Journal of Molecular Graphics and Modelling 28 (2010) 577–591

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Deciphering ligand dependent degree of binding site closure and its implication in inhibitor design: A modeling study on human adenosine kinase

Savita Bhutoria, Nanda Ghoshal *

Structural Biology and Bioinformatics Division, Indian Institute of Chemical Biology (A unit of CSIR), 4 Raja S.C. Mullick Road, Jadavpur, Kolkata 700032, India

ARTICLE INFO

Article history:
Received 4 September 2009
Received in revised form 4 December 2009
Accepted 8 December 2009
Available online 16 December 2009

Keywords:
Adenosine kinase
Semi-open conformation
SAR
Docking
Pharmacophore modeling

ABSTRACT

Protein flexibility plays a significant role in drug research due to its effect on accurate prediction of ligand binding mode and activity. Adenosine kinase (AK) represents a highly flexible binding site and is known to exhibit large conformational changes as a result of substrate or inhibitor binding. Here we propose a semi-open conformation for ligand binding in human AK, in addition to the known closed and open forms. The modeling study illustrates the necessity of thorough understanding of the conformational states of protein for docking and binding mode prediction. It has been shown that predicting activity in the context of correct binding mode can improve the insight into conserved interactions and mechanism of action for inhibition of AK. Integrating the knowledge about the binding modes of ligands in different conformational states of the protein, separate pharmacophore models were generated and used for virtual screening to explore potential novel hits. In addition, 2D descriptor based clustering was done to differentiate the ligands, binding to closed, semi-open and open conformations of human AK. The results indicated that binding of all AK inhibitors cannot be described by same rules, instead, they represent a rule based preference for inhibition. This inference about tubercidins binding to semi-open conformation of human AK may facilitate in finding much extensive space for AK inhibitors.

1. Introduction

Human Adenosine Kinase (AK), a promising therapeutic drug target [1,2] against a variety of diseases including hypertension [3], epilepsy [4], pain [5,6], diabetes [7], and inflammation [8,9], has recently attracted great interest in the search of potent and selective inhibitors. Inhibition of AK results in increased intracellular adenosine, which passes out of the cell via passive diffusion or via nucleoside transporter(s) to activate nearby cell-surface adenosine receptors and produce a wide variety of therapeutically beneficial activities against the said diseases [10,11]. There are a number of reports on synthesis, SAR [12,13] and QSAR studies of nucleosidic analogues as AK inhibitors [14,15]. Nucleoside kinases, such as AK, cytidine kinase, thymidine kinase, catalyze phosphorylation of a variety of nucleosides, including nucleoside-based drugs, giving rise to nucleoside-5′-O-phosphates which are incorporated into RNA or DNA via the corresponding triphosphates [16]. Thus, 5′-O-phosphates of these nucleosides are responsible not only for the pharmacological activity but also for their cytotoxic side effects [12]. Therefore, designing molecules having different scaffold than the nucleosides is demanding. With the objective of finding new scaffolds, docking and pharmacophore analysis were carried out. In the present study, we have discussed and illustrated the crucial interactions for inhibition and proposed a rationale for new inhibitor design.

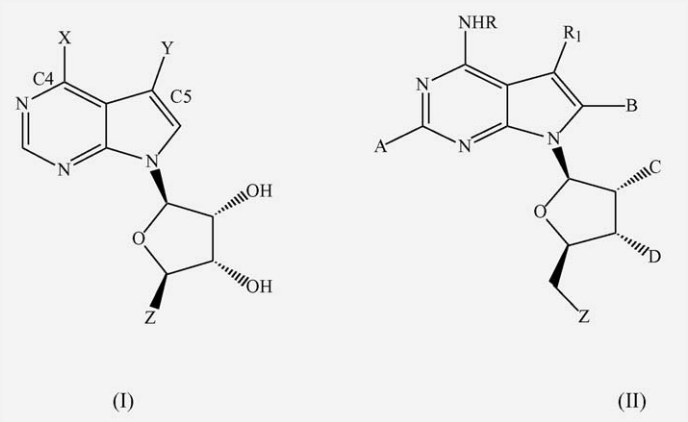
Two categories of compounds are known as inhibitors of AK, nucleosides [17] and non-nucleosides [18]. In an earlier study, a pharmacophore model for nucleoside and non-nucleoside AK inhibitors was developed, considering their similar binding modes [19]. Like other ribokinase family members, adenosine kinase undergoes large conformational changes as a result of ligand binding [20]. Recently, it has been found that nucleoside and non-nucleoside inhibitors bind in two different, closed and open conformations of enzyme, respectively. [20]. Nucleoside binding triggers a hinge bending by Glycine-Glycine (GG) switch on account of sugar anchoring with the apo form of enzyme.

The present study involves a series of aryl and non-aryl nucleosidic inhibitors (tubercidins) with varying degrees of substitutions at C4 and C5 positions (Table 1) [12,13]. Based on an exhaustive docking analysis, using various parameterizations, we have found that due to larger size than the required size in the closed binding site, aryl nucleoside analogues cannot accommodate in the closed binding site as adenosine. This study has illustrated that even all nucleosidic inhibitors cannot bind in similar fashion in the adenosine binding site. Hence, a need for

* Corresponding author. Tel.: +91 33 2473 3491x254/236; fax: +91 33 2473 0284/5197.
E-mail address: nghoshal@iicb.res.in (N. Ghoshal).

1093-3263/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2009.12.001

Table 1
Structures and activity data of nucleosidic analogues under study.



(I) (II)

I

Compd No.	X	Y	Z	AK IC_{50} (μM)[a]
1	NH_2	I	CH_2OH	0.026
2	NH_2	I	CH_3	0.009
3	NH_2	Br	CH_3	0.04
4	NH_2	Br	CH_2OH	0.12
5	NH_2	Cl	CH_2OH	0.21
6	Cl	I	CH_3	0.003
7	Cl	I	CH_2OH	0.024
8	Cl	CH_3	CH_3	0.535
9	Cl	SCH_3	CH_3	0.07
10	Cl	I	CH_2N_3	0.009
11	Cl	Br	CH_2N_3	0.1
12	Cl	I	CH_2OCH_3	0.45
13	NH_2	SCH_3	CH_3	1.35
14	NH_2	COOEt	CH_3	1.0
15	NH_2	I	CH_2N_3	0.035
16	NH_2	Br	CH_2N_3	0.063
17	NH_2	I	$CH{=}CH_2$	0.1
18	NH_2	I	CH_2CH_3	0.06
19	NH_2	Br	CH_2CH_3	0.5
20	$NH\text{-}CH_3$	I	CH_2OH	1.2
21	SH	I	CH_3	6.5
22	SCH_3	I	CH_3	0.045
23	*S-allyl*	I	CH_3	2.8
24	*S-n*-C_4H_9	I	CH_3	1.0
25	*S-benzyl*	I	CH_3	1.0
26	S-p-nitrobenzyl	I	CH_3	2.0
27	NH_2	I	CH_2NH_2	0.0006
28	NH_2	Br	CH_2NH_2	0.0002
29	NH_2	H	CH_2NH_2	13.0
30	Cl	I	CH_2NH_2	0.0001
31	NH_2	CN	CH_2OH	0.31
32	NH_2	CN	CH_3	0.31
33	NH_2	$CONH_2$	CH_2OH	0.47
34	NH_2	$CONH_2$	CH_3	0.46
35	Cl	Br	CH_3	0.05
36	NH_2	CH_3	CH_3	16.5
37	NH_2	I	CH_2OCH_3	1.2

II

Compd No.	R	R_1	A	B	C	D	Z	AK IC_{50} (μM)[a]
38	Ph	I	H	H	OH	OH	H	0.1
39	Ph-CH_2	I	H	H	OH	OH	H	0.8
40	4-F-Ph	I	H	H	OH	OH	H	0.55
41	4-MeO-Ph	I	H	H	OH	OH	H	0.775
42	4-HO-Ph	I	H	H	OH	OH	H	0.2
43	cyclohexyl	I	H	H	OH	OH	H	10
44	4-CN-Ph	I	H	H	OH	OH	H	1.2
45	Ph	I	H	H	OH	OH	OH	0.12
46	Ph	H	H	H	OH	OH	OH	1.25
47	H	Ph	H	H	OH	OH	H	0.32
48	Ph	Ph	H	H	OH	OH	H	0.0005
49	Ph-CH_2	Ph	H	H	OH	OH	H	0.03
50	4-F-Ph	Ph	H	H	OH	OH	H	0.0015
51	4-MeO-Ph	Ph	H	H	OH	OH	H	0.006

Table 1 (*Continued*)

II								
Compd No.	R	R_1	A	B	C	D	Z	AK IC_{50} (μM)[a]
52	4-HO-Ph	Ph	H	H	OH	OH	H	0.001
53	cyclohexyl	Ph	H	H	OH	OH	H	0.25
54	Ph	Ph	H	H	OH	OH	OH	0.0008
55	4-F-Ph	4-F-Ph	H	H	OH	OH	H	0.026
56	Ph	4-Cl-Ph	H	H	OH	OH	H	0.0012
57	4-Cl-Ph	4-Cl-Ph	H	H	OH	OH	H	0.0027
58	4-CN-Ph	Ph	H	H	OH	OH	H	0.014
59	4-CN-Ph	4-MeO-Ph	H	H	OH	OH	H	0.001
60	Ph	2-furanyl	H	H	OH	OH	H	0.0036
61	4-MeO-Ph	2-furanyl	H	H	OH	OH	H	0.009
62	3-pyridyl	Ph	H	H	OH	OH	H	0.025
63	4-Cl-Ph	Ph	H	H	OH	OH	H	0.0023
64	4-Me-Ph	Ph	H	H	OH	OH	H	0.0015
65	Ph	Ph	Me	H	OH	OH	H	0.047
66	Ph	Ph	H	Br	OH	OH	H	0.001
67	Ph	Ph	H	H	OH	OH	N_3	0.0015
68	Ph	Ph	H	H	OH	OH	NH_2	0.0063
69	Ph	Ph	H	H	H	OH	OH	0.3
70	Ph	Ph	H	H	H	OH	H	0.100
71	Ph	Ph	H	H	OH	H	H	0.044

[a] IC_{50} from enzyme inhibition assays performed on human recombinant AK enzyme [4,7].

another relaxed conformation of protein was realized. Earlier report on the existence of semi-open conformation of *T. gondii* AK [21], a homologue of human AK, tempted us to further explore the binding of larger tubercidins in semi-open form of human AK. Ligand induced semi-open form of binding site is also reported recently in case of HIV–RT [22,23].

In this report we have proposed for the first time that, like *T. gondii* AK, human adenosine kinase also possesses a semi-open/semi-close conformation to accommodate the larger analogues. The docking results indicate that the key ligand–protein interactions are same for all tubercidin inhibitors, irrespective of their binding to close or semi-open protein conformations, thereby preserving all crucial interactions for activity. We started this study by presuming that only aryl molecules can not bind in the adenosine binding site due to large aryl groups at C4 and C5 position but at the end we could find that it is not only the presence of aryl groups, instead, few spatial requirements force the molecules to bind in different ways in the active site. Presence of large hydrophobic aromatic group at C4 position of the ligands has been predicted to be the main discriminating factor for separating the aryl binders from the non-aryl binders.

