

Molecular docking guided 3D-QSAR CoMFA analysis of N-4-Pyrimidinyl-1H-indazol-4-amine inhibitors of leukocyte-specific protein tyrosine kinase

Mahendra Awale · C. Gopi Mohan

Received: 7 February 2008 / Accepted: 9 June 2008 / Published online: 15 July 2008
© Springer-Verlag 2008

Abstract Inhibition of leukocyte-specific protein tyrosine kinase (Lck) activity offers one of the approaches for the treatment of T-cell mediated inflammatory disorders including rheumatoid arthritis, transplant rejection and inflammatory bowel disease. To explore the relationship between the structures of the N-4 Pyrimidinyl-1H-indazol-4-amines and their Lck inhibition, 3D-QSAR study using CoMFA analysis have been performed on a dataset of 42 molecules. The bioactive conformation of the template molecule, selected as the most potent molecule **23** from the series was obtained by performing molecular docking at the ATP binding site of Lck, which is then used to build the rest of the molecules in the series. The constructed CoMFA model is robust with r_{cv}^2 of 0.603 and conventional r^2 of 0.983. The predictive power of the developed model was obtained using a test set of 10 molecules, giving predictive correlation coefficient of 0.921. CoMFA contour analysis was performed to obtain useful information about the structural requirements for the Lck inhibitors which could be utilized in its future design.

Keywords Lck · Leukocyte-specific protein tyrosine kinase · CoMFA · QSAR

Introduction

Protein tyrosine kinases (PTKs) are critically involved in signalling pathways that regulate cell growth, differentiation, activation, and transformation [1]. Malfunctions of cellular signalling have been associated with many diseases including cancer and diabetes. PTKs can be divided into receptor tyrosine kinases (RTKs) and non-receptor (cytosolic) tyrosine kinases [2, 3]. Non-receptor tyrosine kinases belonging to the Src family are key players in signal transduction. Some Src kinases (Fyn, Src, and Yes) are found in most cell types whereas others exhibit a more restricted tissue distribution (Lck, Hck, Blk) and have more specific tasks in signal transduction [4]. The induction of PTK activity by the T-cell antigen receptor (TCR) is essential to couple it to downstream pathways which trigger proliferation and differentiation of resting T cells into effector T cells [5]. However, the component subunits of TCR do not contain any intrinsic tyrosine kinase domains. Rather, the TCR initiates signal transduction by interacting sequentially with two different families of cytoplasmic non-receptor PTKs. The first of these is the Src family, in particular Lck and Fyn. Analysis of Lck deficient T-cell lines has indicated that Lck is absolutely required for TCR signalling. Indeed, in the absence of Lck, the TCR fails to induce any tyrosine phosphorylation cytoplasmic proteins, and all downstream signalling events are blocked. The second family of PTKs which interact with the TCR comprises ZAP-70 and Syk.

Lck is a lymphoid-specific cytosolic PTK, which is essential for the T-cell development and function [6, 7]. It is constitutively associated with the cytoplasmic portions of the CD4 and CD8 surface receptors and plays a key role in TCR-linked signal transduction pathways. Association of the TCR with a peptide antigen-bound MHC complex facilitates the interaction of CD4 and CD8 with MHC class I

Electronic supplementary material The online version of this article (doi:10.1007/s00894-008-0334-8) contains supplementary material, which is available to authorized users.

M. Awale · C. G. Mohan (✉)
Centre for Pharmacoinformatics,
National Institute of Pharmaceutical Education and Research (NIPER),
Sector 67, S.A.S. Nagar,
160 062 Punjab, India
e-mail: cmohan@niper.ac.in

and class II molecules, respectively, and thereby recruits the associated Lck to the vicinity of the TCR/CD3 complex [8, 9, 10]. Lck then phosphorylates tyrosine residues within the immunoreceptor tyrosine-based activation motifs (ITAMs) in the cytoplasmic tails of the TCR α -chains and CD3 subunits [11]. The phospho-ITAMs serve as docking sites for Src homology domain 2 (SH2)-containing molecules, predominantly ZAP-70 and Syk. Due to this ZAP-70 undergo tyrosine phosphorylation, become enzymatically active, and further phosphorylate downstream effector molecules [12]. In addition, Lck contributes to signaling by other receptor molecules. It associates directly with the cytoplasmic tail of CD2, and upon engagement of the CD2 molecule, Lck undergoes hyperphosphorylation and activation [13]. Lck activation leads to induction of cytokines including IL-2 which is capable of supporting T-cell proliferation. Therefore inhibition of Lck kinase activity offers one approach to the treatment of T-cell mediated inflammatory disorders including rheumatoid arthritis, transplant rejection and inflammatory bowel disease [14].

There are several reports on Lck inhibitors. Much of the earlier work was on natural products which are mainly related to the non-selective tyrosine kinase inhibitors. Recently other inhibitors have been derived from the tyrphostins, quinazolines, pyrazolopyrimidine PP1, thiazole, benzothiazole, quinolines with 2- methyl 5-hydroxy aniline as a 4-substituent, pyrido [2, 3-d] pyridine inhibitors, which showed potent activity against Lck. 4-Anilino quinolines and quinazolines, are potent inhibitors of tyrosine kinases [15].

