T (2008) Virtual screening of cathepsin K inhibitors using docking and pharmacophore models. *Chem Biol Drug Des* 72(1):79–90. doi:[10.1111/j.1747-0285.2008.00667.x](https://doi.org/10.1111/j.1747-0285.2008.00667.x)
41. Vadivelan S, Sinha BN, Rambabu G, Boppana K, Jagarlapudi SARP (2008) Pharmacophore modeling and virtual screening studies to design some potential histone deacetylase inhibitors as new leads. *J Mol Graph Model* 26(6):935–946. doi:[10.1016/j.jmgm.2007.07.002](https://doi.org/10.1016/j.jmgm.2007.07.002)
42. 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. doi:[10.1021/ci050064z](https://doi.org/10.1021/ci050064z)