Again by comparing the interactions of nucleosides with those of non-nucleosidic inhibitors, we have analyzed the importance of some interactions, which are conserved for all the molecules interacting with the enzyme and play important role in molecular recognition. For all kinds of inhibitors and protein conformations, three separate pharmacophore models for lead identification have been generated. The interpretation of binding interactions was carried out in the light of pharmacophore modeling data. Coherent result for diverse binding patterns of inhibitors was obtained by 2D descriptor based cluster analysis. We have also discussed why and how molecules choose the mode of binding and in what way the information could be utilized for designing new inhibitors. The pharmacophore models were validated and finally used as queries for 3D database mining.

Thus to gain more insight into structure–activity relationships we carried out this docking and pharmacophore analysis. More importantly, we have shown that information encoded in this approach can be effective to create virtual libraries for selectively targeting particular conformation of protein. Incorporation of this knowledge into virtual screening queries represents a much wider chemical space of molecules, which could be identified as potent AK inhibitors.

2. Methods

2.1. Preparation of proteins and ligands

Adenosine and tubercidin analogues are known to bind similarly in the same binding site [24]. Comparison of AK bound to 5-iodotubercidin (PDB code 2I6A) and to adenosine (PDB code 1BX4) shows that the iodine atom of 5-iodotubercidin in 2I6A replaces a conserved and highly stabilized water molecule, which is present in 1BX4, deeply buried into the pocket. Adenosine binding induces protein closing, and subsequent retention of two water molecules W1 and W2 (near C4 and C5 substituents, respectively) in the active site [20]. These waters were used to rationalize their effect on the role of C4 and C5 substituents, during the binding of tubercidins in close binding site. By examining number of X-ray structures of protein ligand complexes, displaceable and conserved waters can be identified [25]. Based on the visual inspection, which included assessing the alignment of the two crystal structures, only these two critical water molecules (W1 and W2) from 1BX4 were selected for docking studies (Fig. S1 in supplementary information).

The docking calculations were done using 1BX4 so that position of water molecules and their role in docking could be ascertained. ADO was taken out, hydrogens were added and correct protonation states were assigned and the protein was then minimized using CHARMM force field. To represent the open binding site of protein, a second conformation of human AK (PDB code 2I6B), complexed with an alkynylpyrimidine inhibitor [26], was prepared using the similar procedure as above without considering the waters.

A third conformation of protein was considered by modeling the protein using a semi-open *T. gondii* structure (PDB code 2A9Y) as template. It is known that though the sequences of the human and *T. gondii* AK are less than 28% identical, their overall global structures are similar and adenosine binding site is structurally conserved [27]. The structure of 2A9Y was used for sequence alignment of human AK (Fig. S2 in supplementary information), using FUGUE [28]. Initial homology model was prepared using MODELLER, which generates protein 3D structures by satisfying

spatial restraints imposed by sequence alignment with the template structure [29]. The quality of the homology model was assessed by calculating RMSD with the template, PROCHECK [30], and Profile 3D analysis. This is followed by manual docking of reference ligand (bound ligand in 2A9Y). Subsequent energy minimization including molecular dynamics using 'Discover' module and Cff91 force field within INSIGHTII [31] was performed to get the most stable complex structure. For energy minimization, a combination of steepest descent and conjugate gradient methods (100 steps each) was used with a convergence criterion of 0.001 kcal/mol. While carrying out energy minimization and molecular dynamics of complex, positional constraint was applied to backbone atoms and only side chains were kept flexible. Finally, the obtained protein model of human AK, representing a semi-open conformation, was utilized for docking (after removing the ligand).

All ligands were modeled, conformationally sampled by dynamic simulation using the annealing dynamics with user defined attributes: Reqtemp, 450 K; dynamic time step, 0.01 ps; steps, 10,000 using constant NVE (constant number of moles, volume, and energy). This was followed by energy minimization using UNIVERSAL 1.02 forcefield. The nucleosidic analogues (Table 1), synthesized and tested on human recombinant AK enzyme for inhibitory activity, were taken from literature [12,13].

2.2. Docking

GOLD is an automated ligand-docking program that uses a genetic algorithm to explore the full range of ligand conformational flexibility and partial flexibility of the protein [32]. Genetic operation settings (crossover, migration, mutation) were set to auto, with 100% search efficiency, where GOLD can automatically calculate an optimal number of operations for a given ligand. The parameters were set as: hydrogen bonding (4.0 Å), van der Waals (2.5 Å). The position of the active site was introduced by selecting all protein atoms within 7 Å of bound ligand atoms. GOLD allows to switch specific water molecules on or off (i.e. a particular water should be present or absent in the protein) or it can automatically determine whether a specific water should be bound or displaced by toggling (On/Off) during the docking run. A constant penalty is added for each water molecule that is switched on and represents the loss of rigid-body entropy on binding to the target, hence water displacement is rewarded [33]. Here closed binding site was considered with two water molecules, W1 and W2, in on and off respectively (I) and toggle (off/on) (II) modes. Clustering option was chosen to check docked solutions for similarity in the poses in terms of their RMSD (clustering distance 0.75 Å). All the docking solutions were analyzed and the best docked conformations were then used for further studies. GOLD was used along with two scoring functions, Modified Chemscore for kinase systems [34], and Goldscore. Scores were evaluated using two-step cluster analysis in SPSS [35] and by correlating with pIC_{50} (mM).

2.3. Generation of pharmacophore models

HypoGen, a module of Catalyst, was chosen for pharmacophore elucidation. The detailed description of algorithm and analysis of cost components is discussed elsewhere [36,37]. The quality of the generated pharmacophore hypotheses were evaluated by considering the three cost functions; fixed cost, null cost and the total hypothesis cost. The difference between total cost of the generated hypothesis and the null hypothesis should be as large as possible; a value of 40–60 bits difference indicates the probability of 75–90% chance of a true correlation in the data set [38]. The total cost of any hypothesis should be close to the value of fixed cost for a significant model. Correlation coefficient is derived from regression of the actual and estimated activity (by geometrical fit to hypothesis). The conformational models for all molecules were generated using the quasi-exhaustive/ConFirm module, with the "best quality" conformational search option. A maximum of 250 conformations were generated using CHARMM force field parameters [36] and a constraint of 20 kcal/mol default settings. Based on the two binding modes for nucleosidic analogues, two separate pharmacophore models were constructed. These two sets of nucleosidic analogues were further divided into training and test sets in such a way to cover full range of structural variation and IC_{50} values. A set of 10 pharmacophore hypotheses was generated for each group of compounds. The spacing parameter was set to 1 pm. Variable weights and variable tolerance options were given as it was found that all the features were not of equal importance. Additionally, for the first hypothesis generation, excluded volumes (EV) were added to improve the hypothesis. CatScramble [36], a Fischer randomization test, was performed at 95% confidence level, to assess the statistical significance of the pharmacophore hypotheses.

2.4. Structure based pharmacophore modeling

A third pharmacophore for non-nucleosidic compounds was modeled considering the conserved interactions and using the structure-based approach. The features of the pharmacophore were placed manually, guided by the X-ray structure of the open conformation of the receptor bound to an alkynylpyrimidine compound (PDB code 2I6B). Validation of pharmacophore was done by mapping the non-nucleosidic (NN) compounds [39] with known activity.

2.5. Database searching for virtual screening

The pharmacophore hypotheses were used as query for screening Maybridge database using fast flexible search. The hit list contained those compounds of the database that fulfill the requirements of any one of the three pharmacophores hypotheses. Hits were further analyzed for identification of new important scaffolds.

2.6. Cluster analysis

Cluster analysis is a method for dividing a set of molecules into subsets, where molecules in a subset share a certain degree of similarity and differ from other subset molecules. For clustering a set of 2D topological descriptors and BCUT descriptors were calculated, using *Cerius*^{2} (4.10) [40] and *MOE* [41], respectively. The usefulness of topological descriptors and BCUT descriptors in QSAR and QSPR studies has been extensively demonstrated and these descriptors have been suggested as a measure of structural similarity or diversity [42,43]. Since the descriptors were of non-uniform dimensions, PCA was performed. Principle components were then subjected to hierarchical cluster analysis using (HCA)–Wards method in *Cerius*^{2} 4.10.

3. Results and discussion

3.1. Overall docking results, binding mode prediction and analysis of scores

To assess the ability of docking strategy, ADO was initially docked as a control and the result was compared with the crystal structure of the enzyme inhibitor complex (1BX4). A docked structure of ADO, the natural substrate, in the protein (Fig. 1a) reveals all interactions of bound mode in 1BX4, with RMSD less than 0.5 Å, in all the four docking runs. The validity of GOLD

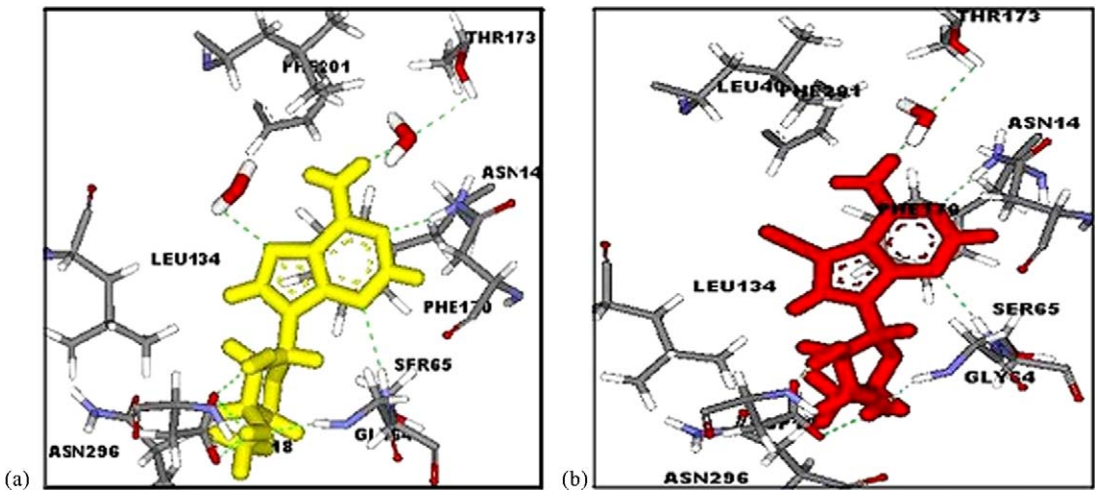


Fig. 1. Docked poses of ADO (a) and 5-iodotubercidin (b) in the closed binding site.

protocol for the water system was assessed by docking 5-iodotubercidin (compound **2** in Table 1) into protein (1BX4 after removing ADO) and comparison with 2I6A. It also showed that GOLD is able to dock this molecule in respective position by preserving the key interactions and correct water occupancy (Fig. 1b).