Review of literature showed various *in silico* studies to understand the mode of action and the relationship between physicochemical properties and inhibitory activity of different class of Lck inhibitors. Bharatham et al. performed pharmacophore modeling and molecular docking of various P56 Lck inhibitors [15]. Structure-activity relationships and binding mode analysis using molecular docking of benzothiazole analogs were also carried out by the same research group [15].

The 2,4-dianilino pyrimidines are another well studied series of potent PTK inhibitors including Lck. Structure-activity relationship study of 2-methyl-5-hydroxyaniline substituents at the 4th position lead to potent compounds but with poor pharmacokinetics. Replacement of this substituent by 4-amino(5-methyl-1H-indazole) yielded compounds having improved pharmacokinetic properties and enzyme potency [16].

The structure-activity relationship of N-4-Pyrimidinyl-1H-indazol-4-amine with several sulphonamide 2-substituents showed that the amide group is not crucial for activity but substituents at the 4th position is significant, which is the focus of the present study [16]. Docking analysis using GOLD software predicted that pyrimidine forms two typical

hydrogen bonds to the hinge of the Lck ATP-site (between the backbone NH and carbonyl of Met319, and the N₁ atom of the pyrimidine and the 2-anilino NH, respectively) [17, 18]. The 2-methyl contributes about 10-fold to the activity and its substitution on the ring at 3rd and 4th-positions did not restore full activity [16]. To further explore the relationship between the structure of the N-4-Pyrimidinyl-1H-indazol-4-amine series of molecules and their Lck inhibitor activity, three dimensional quantitative structure-activity relationship (3D-QSAR) study using comparative molecular field analysis (CoMFA) was performed.

Methodology

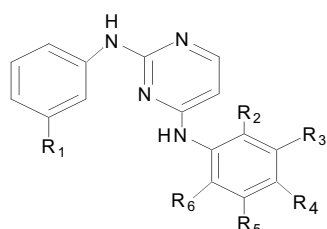
CoMFA

3D-QSAR method CoMFA was introduced by Cramer (1998) [19], in which an assumption is made that the interaction between an inhibitor and its molecular target is primarily non-covalent in nature and shape dependant. Therefore QSAR may be derived by sampling the steric and electrostatic fields surrounding a set of ligands and correlating the difference in these fields to biological activity. CoMFA calculate steric field using Lennard-Jones potential and electrostatic field using coulomb potential.

Computational details

Dataset for analysis

N-4-Pyrimidinyl-1H-indazol-4-amine inhibitors of Lck reported by Bamborough et al. was used [16] to construct CoMFA model and for the analysis of its physico-chemical features. The structure and experimental value of activity (IC₅₀) for the 42 molecules used in this study are shown in Table 1. The IC₅₀ values is measured as the 50% inhibition of the Lck activity which was assessed using time resolved fluorescence resonance energy transfer assay [16]. The molecules under study were built using SYBYL7.1 [20] molecular modeling package installed on a Silicon Graphics Fuel Work station running IRIX 6.5. As the crystal structure of Lck (PDB ID: 1PQC) is available, we have performed docking analysis of the most potent molecule (**23**) at the ATP binding site of Lck. Molecule **23** was initially optimized using PM3 Hamiltonian by applying Gasteiger-Hückel charges [21]. The optimized conformation of molecule **23** was then subjected for docking in order to obtain broadly its bioactive conformation for further CoMFA analysis. The rest of the 41 molecules were built by modifying required substitution on bioactive conformation of the molecule (**23**) obtained from docking analysis;

Table 1 Structures, pIC₅₀ values (actual and predicted) and residuals of substituted N- 4- Pyrimidinyl-1H-indazol-4-amines2, 4 dianilino pyrimidines with different substitutions (Molecules **2-24**)

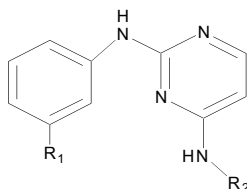
| Mol. ID | R ₁ | R ₂ | R ₃ | R ₄ | R ₅ | R ₆ | pIC ₅₀ ^b | | Residual |
|------------------|---------------------------------|----------------|----------------|----------------|---------------------------------|----------------|--------------------------------|-----------|----------|
| | | | | | | | Actual | Predicted | |
| 2 | CONH ₂ | Me | H | H | OH | H | 7.89 | 8.06 | -0.17 |
| 3 | CONH ₂ | H | H | H | OH | H | 6.89 | 6.80 | 0.09 |
| 4 | CONH ₂ | H | Me | H | OH | H | 7.14 | 6.79 | 0.34 |
| 5 | CONH ₂ | H | H | Me | OH | H | 6.73 | 6.83 | -0.10 |
| 6 | CONH ₂ | Br | H | H | OH | H | 7.39 | 7.42 | -0.03 |
| * ^a 7 | CONH ₂ | iPr | H | H | OH | H | 6.06 | 8.09 | 2.03 |
| 8 | CONH ₂ | Me | H | H | H | H | 7.06 | 6.92 | 0.14 |
| 9 | CONH ₂ | Me | H | OH | H | H | 7.40 | 7.49 | -0.09 |
| ^a 10 | CONH ₂ | Me | H | H | H | OH | 7.00 | 7.30 | 0.30 |
| ^a 11 | CONH ₂ | Me | H | H | F | H | 6.57 | 6.83 | 0.26 |
| ^a 12 | CONH ₂ | Me | H | H | Cl | H | 6.43 | 6.80 | 0.37 |
| 13 | CONH ₂ | Me | H | H | OMe | H | 7.07 | 6.94 | 0.12 |
| 14 | CONH ₂ | Me | H | H | NH ₂ | H | 7.02 | 7.26 | -0.24 |
| ^a 15 | CONH ₂ | Me | H | H | CONH ₂ | H | 5.89 | 5.72 | -0.17 |
| 16 | CONH ₂ | Me | H | H | CONMe ₂ | H | 5.04 | 5.02 | 0.01 |
| 17 | CONH ₂ | Me | H | H | NHSO ₂ Me | H | 5.12 | 5.15 | -0.03 |
| 18 | CONH ₂ | Me | H | H | CH ₂ NH ₂ | H | 5.47 | 5.50 | -0.03 |
| 19 | CONH ₂ | H | H | H | OMe | H | 5.57 | 5.61 | -0.04 |
| 20 | CONH ₂ | H | H | H | F | H | 5.49 | 5.52 | -0.03 |
| 21 | CONH ₂ | H | H | H | N(Me)COMe | H | 5.75 | 5.70 | 0.04 |
| 22 | CONH ₂ | F | H | H | H | H | 5.96 | 5.93 | 0.02 |
| 23 | SO ₂ NH ₂ | Me | H | H | OH | H | 8.08 | 7.84 | 0.23 |