The binding modes of all the ligands were evaluated by comparing the docked positions with the crystallographic structure of bound ligand. Two distinct orientations of tubercidins were identified while docking within the binding site. The first group consisted of non-aryl derivatives, possessing the same orientation of tubercidin nucleus as that of ADO in 1BX4, and second class with aryl analogues possessing nucleus in almost perpendicular orientation to that of ADO. In particular, the pyrimidine ring is in parallel orientation relative to the amino acid Phe170 in all non-aryl ligands, whereas in case of aryl analogues, the aryl substituent is making stacking interaction with this amino acid (Fig. 2).

Analysis of all the docked conformations revealed that whether it is Goldscore or Chemscore based pose selection methodology or presence or absence of water molecules, the non-aryl analogues bind in the same mode as ADO or 5-iodotubercidin. However, the aryl analogues could not produce the pose similar to bound conformation of ADO, though reflected similar binding pattern within them. The consensus coming of one binding mode for all aryl analogues may be a clear indication of their binding pattern different from that of non-aryl analogues. Analysis revealed that it is not the presence of hydrophobic aryl substituent at C4 and/or C5 position which restricts the molecules to accommodate in closed binding site, instead, only the C4 position is determining the binding orientation. This can be illustrated by examining the differences at C4 and C5 substituents in **46** and **47** where NH_2 and phenyl groups are interchanged. Compound **47**, bearing NH_2 group at C4 position and a phenyl group at C5 position, is always coming in the first cluster of non-aryl analogues whereas **46**, with reverse arrangement of these groups, is coming in second cluster of aryl analogues. Analysis of scores shows that aryl analogues have lesser scores (Goldscore and Chemscore) as compared to the non-aryl analogues though the aryl analogues were having high or comparable activity (Table S1 in supplementary information).

The result of two-step cluster analysis with SPSS suggests that scores are able to differentiate between the actives and inactives (pIC_{50} (mM) < 3.93) for first group compounds of non-aryl analogues (Fig. 3). The analysis reveals that Goldscore is better in on-off condition, which could successfully discriminate between active and inactive non-aryl molecules, with the exception of few molecules (Table S2 in supplementary information). Results also suggest that Chemscore is able to characterize the highly active molecules more efficiently than those with lower activity.

Presence of water is not significantly affecting the respective binding orientations of aryl or non-aryl analogues, whether it is toggle or on-off mode. It corroborates with earlier studies which showed that GOLD is sufficiently able to predict the correct binding

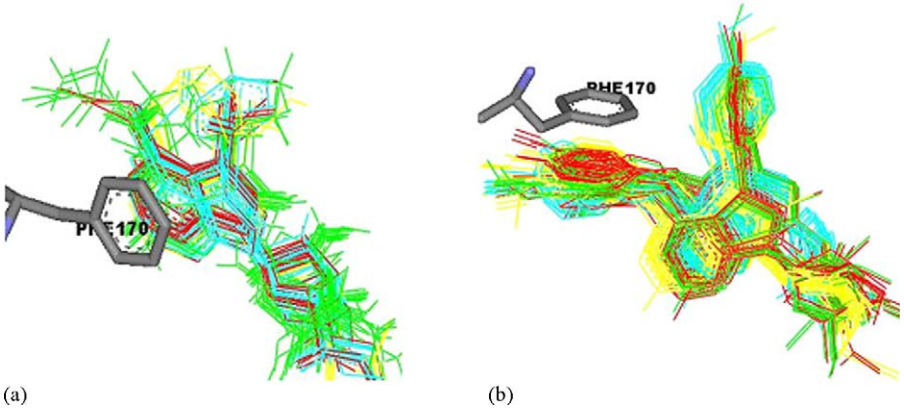


Fig. 2. Orientation of non-aryl (a) and aryl analogues (b) in the closed protein conformation with reference to PHE170, represented by GG on-off: (yellow), GG toggle (red), GC on-off: (cyan), GC toggle (green).

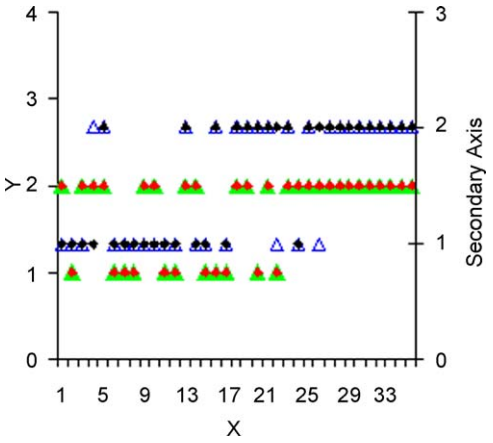


Fig. 3. *X*-axis: Compounds arranged according to increase in activity; *Y* and secondary axes represent clusters, 2: (pIC_{50} (mM) $\geq$ 3.93 as active compounds), 1: (pIC_{50} (mM) $<$ 3.93 as inactive compounds). Series (◆) GC On-Off, (▲) GC toggle, (◆) GG On-Off, (▲) GG toggle.

modes in the presence or in the situation when waters are not considered for docking [44]. Examination of the docked structures in closed binding site revealed that when waters were toggled, all non-aryl molecules (except **29**) were replacing the W2 water molecule (Table 2). Ability to displace water could be effective in assessing the activity, as the molecules, which could not displace water, will certainly have lesser score as compared to the one displacing the water. For example **29,** which could not displace any water in GG and GC methods (Table 2), is inactive and also having lesser score. Behavior of few molecules was found to be anomalous, e.g. **13**, having SCH_3 group at C5 position, displaces W2 and possesses high scores, still less active. Comparison of two molecules, **2** and **36**, showed that only substituting iodo group with methyl at C5 position is responsible for lower activity of **36**, although both have similar Chemscore and a small difference in Goldscore (Table S2 in supplementary information). At C5 position, bulk substitution is required for interaction. However, along with the bulk, electronic attributes also play important role. It seems that iodine at this position provides an adequate balance between bulk and electronic properties for interaction. Although the scores do not show very diverse figures as their divergent activity values, GOLD docking was found to be very informative regarding the understanding of interactions of inhibitors in AK binding site.

In case of aryl analogues, docking studies showed that whether water is displaced or not, the molecules were not able to bind as ADO. Hence, it was realized that these aryl molecules must possess some other orientation in the binding site. Aryl molecules in the closed binding site possess clashes or bumps with the protein residues, signifying that they are less likely to correspond to physiologically relevant ligand binding position (Fig. 4a). AK, like other related enzymes, shows considerable degree of flexibility as represented by their open and closed conformations [20]. It is plausible that if the clashes could be overcome by flexibility in the system and favourable contacts increase, it would be beneficial for the molecules. So, finally, we disregarded the possibility of the aryl tubercidins binding to the closed binding site and carried out the search for other alternatives. It is also known that considering single receptor conformation in a flexible system during docking experiments can lead to errors in identification of binding modes and no clear relationship between docking and scoring could be found [45]. Further, cross-docking studies of alkynylpyrimidine inhibitor into close form and aryl compounds into open form of the AK were performed. In ADO binding site (toggled waters) alkynylpyrimidine inhibitor showed similar interactions as those of aryl molecules, with slight displacement of pyrimidine ring and aryl groups (Fig. 4a). Certainly the interactions are not conceivable, as this molecule is known to bind in open form of the protein. This experiment gave us a clue to think that if these aryl molecules are allowed to bind in relaxed protein conformation, more reasonable binding orientation could be obtained. The docking of aryl molecules into open form of the binding site suggested that the resulting interactions are hybrid of two bound ligands (1BX4 and 2I6B). The best GOLD solution suggested that the sugar part interacts with Asp18, while pyrimidine ring is not exactly parallel to Phe170, instead, it is displaced (Fig. 4b). Thus, according to GOLD docking, the docked solutions of aryl tubercidins emerged to be different from those of adenosine and alkynylpyrimidine inhibitors in the AK closed and open binding site, respectively.

In AK, initial anchoring of substrate or inhibitors to open binding pocket is facilitated by lining residues Leu40, Leu139, Phe170 and Phe201. After binding of inhibitors, these residues are brought into opposition to close the site between the two domains and sequester from bulk solvent. Ribokinase family uses sugar as substrate recognition and consequent closing of binding site that involves Glycine-Glycine switch [20]. It is known that human and *T. gondii* AK structures show 30 degrees of bending in closed form as compared to open conformation [20]. Comparison of apo and adenosine bound conformations revealed that dihedral angles between Gly68 and Gly69 in *T. gondii* AK and Gly63 and Gly64 in human AK undergo large torsional change, necessary for hinge bending motion [20]. This hinge bending also induces a semi-open conformation in *T. gondii* [21]. Based on these reports, we hypothesized that the aryl tubercidins, after binding, must start

Table 2
Cluster representation and water occupancy[a], suggested by GOLD.

Pyrimidine ring making pi–pi interaction with Phe170	Group 1	1–37 (excluding 24, 25, 26) and 47
N-substituted aryl ring forming pi–pi interaction with Phe170	Group 2	38–71 (excluding 47) and 24, 25, 26

	GS (Toggle)	
	Group 1	Group 2
W1 off–W2 on	None	None
W1 on–W2 off	1–19, 21, 27, 28, 30–37	25, 26, 43, 46–50, 52–55, 56, 57, 60, 61, 63–71
W1 off–W2 off	20, 22, 23	51, 58, 59, 45
W1 on–W2 on	29	24, 38, 39, 40–42, 44, 62

	CS (Toggle)	
	Group 1	Group 2
W1 off–W2 on	None	None
W1 on–W2 off	1–22, 27, 28, 30–37	25, 38–48, 50–56, 58, 60–71
W1 off–W2 off	23	24, 26
W1 on–W2 on	29	49, 57, 59

[a] On: water could not be displaced, Off: water displaced.

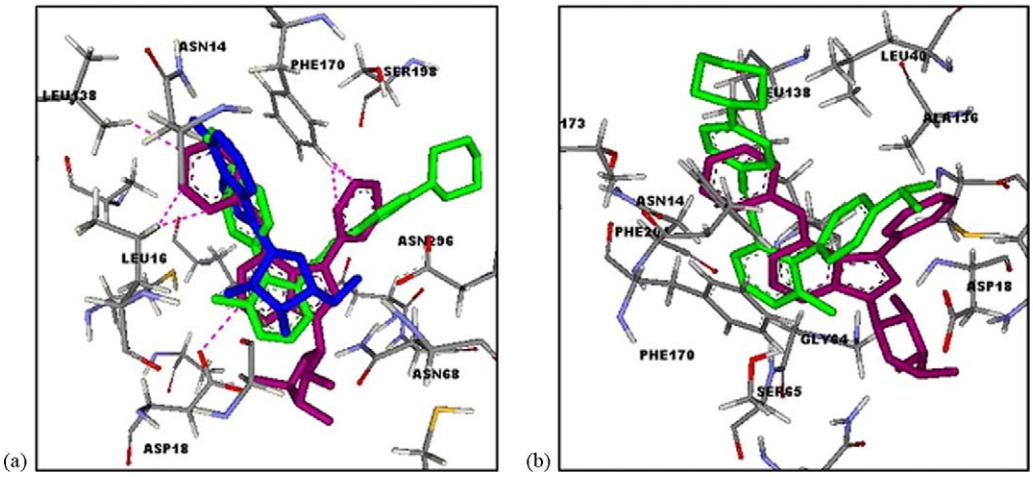


Fig. 4. (a) Docking alignment of adenosine (blue), compound **48**, an aryl inhibitor (magenta, purple lines represent the bumps) and alkynylpyrimidine inhibitor (green) in close binding site. (b) Docked pose of compound **48** and alkynylpyrimidine inhibitor in open binding site.

closing of binding site involving GG switch. However, instead of complete closing, it would stop at some conformation in between the closed and the open forms (as in *T. gondii* [21]), which is highly favourable for the ligands.