showing satisfying hydrogen bonding interactions and spatial constrains. After building molecules Gasteiger-Hückel charges were applied to all the molecules and subjected to constrain minimization in the gaseous phase, in

which core bioactive conformation of the molecules was kept un-optimized while rest of the substitution was subjected to energy minimization. These molecules were then used to construct 3D-QSAR model. The IC₅₀ value in

Table 1 (continued)

| | | | | | | | | | |
|-----------------|---------------------------------|----|---|---|-----|---|------|------|------|
| ^a 24 | SO ₂ NH ₂ | Me | H | H | OMe | H | 7.27 | 7.52 | 0.25 |
|-----------------|---------------------------------|----|---|---|-----|---|------|------|------|

2-anilino pyrimidines with different substitutions (Molecules **26-43**)

| MOL ID | R ₁ | R ₂ | pIC ₅₀ ^b | | Residuals |
|-----------------|-------------------|----------------|--------------------------------|-----------|-----------|
| | | | Actual | Predicted | |
| 26 | CONH ₂ | | 5.53 | 5.45 | 0.07 |
| 27 | CONH ₂ | | 4.8 | 4.81 | -0.01 |
| 28 | CONH ₂ | | 5.53 | 5.51 | 0.01 |
| ^a 29 | CONH ₂ | | 5.75 | 5.55 | -0.20 |
| ^a 30 | CONH ₂ | | 5.36 | 5.11 | -0.25 |
| 31 | CONH ₂ | | 5.09 | 5.12 | -0.03 |

Table 1 (continued)

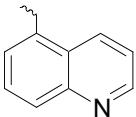
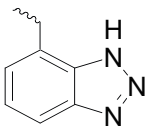
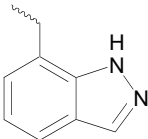
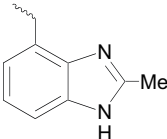
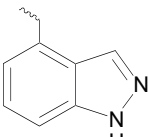
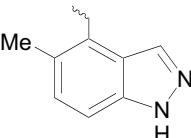
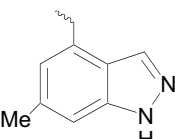
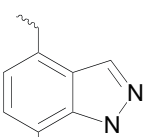
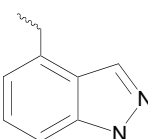
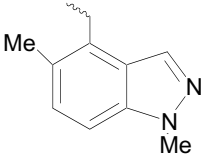
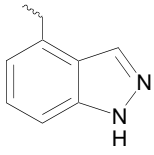
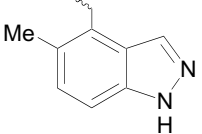
| 32 | CONH ₂ |  | 6.26 | 6.27 | -0.01 |
|-----------------|-------------------|---|--------------------------------|-----------|-----------|
| 33 | CONH ₂ |  | 6.31 | 6.31 | 0.00 |
| ^a 34 | CONH ₂ |  | 5.86 | 6.08 | 0.22 |
| MOL ID | R ₁ | R ₂ | pIC ₅₀ ^b | | Residuals |
| | | | Actual | Predicted | |
| 35 | CONH ₂ |  | 5.26 | 5.17 | 0.08 |
| 36 | CONH ₂ |  | 7 | 6.99 | 0.01 |
| 37 | CONH ₂ |  | 7.93 | 7.68 | 0.24 |
| 38 | CONH ₂ |  | 6.29 | 6.48 | -0.19 |
| 39 | CONH ₂ |  | 7.26 | 7.34 | -0.08 |
| 40 | CONH ₂ |  | 5.8 | 5.82 | -0.02 |

Table 1 (continued)

| | | | | | |
|-----------------|---------------------------------|---|------|------|-------|
| 41 | CONH ₂ |  | 5.56 | 5.58 | -0.02 |
| ^a 42 | SO ₂ NH ₂ |  | 7.49 | 7.32 | -0.17 |
| 43 | SO ₂ NH ₂ |  | 7.73 | 7.81 | -0.08 |