Ligand binding information can be gleaned from either exhaustive MD or by examining multiple conformational states of protein, which are already known to exist. The conformation, which is already known, gives more essence as it represents a minimum energy complex already existing in nature [46]. This second alternative was considered in further docking studies using the homology modeled semi-open conformation of protein. The root mean square deviation of the Ca atoms (Ca RMSD) between the modeled protein and template was 0.8 Å. The homology model was found to be valid in terms of phi-psi occupancy as represented by Ramachandran plot (Fig. S3 in supplementary information) showing 87% of residues occupying the most favoured allowed region and 8.4% in the additional allowed region. Only two residues were found in disallowed region, however, none of these residues were part of the binding site. Moreover, out of these two residues, one (SER198) is also reported to be in the disallowed region of the Ramachandran plot of the template (2A9Y) and the other residue is far from the active site (Fig. S4 in supplementary information). The initial homology model was having a side chain (Phe170) deviated from the crystal structure (ligand bound semi-open AK of *T. gondii*) (Fig. 5), which could be on account of ligand binding. It is known from earlier studies that ligand based minimization of the homology model could give the correct definition of binding site region [47]. With this rationale, the modeled protein was minimized in the presence of reference ligand. Profile-3D, the self-compatibility score, for this protein is 130.71 which is much higher than the lowest score 67.00 and close to the top score 148.90. Compatibility scores of residues above zero reflect acceptable side chain environment. Most of the residues possess reasonable compatibility scores (Fig. S5 in supplementary information). Only few residues 251, 277–281 and 284–287 are built poorly in homology model. Fortunately, these residues are located far from the active site of AK and hence these residues would have no influence on the following study (Fig. S4 in supplementary information).

The docking solutions obtained with the GS and CS methods for aryl molecules were qualitatively similar with respect to interactions in semi-open binding site. Binding of tubercidins seems to be stabilized by planar adenine ring, which maintains the alignment of molecules in the groove. This alignment is supported by sugar moiety, with both O2′ and O3′ of sugar moiety involved in salt bridge interaction with the side-chain oxygen atoms of Asp18 (Fig. 6).

It is striking that despite the structural variability in ligands and modified protein conformations, the important interactions were found to be conserved. Sugar moiety is binding exactly in the manner as that of adenosine or non-aryl inhibitors while pyrimidine ring is forming interaction with Phe170, Phe210, Leu40 and Leu138 amino acids as in non-aryl or alkynylpyrimidine inhibitors. An inspection of the score spectra indicated that the Chemscore scoring function discriminated between active and inactive compounds (Fig. S6 in supplementary information). The diverse biological activities of the compounds and Chemscore values were found to be correlated ($r = 0.62$) with few major outliers as **49** and **59**, removal of which led to improved correlation ($r = 0.74$). The compounds differ mainly in hydrophobic portion while the H-bonding terms are mostly constant throughout the group. As increase in activity is associated with increase in

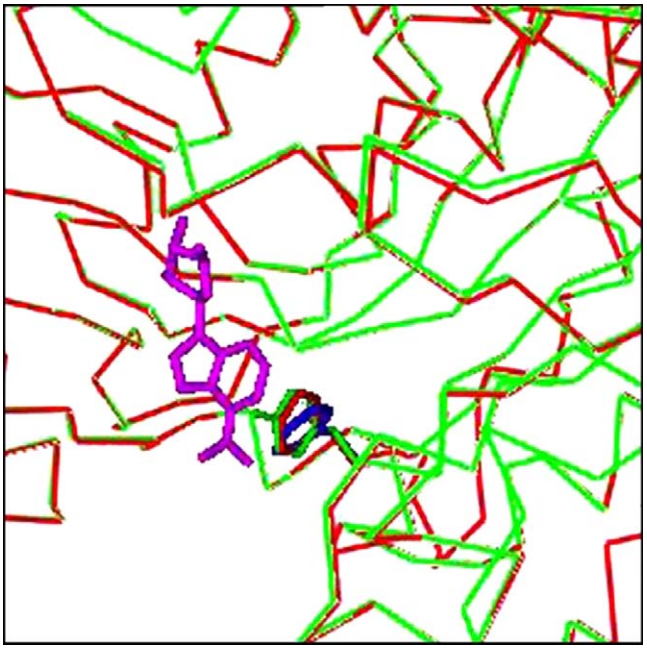


Fig. 5. Homology model of human AK in semi-open form after minimization (red); template 2A9Y (green); Phe170 side-chain before minimization (blue); bound ligand (magenta). Only binding site is shown for clarity.

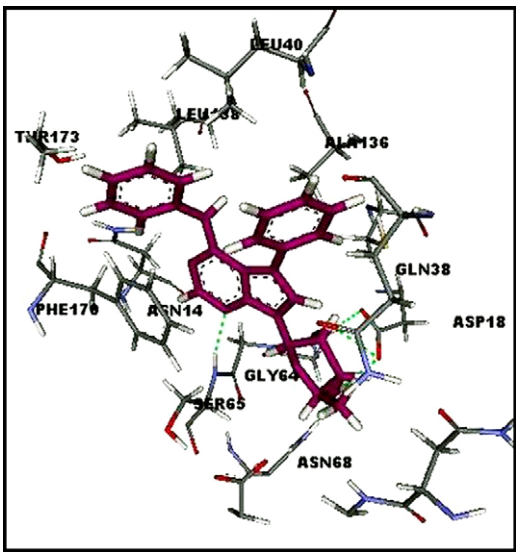


Fig. 6. Docked pose of the most active aryl molecule **48** by Chemscore in modeled semi-open binding site; green dotted lines represent hydrogen bonds.

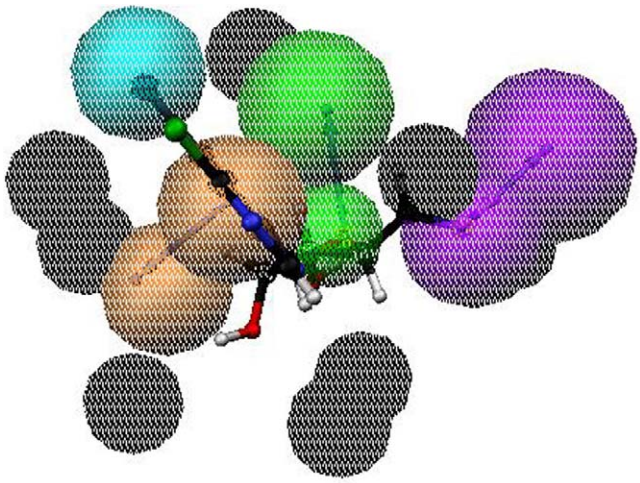


Fig. 7. Pharmacophore mapping of compound **30** with Hypo1. Pharmacophore features: for hydrogen bond donors (HBD), purple; ring aromatic (RA), brown; hydrophobic aromatic (HA), green; hydrophobic, cyan; excluded volume, black.

lipophilic character, activity dependence on hydrophobicity is not reflected in Goldscore, which gives more weightage to H-bonding.

It is justified from the docking results that the semi-open conformation, which the binding site acquires by a smaller degree of movement, is sufficient to accommodate aryl compounds. It is apparent that, like *T. gondii* AK, human adenosine kinase also possesses a semi-open or semi-close conformation to accommodate the larger analogues. Based on the above results, we inferred that tubercidins, depending on their size, bind to two binding conformations of AK instead of a single. These two groups of tubercidin inhibitors were separately utilized for ligand based pharmacophore modeling.

3.2. Pharmacophoric elements

3.2.1. Pharmacophoric elements for closed and semi-open state binders

For the non-aryl compounds, care was taken to choose the maximum diversity points in the training set. Presence of compounds **36**, **6** and **22** in the training set could not give good hypothesis in terms of prediction and statistical significance. So these compounds were not considered in final hypothesis generation run. Finally 16 compounds were selected (Table 3) based on visual inspection of their structural and activity profile. Initially, Hypogen runs were performed using default chemical features with excluded volume but the addition of excluded volume has improved the prediction and significance. To understand the degree of significance of various features, variable weights and tolerances were considered in pharmacophore modeling. The first ranked hypothesis (Hypo-1) showed a good correlation ($r = 0.982$) between actual and estimated activity of training set (Table 3). Hypo-1 contains four features: one hydrophobic (**H**), one ring aromatic (**RA**), one hydrogen bond acceptor (**A**) and one hydrogen bond donor (**D**). The mapping of the best training set compound **30** to Hypo-1 is shown in Fig. 7. The hypothesis was then tested against all test set molecules, which included **36, 6** and **22**. These three compounds were again not predicted well (Table S4 in supplementary information) and affected the predictivity of the hypothesis ($r = 0.47$). Removal of these outliers (**36**, **6** and **22**) gave a reasonably good regression correlation ($r = 0.71$).

The most likely explanation for their incorrect prediction could be ascertained by examination of structural features. Molecules have special preference for iodine over methyl or any other group at C5, as iodine replaces water and nicely fit into the active site by strong van der Waals interactions. At this position halogen substituents or other hydrophobic groups like methyl are represented by same hydrophobic feature by Catalyst. Presence of iodine instead of methyl group is not differentiable in terms of hydrophobic feature and this is the reason for incorrect prediction of **36**. By looking into the prediction of **6** (one of the very active compounds), the lower activity prediction may be attributed to the presence of CH_3 group at C5′ position, which is not represented by the pharmacophoric feature at that point, as the CH_3 group is also

Table 3
Information of statistical significance and predictive power presented in cost values[a] for top 10 hypotheses.

Hypothesis no.	Features, Ex.V[b]	Total cost	rms	Correlation (r)
1	A D H RA, 9	97.0675	0.697664	0.982802
2	A D H RA, 10	104.351	1.10554	0.957493
3	A D H RA, 2	114.465	1.40285	0.934668
4	A D H RA, 3	120.28	1.63726	0.908467
5	D D H RA, 1	132.807	2.10202	0.83518
6	A A D D, 1	132.815	2.17397	0.818942
7	A D D D, 2	133.852	2.19344	0.815659
8	D D H RA, 1	134.128	2.18345	0.819073
9	A A D D, –	136.949	2.20258	0.818181
10	A D D D, 2	136.962	2.3031	0.793462

[a] Fixed total cost = 92.5412, Total cost for null hypothesis = 164.342.
[b] Ex.V: excluded volumes. Training set compounds include **1**, **10**, **11**, **20**, **33**, **35**, **9**, **21**, **28**, **30**, **37**, **23**, **31**, **2**, **29**, **13** from group 1 (non-aryl analogues); rest of the group 1 compounds are included in test set.