^aTest set molecules;

^bActual IC₅₀ (nM) is from Ref. [16] and is converted to pIC₅₀ using Equation (1)

*Outlier molecule

nano molar (nM) range for 42 molecules reported by Bamborough et al. [16] were converted into molar (M) range and then its logarithmic scale (pIC₅₀, M) were then used for subsequent QSAR analysis (Eq. 1) as the response variable

$$\text{pIC}_{50} = -\log \text{IC}_{50} \quad (1)$$

Molecular docking

Most potent molecule (**23**) was docked into the ATP binding site of the Lck crystal structure (PDB ID 1PQC), solved at 1.60 Å resolution, using FlexX program. FlexX is a fast, flexible docking method and uses incremental construction algorithm to place ligands into the active site. Standard default parameters of the FlexX program as implemented in SYBYL 7.1 were used during docking analysis [20]. All the flexibilities of the rotatable bonds of molecule (**23**) were considered and interaction and spatial constrains were applied in the process of docking, for identifying the best binding conformation of the molecule in the active site of Lck. Molecule **23** make strong hydrogen bonding interaction with Met319 and Asp382 active site amino acid residues of Lck, and is shown in Fig. 1. Bamborough et al. have also performed docking analysis on Lck using GOLD software, which supports the above result [16]. This docking guided conformation of molecule (**23**) were then further used for CoMFA analysis.

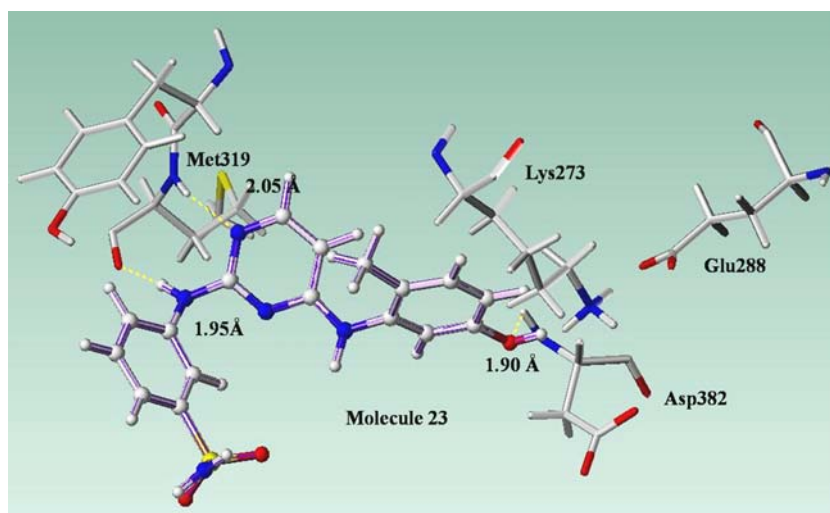
Molecular alignment

Molecular alignment is the most sensitive parameter in 3D-QSAR analysis. This renders the spatial alignment of molecules under study as one of the most sensitive and determining factors in obtaining robust and meaningful models. In the present study docking guided bioactive conformation of the template molecule (**23**) was used for aligning other 41 molecules from the series by common substructure alignment, using ALIGN DATABASE command in SYBYL. The common substructure used for alignment and superimposed structure after alignment is presented in Figs. 2a and b respectively.

CoMFA interaction energies

CoMFA electrostatic potential and steric fields were calculated at each grid intersection on a regularly spaced grid of 2.0 Å units in x, y, and z directions on the aligned dataset. The pattern of 3D cubic lattice (or grid) generated automatically by SYBYL/CoMFA routine, extended at least 4.0 Å beyond the volumes of all investigated molecules along these axes. The van der Waals potential and Coulombic terms, which represent steric and electrostatic fields respectively, were calculated using the standard Tripos force field method [22]. A distance dependent dielectric constant of 1.0 was used and sp³ hybridized carbon atom with van der waals radius of 1.52 Å and +1

Fig. 1 Hydrogen bonding interaction of molecule 23 (ball and stick) with ATP binding site residues in Lck. Hydrogen bonds are shown in yellow dotted lines



charge served as probe atom to calculate the steric and electrostatic fields. The steric and electrostatic cut offs are both set to the default 30 kcal/mol to calculate standard Tripos CoMFA fields [22].

CoMFA interaction energies is usually represented as 3D coefficient iso-contour maps which surround all lattice point, where QSAR is found to strongly deal with the change in interaction energy or binding affinity with respect to its structural changes. In simple words, iso-contour maps shows how the variation of steric or electrostatic property in structural features of molecules contained in the training set leads to increase or decrease in activity. The contour lines connect points (terms) in lattice (or grid) space having similar values (iso-contour). The colored polyhedra produced surround lattice (or grid) points where the scalar products of the associated QSAR coefficient and the standard deviation of all values in the corresponding column of the data matrix are higher or lower than the user-specified value. Also the resulting iso-contoured surfaces are usually closed polyhedra when CoMFA QSAR-coefficients take on extreme high or low values.