Table 4
Information of statistical significance and predictive power presented in cost values[a] and clusters for top 10 hypotheses from group 2 (aryl compounds).

Hypo no.	Features	Total cost	RMS	Correlation (r) training[b] (test[c])	HCA clusters							
					2nd	3rd	4th	5th	6th	7th	8th	9th
1	A D Ha RA	101.769	1.09	0.952, (0.891)	1	1	1	1	1	1	1	1
2	A A Ha RA	102.703	1.13	0.947, (0.920)	2	3	4	5	6	7	8	9
3	D Ha Ha RA	103.253	1.15	0.946, (0.888)	1	2	2	3	4	5	6	6
4	D Ha Ha RA	103.39	1.16	0.945, (0.820)	1	2	2	3	4	4	5	5
5	D Ha RA RA	103.489	1.15	0.945, (0.891)	1	2	3	4	5	6	7	8
6	D Ha Ha RA	103.554	1.15	0.946, (0.885)	1	2	2	3	4	5	6	7
7	D Ha RA RA	103.52	1.15	0.946, (0.884)	1	2	3	4	5	6	7	8
8	A D Ha Ha	103.596	1.15	0.946, (0.862)	1	1	1	2	3	3	3	3
9	A D Ha RA	103.68	1.16	0.945, (0.881)	1	1	1	1	2	2	2	2
10	D Ha Ha RA	103.698	1.17	0.944, (0.831)	1	2	2	3	4	4	4	4

[a] Fixed total cost = 91.7576, Total cost for null hypothesis = 150.444.
[b] Training set compounds include **48**, **66**, **57**, **60**, **51**, **68**, **61**, **58**, **55**, **71**, **45**, **69**, **39**, **25**, **26**, **43** from group 2; rest of the group 2 compounds are included in test set.
[c] Training set compounds include **48**, **66**, **57**, **60**, **51**, **68**, **61**, **58**, **55**, **71**, **45**, **69**, **39**, **25**, **26**, **43** from group 2; rest of the group 2 compounds are included in test set.

associated with the compounds having lower activity (**21** and **23**). Careful analysis of the dataset suggested that the molecules which possess CH_3 group at C5′ position along with the groups other than NH_2/Cl at C4 position, are less active. However, CH_3 group gives advantageous effect to molecules in cases where NH_2/Cl group at C4 position is present (e.g. **32**, **34**) and the effect is still more pronounced by the presence of halogen at position C5 (e.g. **2**, **3**). This indicates that it is not the presence of CH_3 group at C5′, instead the absence of NH_2/Cl group (e.g. thio substituted groups in **21**, **23**) at C4 position might be associated with lower activity of the respective compounds. In terms of features, NH_2 and Cl cannot be represented by same feature in Catalyst and thus could not be selected by Hypogen algorithm in hypothesis generation run. Thus, nonappearance of this required feature is responsible for lower prediction of compound **6**. Incorrect prediction of **22** by this hypothesis is justified as it binds to semi-open conformation of AK (discussed in latter section).

For the second class, 16 aryl compounds were selected for hypothesis generation (Table 4). All the generated hypotheses were found to be minor variants of each other in terms of cost values. The best hypothesis consists of four features: one ring aromatic (RA), one hydrogen bond acceptor (HBA), one hydrogen bond donor (HBD) and one hydrophobic aromatic (HA). The IC_{50} values of all the compounds of training set and test set were predicted within the measured limits (r values shown in Table 4). So, instead of studying the first best hypothesis, we performed a cluster analysis by hierarchical average linkage method to categorize the obtained pharmacophoric models into clusters. At second cluster level, first and second hypothesis are distantly related, by the presence of HBD instead of HBA feature, mapped by 2′OH of the sugar moiety. This indicates that at this position a donor and acceptor must be present simultaneously and hence suggests the importance of an OH group, having both the functionalities. At 3rd cluster level Hypo1, 8 and 9 fall in same cluster and Hypo3, 4, 5, 6, 7 and 10 in another, illustrating the difference in ring aromatic mapped by pyridine ring or HBA feature mapped by ribose oxygen of tubercidins. The ribose oxygen is proposed to be important in orientating molecules in the active site [24]. This suggests that alignment of pyrimidine ring might facilitate the orientation of sugar with such conformation or vice versa. Other clustering steps suggest interchangeable positions of HA and RA. Hypo 8 and 9 have another characteristic feature of HBA mapped by nitrogen of pyrimidine ring (for interaction with Ser65 in docked mode), instead of RA representing pyrimidine ring or HBA feature representing ribose oxygen. Based on the result of cluster analysis, we have merged the important features and proposed a combined hypothesis (Hypo2) for virtual screening (Fig. 8a).

CatScramble results indicated that in both the cases, the best hypotheses were 95% significant as none of the outcome hypotheses had a cost score lower than the initial hypotheses (Table S3 in supplementary information). The pharmacophore models clearly highlighted that the proposed hypotheses are capable in explaining the SAR for nucleosidic analogues. Based on the detailed analysis of the activity predictions by Hypo1, it is apparent that there is a requirement of a halogen substituent (I/Br/Cl) at C5 position and

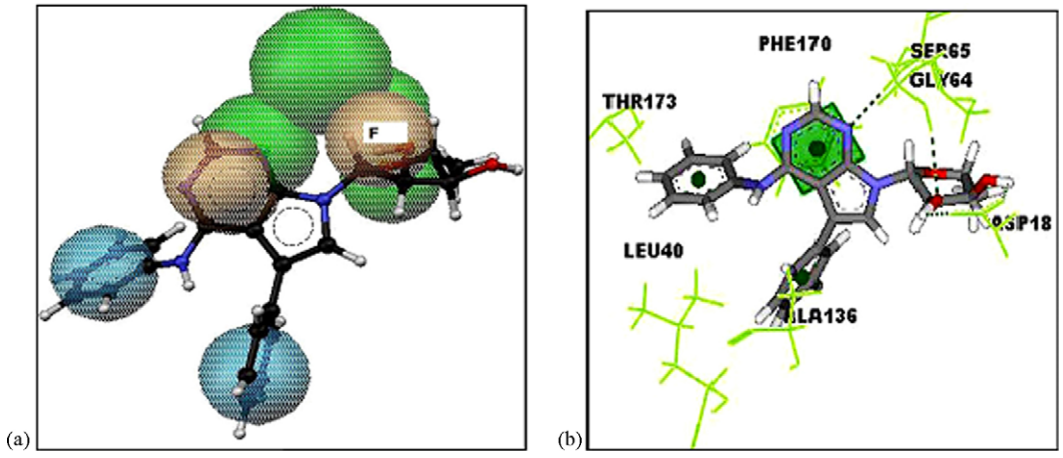


Fig. 8. (a) Hypo2, pharmacophore based on merging all important features. Pharmacophore features are color-coded (RA: brown; HA: blue, HBA: green), F: (brown, representing a HBD and HBA feature). (b) The most important interactions between aryl analogue (**48**) and active site residues; green dotted lines show the hydrogen bonds with key residues, stacking interaction is shown by green plane.

NH2/Cl substituent at C4 position. It seems that the corresponding positions in the binding site are most compatible with these functional groups. Hence, any hit with novel scaffold bearing these substituents, at respective positions, will be prospective inhibitor. Analysis of the aryl pharmacophore (Hypo2) shows that the model generally matches well with the active site pocket in the docked state (Fig. 8b). Aryl groups are placed in the clefts formed by residues Leu40, Ala136, Thr173 and Phe170. A hydroxyl group, which shows interaction with the amino acids Asp18 and Gly64, represents HBD and HBA, respectively. The central aromatic feature is making stacking interaction with Phe170. The presence of HBA feature, represented by ribosyl oxygen, is not directly representing the acceptor feature in the active site. It is also reported that the nucleocidic ribose oxygen is having importance in maintaining the ring structure for proper orientation of molecules in AK binding site [24], which signifies the appearance of this pharmacophoric feature.

3.2.2. Pharmacophoric elements for open state binders

A structure based pharmacophore has been derived by using the open form of AK complexed with alkynyl inhibitor (2I6b). The ligand in the crystal structure showed same aromatic interaction with Phe170 and anchoring with Ser65, which was detected by Hypo2 for aryl molecules. Comparison of pharmacophoric features for aryl molecules and interactions of bound/docked non-nucleosidic inhibitor suggests that an aromatic ring and an anchoring point are consistently required. This indicates that these two interactions represent a conserved scaffold interaction in case of aryl, non-aryl and non-nucleoside inhibitors. Also, binding site with hydrophobic residues poses the requirement for hydrophobic bulk flanking the central ring in all inhibitors. The confined bulk is necessary as spread in the bulk might destabilize the narrow groove of the binding site. Thus, like aryl molecules, the aromatic feature seems to be important for open state binders too. Hydrophobic aromatic feature represents hydrophobic bulk, which is not largely distributed, thus, represents a specific filter than a general hydrophobic feature. Hence, such constraint could be useful to reveal more probability of finding true positives in VS. Based on this analysis, the pharmacophore (Hypo3) for open state binders was derived by manually putting, an anchoring HBA which stabilizes an aromatic ring, flanked by two hydrophobic aromatic features, for packing into the hydrophobic groove (Fig. 9a). It is to be noted that in structure based pharmacophore generation, the intrinsic systematic errors of the crystal structure may bleed over into pharmacophore. However, it may not seriously affect the precision of the pharmacophoric approach as some tolerance is allowed in the distances between pharmacophoric features in 3D space. The validity of this pharmacophore model was assessed by mapping and correlating the fit value with inhibitory activity (log IC_{50} (nM)) of the non-nucleoside (NN) analogs (Table S5, supplementary information). This simple four feature hypothesis is able to describe the major structure activity pattern (activity difference up to 4 log units) within the non-nucleoside inhibitors (Fig. 9b).

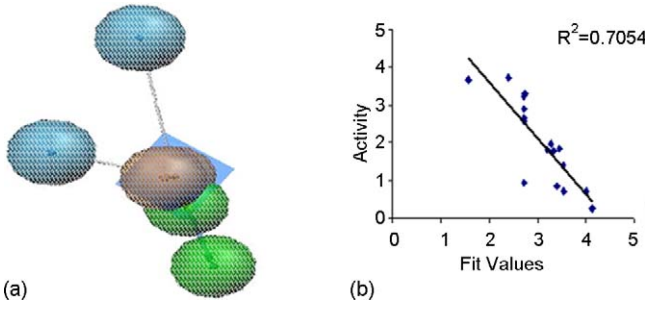


Fig. 9. (a) Structure based pharmacophore (Hypo3) derived for open form of the enzyme. (b) Graph representing the correlation between the fit values and activity (binding affinity in log IC_{50} [nM]).

3.3. Clustering

All nucleoside and non-nucleoside molecules could be divided into three clusters, based on their 2D similarity, using HCA-Ward clustering as shown in 3D PCA plot (Fig. 10). From the clustering results it was apparent that molecules possess at least three types of structural pattern and they are not distributed evenly in the space. Instead, a clear demarcation can be drawn which suggests that they may not bind in similar fashion and cannot be predicted by a single universal hypothesis. Only 4, out of 88 compounds, were not placed in their respective clusters (Fig. 10). This clustering method was found to be quite efficient to classify compounds under study in three groups according to their binding and stabilization to different final conformations of the receptor.