SYBYL [20] setting used different color to represent electrostatic iso-contour (blue and red) and steric (yellow and green) iso-contour. In the case of electrostatic iso-contour map, the region where presence or introduction of negative charge leads to increase in activity of the molecule, and is represented by red polyhedra (80% contribution); while the region with the presence or introduction of positive charge leads to better activity is represented as blue polyhedra (20% contribution). On the other hand in the case of CoMFA steric iso-contour, yellow polyhedra region (20% contribution) represents sterically disfavored point, and addition of bulky group at that point leads to decrease in activity. While green polyhedra regions (80% contribution) represent sterically favored point, where the bulky

group substitution at this region could lead to increase in activity of the molecule [20].

Partial least square (PLS) analysis

The relationship between the structural parameters (CoMFA interaction energies) and the Lck inhibitory activities has been quantified by the PLS algorithm [23]. The cross-validation analysis was performed using leave-one-out

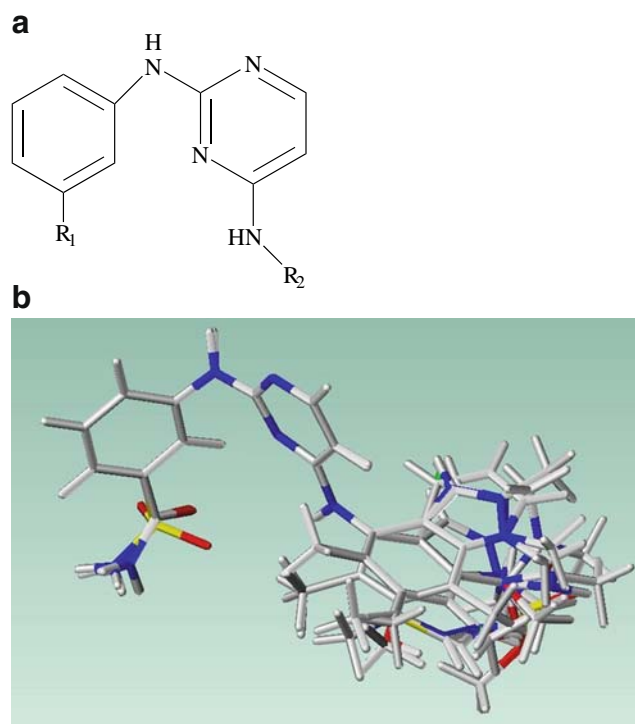


Fig. 2 a. Common substructure used for alignment b. Alignment of molecules

(LOO) method, where one molecule is removed from the dataset and its Lck activity is predicted using the model derived from the rest of the dataset. The cross-validated r_{cv}^2 that resulted in optimum number of components and lowest standard error of prediction was selected. To speed up the analysis with reduced noise, a minimum column filtering value of 2.0 kcal/mol was used for the cross-validation. Final analysis (non-cross-validation) was performed to calculate non-cross-validated (r^2) using the optimum number of components obtained from the LOO cross-validation analysis. To further assess the statistical confidence and robustness of the derived model, a 100-cycle bootstrap analysis was performed.

Predictive correlation coefficient (r_{pred}^2)

The predictive power of the 3D-QSAR models were determined from a set of 10 molecules that were excluded during model development. The optimization, alignment and all other steps of these test set molecules were the same as that of the training set molecules, described above, and their activities were predicted using the model produced by the training set. The predictive correlation (r_{pred}^2) based on the test set molecules, is computed using (Eq. 2)

$$r_{pred}^2 = \frac{(SD - PRESS)}{SD} \quad (2)$$

Where, SD is defined as the sum of the squared deviations between the biological activity of the test set and mean activity of the training set molecules and PRESS is the sum of the squared deviation between the predicted and actual activity values for each molecule in the test set.

Results and discussion

CoMFA models

The CoMFA analysis was performed to explore the structure-activity relationship of N-4-Pyrimidinyl-1H-indazol-4-amine inhibitors of Lck. The dataset consisting of 42 molecules was divided into training set of 32 molecules and test set of 10 molecules by considering the structural diversity and activity range.

Initially we have developed three different CoMFA models: Model 1 is constructed using both steric and electrostatic fields, model 2 with only steric fields and model 3 with only electrostatic fields is presented in Table 2. Model 1 with both steric and electrostatic field was selected as the best model due to its statistical significance and predictive power (Table 2). Model 1 was used for further counter map analysis for QSAR predictions.

Table 2 PLS result summary

| Parameters | SE | S | E |
|---------------------------------|---------|-------|-------|
| PRESS | 0.421 | 6.60 | 4.06 |
| r_{cv}^2 | 0.603 | 0.418 | 0.043 |
| NOC | 7 | 9 | 2 |
| SEE | 0.143 | 0.225 | 0.757 |
| R^2 | 0.983 | 0.962 | 0.428 |
| F-test | 200.680 | 61.53 | 10.84 |
| r_{bs}^2 | 0.992 | | |
| SD _{bs} | 0.004 | | |
| r_{pred}^2 | 0.921 | | |
| Fraction of field contributions | | | |
| Steric | 0.622 | | |
| Electrostatic | 0.378 | | |