3.4. Overall binding modes and structure activity interpretations

From the docking study we could infer that the small cavity in the AK binding site that has been known to accommodate the small non-aryl derivatives is not sufficient for aryl analogues. Among nucleosidic inhibitors, the aryl ligands in general possess higher activities and activity increases with increase in aryl portion. This study indirectly gives an important clue that semi-open binding site should be considered for designing more potent compounds. It has already been speculated that binding through aromatic hydrophobic interactions may change the orientation of aryl AKIs relative to that of the natural substrate adenosine/non-aryl AKIs [7]. The importance of hydrophobic interactions could be ascertained by the hydrophobic nature of most of the lining residues. Engulfing of hydrophobic ligands is beneficial for activity, as also evident from the pharmacophoric elucidations. The tubercidin molecules with two phenyl substitutions were highly potent AKIs (e.g., **48**, IC_{50} = 0.0005 μM). Replacement of the phenyl

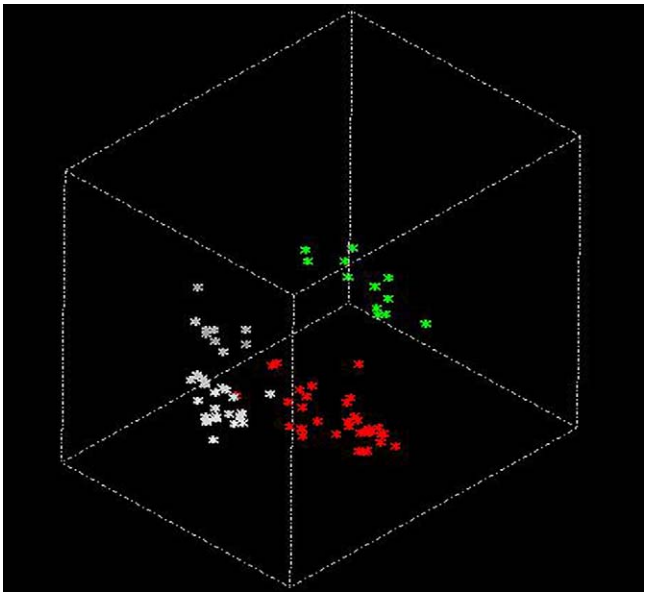


Fig. 10. Clusters obtained for the non-nucleoside and nucleoside compounds using HCA-WARD method.

Cluster number	Cluster member
Cluster-1 (White)	Group1, 24, and NN14
Cluster-2 (Red)	Group2 (except 24), NN10, NN18
Cluster-3 (Green)	Non-nucleosides (NN) (except NN10, NN14, NN18)

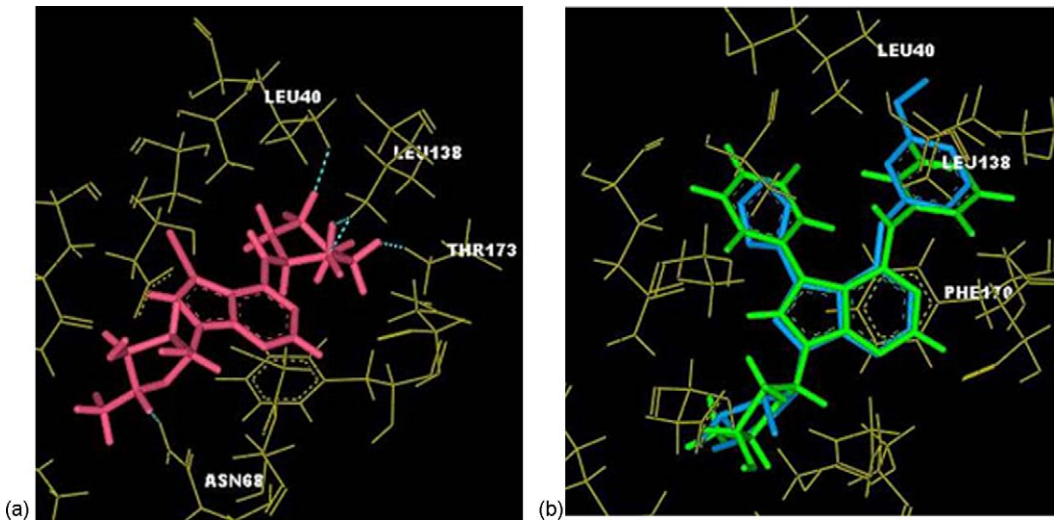


Fig. 11. Docked poses of (a) **43** (blue lines represent the bumps with the protein), (b) **59** (green) and **58** (blue) in semi-open conformation of human AK.

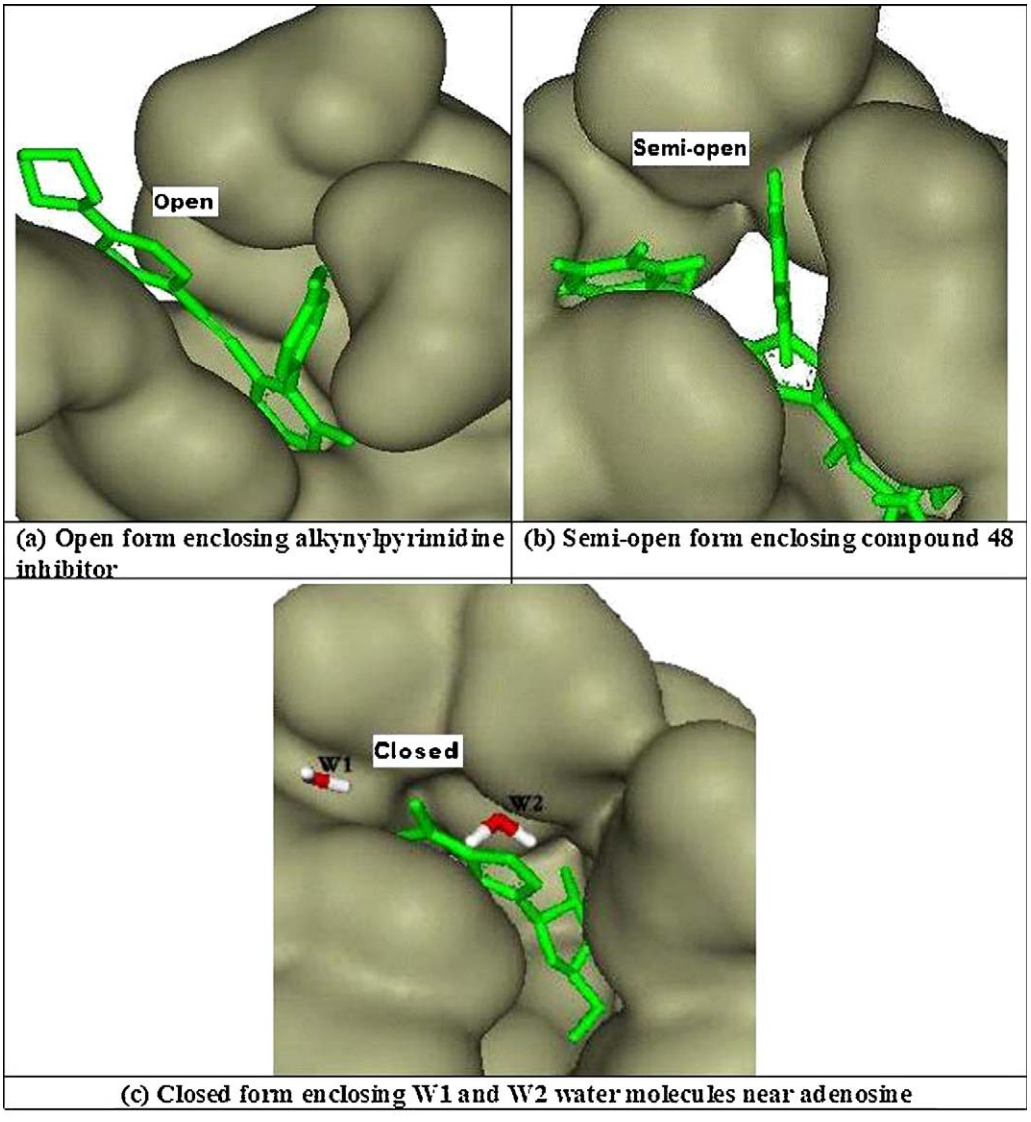


Fig. 12. Degree of closing in human adenosine kinase (a) open form (b) semi-open form (c) closed form.

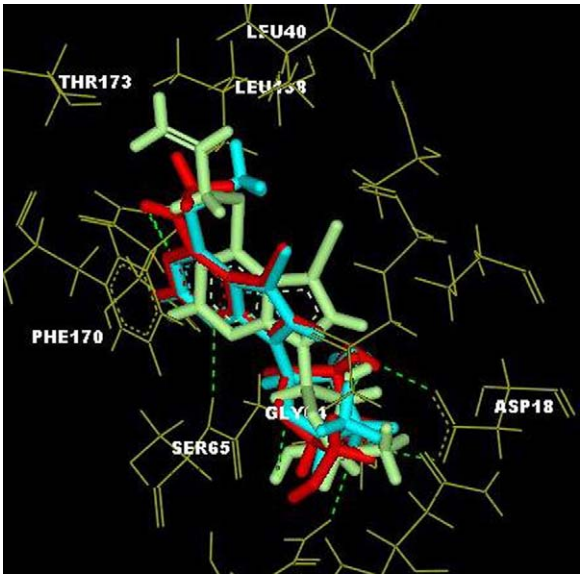
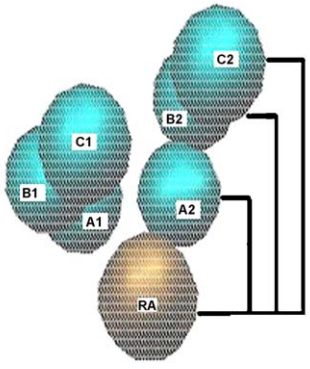


Fig. 13. Docked poses of compound **22** (cyan), **20** (red) and **23** (light green) in semi-open conformation of human AK. Green lines represent H bonding interactions of **22**.

group on the C4-amino with cyclohexyl in case of **43** showed a significant loss in potency, suggesting that the shallow groove of the active site laid the restriction for molecular bulk to be confined or less spread (Fig. 11a). Loss in potency of 3-pyridyl bearing compound **59** is due to the ring nitrogen, which makes it less hydrophobic (Fig. 11b). Similarly in other compounds for example **50** and **51**, the presence of polar side chains attached to aryl part are responsible for their lower activity as compared to unsubstituted aryl groups. In general, extensions to the aryl groups (as in **44**, **55**, **58**) is associated with the lower activity as compared to unsubstituted aryl groups, because of steric crowding of amino acids Leu138 and Leu40 (Fig. 11b).

Keeping the fact in mind that interaction of ribose moiety is same for all tubercidins, the arrangement of other groups decide which conformation of protein is involved. Every molecule with its sugar portion binds and GG switch gets activated, causing sequential closing of binding site. Thus, there is equilibrium of various conformations of protein. The size and shape of the molecules, together with the presence of groups capable for interactions with the waters, are important factors for ligand activity, and presence of larger group at C4 position is important for their placement in the close or semi-open forms. Binding involves similar interactions in all the tubercidins, whether aryl or non-aryl. On the basis of above observations we propose that, like *T. gondii* AK, human adenosine kinase also possesses a semi-open or semi-close conformation, not yet known, to accommodate the larger analogues. Only one site (Fig. 12), delineated by different degrees of closing, is found as crucial in the binding mechanism of human AK inhibitors.

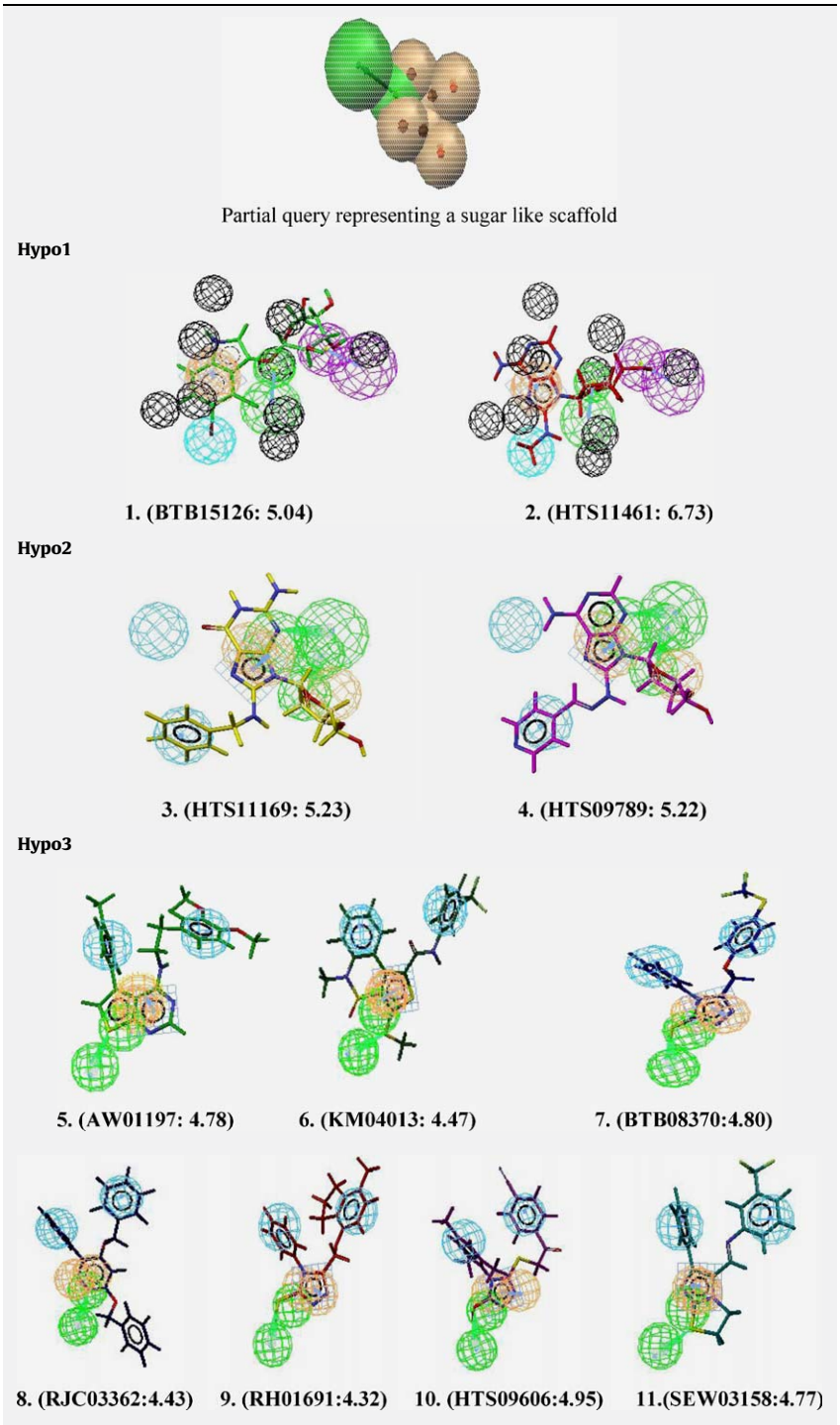
Activity of inhibitor depends on how the system stabilizes in any of those conformations. If C4 position is having a single heavy atom, closure of binding site will occur. A relaxed conformation would be realized if more than two heavy atoms at C4 substituent are present. **22**, containing two heavy atoms at C4 amino, is similar to the corresponding bound ligand from 2A9Y (T. *gondii* semi-open AK), suggesting its possibility of binding in semi-open conformation of human AK. Further, docking of **22** in semi-open conformation of human AK showed interactions comparable to those of larger aryl analogues (Fig. 13). Its incorrect prediction by Hypo 1 gives strength to this possibility. More than one atom will cause the lower probability of closed complex formation. **20** is also having two heavy atoms, but is less active. The reason may be attributed to weaker hydrogen bonding in the closed form as compared to unsubstituted C4 amino analogues (e.g. compound **1**) and steric crowding of leucine residues around the methyl group. **20**, **22** and **23** are clustered into the first group by docking, as GOLD uses soft van der Waals potential. **23**, when docked into the closed binding site, can be accommodated showing high score only in the absence of water (W1), when allowed to toggle. Its lesser score in presence of water (Table S1 in supplementary information) is in accordance with the observed activity, suggesting that the molecule cannot remove the highly stable water. Further, binding of **20** and **23** to semi-open form of the receptor were also explored. It was found that in **20**, the methyl group substitution at amino of C4 was inadequate for interaction, whereas, in case of **23** there is increase in number of heavy atoms at C4 position along with the conformational extended groups as compared to other larger analogues which prevents it to orient properly in the active site (Fig. 13). It may be inferred that **20** and **23** are not able to interact properly to stabilize either closed or semi-open form of the protein ligand complex, hence possessing less activity.



Scheme	Preferential Groups (1,2)	Fit values with mappings to three schemes		
		Alkynyl-pyrimidine	48	5-Iodo-tubercidin
RA,C1, C2	C1, C2: Hydrophobic aromatic	4.22	3.5	2.7
RA,B1, B2	B1, B2: Hydrophobic aromatic	2.52	2.998	1.74
RA,A1, A2	A1: hydrophobic/ polar,A2: hydrophobic	2.82	2.86	3.10

Fig. 14. Representation of three featured scheme for open, semi-open and close forms; mapping results of alkynylpyrimidine, **48** and 5-iodotubercidin with these pharmacophoric schemes; RA: ring aromatic, A1 and A2: for close, B1 and B2: for semi-open, C1 and C2: for open.

Table 5
The sugar like scaffold and mapping of the hits obtained from virtual screening with fit values.



In view of the above, the ability of binding of inhibitors is dependent on hinge-based binding site end points and most importantly presence or absence of sugar scaffold to reach an equilibrium state (closed, semi-open or open), leading to a stable complex. Increased size of the substituent at C4 position guides the molecule to bind in much relaxed conformation. Tubercidin molecules represent a rule based preference at C4 position for binding as (1) if having one heavy atom (hydrophobic or polar), very active, (2) if having more than two heavy atoms, there can be two implications, stabilizing if bulky but less flexible, otherwise less active. In general, presence of anchoring sugar with nonplanar hydrophobic groups, distant from central pyrimidine ring, will destabilize the system. Absence of sugar with less distant hydrophobic groups from central pyrimidine ring, will lead to lesser or no activity.

The three phamacophores represent a recurrent scheme as shown in Fig. 14. Molecules have preference for their respective pharmacophore and can not be predicted correctly with other pharmacophores. To illustrate this, a comparative evaluation of mapping of 5-iodotubercidin, alkynylpyrimidine inhibitor and

compound **48** with the three features in the pharmacophores of these three classes, has been performed. Molecules mapped in the best way with their respective pharmacophores as indicated by fit values (Fig. 14). Based on this study it was inferred that three separate pharamcophores are required for VS of AK inhibitors.

3.5. Virtual screening: utility of three pharmacophore hypothesis

The above three validated pharmacophores were used for virtual screening of novel inhibitors. It is necessary to include the sugar-like scaffold for virtual screening of closed and semi-open state binders, as GG switch is primary requirement for enzyme closing. To increase the probability of finding hits for close and semi-open conformations of AK, initially a partial query representing a sugar like scaffold was created. This query consisted of carbon backbone atoms, two oxygen atoms bearing single hydrogen, and a HBA feature, merged with positional constraints (Table 5).

The preliminary search from this query resulted in 136 hits, bearing anchoring moiety. Next, both the pharmacophores, Hypo1 and Hypo2 were used to screen and evaluate those 136 molecules based on their plausible binding to the close and semi-open binding site, respectively. In this virtual screening procedure these pharmacophore models were revealed to be too restrictive or selective and representative for nucleoside derivatives. Compounds, which have given highest fit values, belong to the category of nucleosides only, with other category compounds possessing low fit values (e.g. hit 1 in Table 5). The first hypothesis could retrieve the adenosine molecule (substrate for the enzyme). This again suggests that this scaffold is highly specific and achieving better binding for compounds other than nucleoside seems difficult. The top most compounds, obtained using Hypo1 and Hypo2, belong to the already known scaffold. Hence, it was realized that if the open conformation of protein is targeted, the limitation of requirement for a sugar like moiety could be overcome. This shows the importance of Hypo3 in virtual screening for open state binders. The potential hits obtained by virtual screening, using Hypo1, 2, and 3 are shown in Table 5.

From the aforesaid analysis of binding of nucleosidic inhibitors, we could get insights about the requirements for non-nucleoside type of inhibitors too. This is due to the fact that some interactions are conserved irrespective of the protein conformation. As there is no unique answer to this complex problem of ligand binding and stabilization of the complex in flexible site of AK, multiple pharmacophores generated from different subsets could represent an appropriate solution that should be pursued in the future.

4. Conclusions

The purpose of this study was (1) to know the binding modes within the tubercidin analogues, (2) generate pharmacophores for validation of different binding modes for tubercidins and utilization of those pharmacophores as search tools to identify potent chemical entities. This work illustrates that the molecules of same chemical series, having different degrees of substitutions, bind to different protein conformations, showing similar interactions. As far our knowledge, this is the first report for tubercidins binding to semi-open conformation of human AK, based on modeling studies. In docking studies Goldscore appears to perform better for non-aryl analogues and the Chemscore function performs better for aryl analogues. Comparison of interactions suggest that aryl molecules if bind to semi-open conformation, preserve the key interactions with the residues. 2D clustering indicated that topological pattern corresponding to ligand–receptor interactions, responsible for AK inhibitory activity, is not uniform throughout the data set. Substantial conclusion drawn was that a single pharmacophore for enzyme inhibitors is not sufficient due to multi specific requirements of flexible binding site and interaction mechanism. This modeling experiment provided a rationale for molecular recognition and interactions, which will heighten the design of potent inhibitors and improve the understanding of behavior of inhibitors within close representatives of ribokinase family.