S = Steric field, E = Electrostatic field, PRESS = Predictive residual sum of squares for training set; r_{cv}^2 = Cross-validated correlation coefficient by PLS LOO method; NOC = Optimum number of components as determined by PLS LOO cross validation study; SEE = Standard error of estimate, r^2 = Conventional correlation coefficient; r_{bs}^2 = Correlation coefficient after 100 runs of bootstrapping; SD_{bs} = Standard deviation from 100 runs of bootstrapping; r_{pred}^2 = Predictive correlation coefficient

CoMFA model 1 analysis

3D-QSAR CoMFA is an efficient in silico methodology employing both statistical techniques (PLS) and interactive graphics (contour maps) for correlating shapes of molecules with their observed biological activities. For each molecule of a series of known substrates the steric and electrostatic interaction energies of the atoms with the test probe is calculated at spatial coordinates around the molecule. The resulting interaction energies calculated at all intersections in a three-dimensional grid or lattice surrounding the molecule form the quantitative shape descriptors entered along with the molecule's measured biological activity as a row in a data table. Subsequent analysis of the data table by a partial least squares (PLS) cross-validation technique yields a set of coefficients which reflect the relative contribution of the shape elements of the molecular series to differences in biological activities.

The statistical results of CoMFA PLS analysis is presented in Table 2. The constructed CoMFA model is robust with r_{cv}^2 of 0.603 and conventional r^2 of 0.983 having 7 optimum number of components (NOC). The r_{cv}^2 represents goodness of internal prediction whereas r^2 represent the goodness of fit of a QSAR model. The external predictive capability of a QSAR model is generally checked using test set molecules. All other procedures including geometry optimization, charge calculation, interaction energy calculation and alignment of the test set molecules were done in manner analogues to the training set molecules. The predictive power (r_{pred}^2) of developed

CoMFA model is good having 0.921 values. Further the statistical validity and stability of the CoMFA model was assessed by running boot strap analysis for 100 runs. The higher r_{bs}^2 value of 0.992 obtained after 100 runs of bootstrapping supports this analysis. In addition to this low value of standard error of estimates (0.143) supports the significance of the developed model. The steric to electrostatic fields were found to be (62.2:37.8), which implies that steric part have more contribution for interaction with Lck than electrostatic part.

During rigorous cycle of CoMFA model 1 development, we found molecule 7 behaving as outlier and not to fit either test set or training set. Their inclusion in training set disturb the statistics of developed model while there inclusion in test set although improving PLS statistics showed poor predictive power. As the molecule 7 behaving as outlier, was not taken into account while calculating r^2 predictive value. One of the possible reasons for such behavior is that molecule 7 has bulkier isopropyl substitution at the 2nd position which is different or more bulky than the rest of the molecules from the series. It was already known that such a bulkier group at 2nd position not only affects the orientation of aromatic group into back ATP binding pocket, but also has not been able to accommodate (due to steric collision) in small lipophilic hole in the active site of Lck [16]. Outlier generally occurs when the bio-active conformation is different from the rest of the molecules in series, as seen in the case of molecule 7.

The scatter plots of actual versus predicted activity of training and test set molecules are shown in Figs. 3a and b respectively. Table 1 shows the structures and the corresponding actual and predicted pIC_{50} values of the molecules under study for CoMFA model development. The actual IC_{50} of all the molecules are reported by Bamborough et al. [16].

Contour map analysis of CoMFA model 1

Electrostatic and steric contour maps obtained using CoMFA method is a powerful tool to explore the protein-ligand interactions. The 3D contour maps generated have color coded regions (polyhedrons) in the vicinity of the molecule, that requires particular physicochemical property to improve its binding affinity. The physicochemical properties of different colored polyhedrons representing electrostatic and steric features of the molecule was discussed earlier in CoMFA interaction energy section.

The CoMFA electrostatic and steric contours of most active molecule 23 is displayed in Figs. 4a and b, while the same for molecule 37 are provided in the supplementary material Figs. 1sa and 1sb respectively.

The CoMFA electrostatic contour map shown in Fig. 4a (1sa) displayed one blue and three red polyhedrons (two are

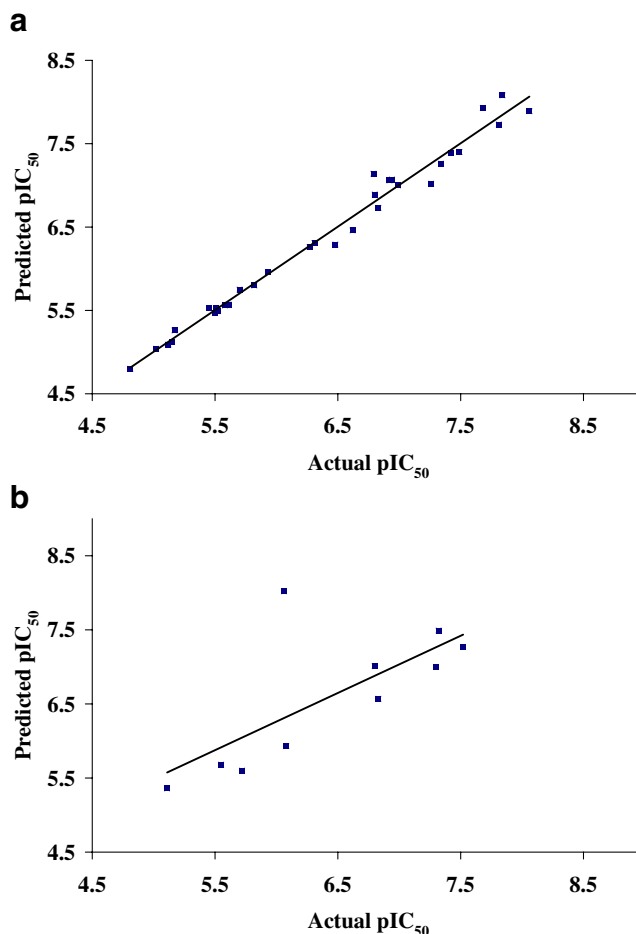


Fig. 3 Scatter plot of actual versus predicted pIC_{50} of the training set (a) and test set (b) molecules