Acknowledgements

The authors thank Council of Scientific and Industrial Research (CSIR), New Delhi, for providing financial grant under Mission Mode Program CMM 0017. SB thanks UGC for Senior Research Fellowship. We wish to thank Dr Peter Carlqvist and Gary Battle from CCDC for the useful suggestions, Katalin Nadassy from Accelrys for helping on some matters while dealing with Catalyst.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2009.12.001.

References

[1] D. Boison, Adenosine as a modular of brain activity, Drug News Perspect. 20 (2007) 607–611.
[2] E.A. Kowaluk, M.F. Jarvis, Therapeutic potential of adenosine kinase inhibitors, Expert Opin. Investig. Drugs 9 (2000) 555–564.
[3] K.J. Broadley, Drugs modulating adenosine receptors as potential therapeutic agents for cardiovascular diseases, Expert Opin. Ther. Pat. 10 (2005) 1669–1692.
[4] D. Boison, Adenosine kinase, epilepsy and stroke: mechanisms and therapies, Trends Pharmacol. Sci. 27 (2006) 652–658.
[5] M.F. Jarvis, Purinergic modulation of pain, Rev. Analgesia 7 (2003) 1–23.
[6] J. Sawynok, G. Doak, A. Poon, Adenosine and pain: recent findings with directly and indirectly acting agent, Drug Dev. Res. 45 (1998) 304–311.
[7] M. Sakowicz-Burkiewicz, K. Kocbuch, M. Grden, A. Szutowicz, T. Pawelczyk, Diabetes-induced decrease of adenosine kinase expression impairs the proliferation potential of diabetic rat T lymphocytes, Immunology 118 (2006) 402–412.
[8] S. McGaraughty, M. Cowart, M.F. Jarvis, Recent developments in the discovery of novel adenosine kinase inhibitors: mechanism of action and therapeutic potential, CNS Drug Rev. 7 (2001) 415–432.
[9] H.B. Cottam, New adenosine kinase inhibitors with oral antiinflammatory activity, Drugs Future 19 (1994) 485–491.
[10] M.A. Pak, H.L. Hass, U.K.M. Decking, J. Schrader, Inhibition of adenosine kinase increases endogenous adenosine and depresses neuronal activity in hippocampal slices, Neuropharmacology 33 (1994) 1049–1053.
[11] K.A. Jacobson, P.J.M. van Galen, M. Williams, Adenosine receptors: pharmacology, structure-activity relationships, and therapeutic potential, J. Med. Chem. 35 (1992) 407–422.
[12] B.G. Ugarkar, J.M. DaRe, J.J. Kopcho, C.E. Browne III, J.M. Schanzer, J.B. Wiesner, M.D. Erion, Adenosine kinase inhibitors. 1. Synthesis, enzyme inhibition and antiseizure activity of 5-iodotubercidin analogues, J. Med. Chem. 43 (2000) 2883–2893.
[13] B.G. Ugarkar, A.J. Castellino, J.M. DaRe, J.J. Kopcho, J.B. Wiesner, J.M. Schanzer, M.D. Erion, Adenosine kinase inhibitors. 2. Synthesis, enzyme inhibition and antiseizure activity of diaryltubercidin analogues, J. Med. Chem. 43 (2000) 2894–2905.
[14] J. Caballero, M. Fernandez, F.D. Gonzalez-Niloa, A CoMSIA study on the adenosine kinase inhibition of pyrrolo[2,3-d]pyrimidine nucleoside analogues, Bioorg. Med. Chem. 16 (2008) 5103–5108.
[15] M.P. Gonzalez, M.C.T. Moldes, A TOPS-MODE approach to predict adenosine kinase inhibition, Bioorg. Med. Chem. Lett. 14 (2004) 3077–3079.
[16] B.G. Ugarkar, A.J. Castellino, J.M. DaRe, J.J. Kopcho, S. Rosengren, M.D. Erion, Adenosine kinase inhibitors. 3. Synthesis, SAR, and antiinflammatory activity of a series of L-lyxofuranosyl nucleosides, J. Med. Chem. 46 (2003) 4750–4760.
[17] E.A. Kowaluk, S.S. Bhagwat, M.F. Jarvis, Adenosine kinase inhibitors, Curr. Pharm. Des. 4 (1998) 403–416.
[18] A. Gomtsyan, C.-H. Lee, Nonnucleoside inhibitors of adenosine kinase, Curr. Pharm. Des. 10 (2004) 1093–1103.
[19] Y. Lee, N. Bharatham, K. Bharatham, K.W. Lee, Adenosine kinase inhibitor design based on pharmacophore modeling, Bull. Korean Chem. Soc. 28 (2007) 561–566.
[20] S.W. Muchmore, R.A. Smith, A.O. Stewart, M.D. Cowart, A. Gomtsyan, M.A. Matulenko, H. Yu, J.M. Severin, S.S. Bhagwat, C. Lee, E.A. Kowaluk, M.F. Jarvis, C.L. Jakob, Crystal structures of human adenosine kinase inhibitor complexes reveal two distinct binding modes, J. Med. Chem. 49 (2006) 6726–6731.
[21] Y. Zhang, M.H. Kounib, S.E. Ealickc, Substrate analogs induce an intermediate conformational change in Toxoplasma gondii adenosine kinase, Acta Crystallogr. D 63 (2007) 126–134.
[22] S. Durdagi, C.T. Supuran, T.A. Strom, N. Doostdar, M.K. Kumar, A.R. Barron, T. Mavromoustakos, M.G. Papadopoulos, *In silico* drug screening approach for the

design of magic bullets: a successful with anti-HIV fullerene derivatized amino acids, J. Chem. Inf. Model. 49 (2009) 1139–1143.
[23] S. Durdagi, T. Mavromoustakos, N. Chronakis, M.G. Papadopoulos, Computational design of novel fullerene analogues as potential HIV-1 PR inhibitors: analysis of the binding interactions between fullerene inhibitors and HIV-1 PR residues using 3D QSAR, molecular docking and molecular dynamics simulations, Bioorg. Med. Chem. 16 (2008) 9957–9974.
[24] I.I. Mathews, M.D. Erion, S.E. Ealick, Structure of human adenosine kinase at 1.5 Å resolution, Biochemistry 37 (1998) 15607–15620.
[25] J. Gunther, A. Bergner, M. Hendlich, G. Klebe, Utilising structural knowledge in drug design strategies: applications using Relibase, J. Mol. Biol. 326 (2003) 621–636.
[26] M.A. Matulenko, E.S. Paight, R.R. Frey, A. Gomtsyan, S. DiDomenico Jr., M. Jiang, C.-H. Lee, A.O. Stewart, H. Yu, K.L. Kohlhaas, K.M. Alexander, S. McGaraughty, J. Mikusa, K.C. Marsh, S.W. Muchmore, C.L. Jakob, E.A. Kowaluk, M.F. Jarvis, S.S. Bhagwat, 4-Amino-5-aryl-6-arylethynylpyrimidines: structure-activity relationships of non-nucleoside adenosine kinase inhibitors, Bioorg. Med. Chem. 15 (2007) 1586–1605.
[27] M.A. Schumacher, D.M. Scott, I.I. Mathews, S.E. Ealick, D.S. Roos, B. Ullman, R.G. Brennan, Crystal structures of toxoplasma gondii adenosine kinase reveal a novel catalytic mechanism and prodrug binding, J. Mol. Biol. 296 (2000) 549–567.
[28] J. Shi, T.L. Blundell, K. Mizuguchi, FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties, J. Mol. Biol. 310 (2001) 243–257.
[29] A. Sali, T.L. Blundell, Comparative protein modeling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[30] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[31] InsightII, Version 4.1, 2007, Accelrys Inc., 9685 Scranton Rd, San Diego, CA 92121, U.S.A.
[32] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.
[33] M.L. Verdonk, G. Chessari, J.C. Cole, M.J. Hartshorn, C.W. Murray, Modeling water molecules in protein-ligand docking using GOLD, J. Med. Chem. 48 (2005) 6504–6515.
[34] M.L. Verdonk, V. Berdini, M.J. Hartshorn, W.T.M. Mooij, C.W. Murray, R.D. Taylor, P. Watson, Virtual screening using protein-ligand docking: avoiding artificial enrichment, J. Chem. Inf. Comput. Sci. 44 (2004) 793–806.
[35] SPSS, Version 15.0, 2006, SPSS Inc., Chicago, Illinois 60606, U.S.A.
[36] Catalyst, Version 4.1; 2006, Accelrys Inc., 9685 Scranton Rd, San Diego, CA 92121, U.S.A.
[37] E.M. Krovat, T. Langer, Non-peptide angiotensin II receptor antagonists: chemical feature based pharmacophore identification, J. Med. Chem. 46 (2003) 716–726.
[38] A.K. Debnath, Generation of predictive pharmacophore models for CCR5 antagonists: study with piperidine- and piperazine-based compounds as a new class of HIV-1 entry inhibitors, J. Med. Chem. 46 (2003) 4501–4515.
[39] C. Lee, M. Jiang, M. Cowart, G. Gfesser, R. Perner, K.H. Kim, Y.G. Gu, M. Williams, M.F. Jarvis, E.A. Kowaluk, A.O. Stewart, S.S. Bhagwat, Discovery of 4-amino-5-(3-bromophenyl)-7-(6-morpholino-pyridin-3-yl) pyrido[2,3-*d*] pyrimidine, an orally active, non-nucleoside adenosine kinase inhibitor, J. Med. Chem. 44 (2001) 2133–2138.
[40] Cerius^2, Version 4.10, 2006, Accelrys Inc., 9685 Scranton Rd, San Diego, CA 92121, U.S.A.
[41] MOE, Version 2008.10, 2008, Chemical Computing Group Inc., 1010 Sherbrooke Street W, Suite 910, Montreal, Canada.
[42] D.K. Agrafiotis, J.C. Myslik, F.R. Salemme, Advances in diversity profiling and combinatorial series design, Mol. Diversity 4 (1999) 1–22.
[43] D.T. Stanton, Evaluation and use of BCUT descriptors in QSAR and QSPR studies, J. Chem. Inf. Comput. Sci. 39 (1999) 11–20.
[44] L. Birch, C.W. Murray, M.J. Hartshorn, I.J. Tickle, M.L. Verdonk, Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase, J. Comput.-Aided Mol. Des. 16 (2002) 855–869.
[45] C.W. Murray, C.A. Baxter, A.D. Frenkel, The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase, J. Comput.-Aided Mol. Des. 13 (1999) 547–562.
[46] G. Klebe, Virtual ligand screening: strategies, perspectives and limitations, Drug Discov. Today 11 (2006) 580–594.
[47] A. Evers, G. Klebe, Ligand-supported homology modeling of G-protein-coupled receptor sites: models sufficient for successful virtual screening, Angew. Chem. Int. Ed. 43 (2004) 248–251.