very small). The one big blue polyhedron around the 3rd and 4th position of phenyl ring of 4-aniline substituent relative to linker amine suggest that replacement of hydroxyl group by another basic electropositive group (capable of forming hydrogen bond) was able to restore activity of molecules. The blue polyhedron mainly concentrates on N_1 nitrogen atom of indazole (shown in Figure 1sa) which is known to form hydrogen bond with the side chain oxygen of Glu288. This blue polyhedron at N_1 atom of indazole ring accounts for activity of most of the indazole ring substituted molecules (32–33, 36–39). The small red polyhedron between 5th and 6th positions of phenyl ring of 4-aniline substituent shows that electronegative group (able to form hydrogen bond) at 5th position of phenyl was able to retain activity, which is known to form strong hydrogen bonding interaction with the backbone NH of Asp382 and side chain of Glu288 [16]. This accounts for better activity of most potent molecule (23) as compared to other molecules (molecules 15–22). Molecule 14 having basic amine group was able to restore some activity as

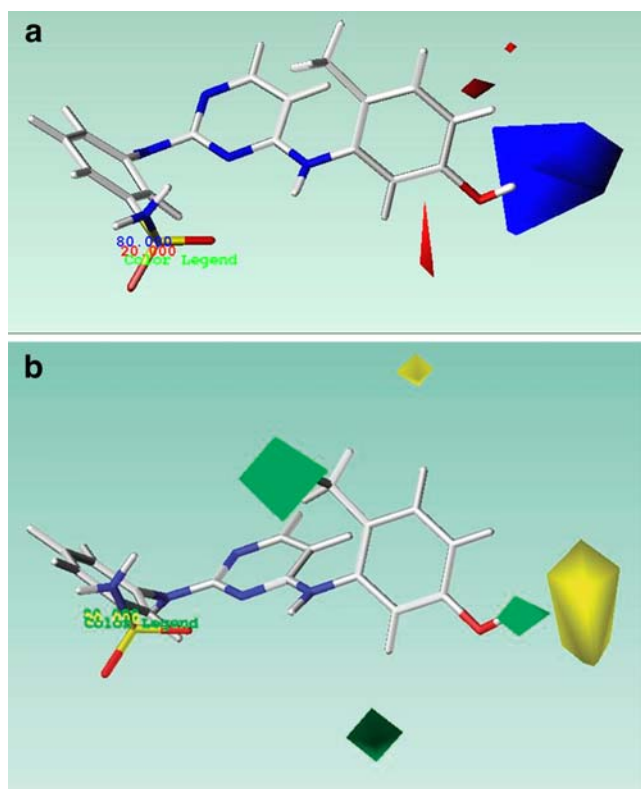


Fig. 4 **a.** CoMFA electrostatic contour map. Positive charge favored areas (contribution level 80%) are represented by blue polyhedra. Negative charge favored areas (contribution level 20%) are represented by red polyhedra. The active molecule 23 shown in capped sticks. **b.** CoMFA steric contour map. Sterically favored areas (contribution level 80%) are represented by green polyhedra. Sterically disfavored areas (contribution level 20%) are represented by yellow polyhedra. The active molecule 23 shown in capped sticks

compared to other molecules (**15–22**), which although having basic electropositive group was not able to restore the activity, as discussed below. It gives the way to consider other parameter (steric) in addition to electrostatic property of the molecules alone.

The CoMFA steric contour plot presented in Fig. 4b (**1sb**) showed three green and two yellow polyhedrons. The one green polyhedron surrounds methyl at 2nd position of the phenyl ring of 4-aniline substituents. In addition to that, there is a small yellow polyhedron around 2nd and 3rd position of phenyl ring of 4-aniline substituents. It suggests that optimum bulk at 2nd position of phenyl ring is essential for activity, which is known to be extended (fill) into small lipophilic hole created by side chains of the Lck N-terminal lobe (Val259, Ala271, and Lys273) [16]. Substituents at 2nd position are also known to affect the orientation of the phenyl ring which extends into the hydrophobic ATP binding region and hence is critical for its activity. The methyl group is most appropriate substituent at 2nd position of phenyl ring which accounts for

better activity of molecule **23** as compared to other molecules (**3–6**).

The molecule **7** with isopropyl substituent at 2nd position of phenyl ring is unable to restore the complete activity as compare to molecule **2**, because isopropyl substituent is unable to fill small lipophilic hole created by side chains of the Lck N-terminal lobe. The importance of green polyhedron (2-methyl substituent) was also observed in the case of indazole ring substituted molecules. Molecule **37** having 2-methyl substituent is more potent than molecule **36** which lacks this substituent. In contrary to these molecules (**15–22**) besides having 2-methyl substituent are less potent. Molecules (**15–22**) have yellow polyhedron which surrounds 5th-position of phenyl ring of 4-aniline substituent, suggesting that size of substituents at this position is limited. Hence although molecules (**15–22**) having appropriate 2-methyl substituent and basic electropositive group at 5th position of phenyl ring, its steric property make these molecules less potent as compared to other molecules such as molecule **14** having amino substitution at the 5th position.

In the case of indazole ring substituted molecules (**32–39**) the yellow polyhedron mainly concentrates on N₁ nitrogen atom of indazole ring. This account for better activity of potent molecule **37** as compared to other molecules which have methyl group substituted on N₁ nitrogen of indazole ring (molecules **40** and **41**). There is also one small green polyhedron at the 4th position of the phenyl ring of 4-aniline substituent (4-indazole substituent) indicating that substitution of small group at this position can be tolerated. This accounts for better activity of molecule **39** which has methyl substituent at the 4th position of indazole ring as compared to other molecules which either has such a group at the 3rd position (molecule **38**) or which lack such a group (molecule **36**). The small green polyhedron surrounding ortho position to linker suggests that substitution of small bulkier group at this position add better to the activity (molecule **37**).

Conclusions

To explore the structure-activity relationships of N-4-Pyrimidinyl-1H-indazol-4-amines for their Lck inhibitory activity and to build the statistically significant model with good predictive power 3D-QSAR CoMFA study was performed. The developed CoMFA model showed good predictive power with low residuals for test set molecules. Counter map analysis shows that substitution of electropositive group around the 3rd and 4th positions of phenyl ring of 4-aniline substituent relative to linker amine was able to restore activity of the molecules. This electropositive group account for the activity of most of the indazole a

substituted molecule which is known to be interacting with Glu288 of Lck. Also substitution of electronegative group at positions 5 and 6 of phenyl ring of 4-aniline substituent was able to restore the activity of molecules. In addition to this optimum bulk substitution at the 2nd position of phenyl ring is essential for activity, which is known to be extended (fill) into small lipophilic space created by side chains of the Lck N-terminal lobe. The methyl group is most appropriate substituent at the 2nd position of phenyl ring. The size of substitution surrounding the 5th position of phenyl ring of 4-aniline substituent should be limited for better activity, while a smaller group could be tolerated at the 4th position. Thus, CoMFA contour analysis have provided useful information about the structural requirements for the Lck inhibitors, which could be utilized in the future design of more potent Lck inhibitors.

References

1. Isakov N, Biesinger B (2000) *Eur J Biochem* 267:3413–3421
2. Neet K, Hunter T (1996) *Genes to Cells* 1:147–169
3. Robinson DR (2000) *Oncogene* 19:5548–5557
4. Schweimer K, Hoffmann S, Bauer F, Friedrich U, Kardinal C, Feller Sm, Biesinger B, Sticht H (2002) *Biochemistry* 41:5120–5130
5. Weiss A, Littman DR (1994) *Cell* 76:263–274
6. Sefton BM (1991) *Oncogene* 6:683–686
7. Straus DB, Weiss A (1992) *Cell* 70:585–593
8. Rudd CE, Trevillyan JM, Dasgupta JD, Wong LL, Schlossman SF (1988) *Proc Natl Acad Sci of America* 85:5190–5194
9. Shaw AS, Amrein KE, Hammond C, Stern DF, Sefton BM, Rose JK (1989) *Cell* 59:627–636
10. Veillette A, Bookman MA, Horak EM, Bolen JB (1988) *Cell* 55:301–308
11. Isakov N (1998) *Adv Immunol* 69:183–247
12. Chan AC, Dalton M, Johnson R, Kong GH, Wang T, Thoma R, Kurosaki T (1995) *EMBO J* 14:2499–2508
13. Carmo AM, Mason DW, Beyers AD (1993) *Eur J Immunol* 23:2196–2201
14. Bolen JB, Brugge JS (1997) *Annu Rev Immunol* 15:371–404
15. Bharatham N, Bharatham K, Lee KW (2007) *Bull Korean Chem Soc* 28:200–206
16. Bamborough P, Angell RM, Bhamra I, Brown D, Bull J, Christopher JA, Cooper AWJ, Fazal LH, Giordano I, Hind L, Patel VK, Ranshaw LE, Sims MJ, Skone PA, Smith KJ, Vickerstaff E, Washington M (2007) *Bioorg Med Chem Lett* 17:4363–4368
17. Zhu X, Kim J, Rose P, Stover D, Toledo L, Zhao H, Morgenstern A (1999) *Structure* 7:651–661
18. Jones G, Willet P, Glen R, Leach A, Taylor R (1997) *J Mol Biol* 267:727–748
19. Cramer RD, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110:5959–5967
20. Sybyl 7.1; Tripos Inc. SI, Mo 63144 USA
21. Gasteiger J, Marsili M (1980) *Tetrahedron* 36:3219–3228
22. Clark M, Cramer RD III, Van Opdenbosch N (1989) *J Comput Chem* 10:982–1012
23. Cramer RD, Patterson DE, Frank IE (1988) *Quant Struct Act Relat* 7:18–25