BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

BIOMEDICAL INFORMATICS

www.bioinformation.net

Volume 12(3)

Hypothesis

Prediction of kinase-inhibitor binding affinity using energetic parameters

Singaravelu Usha and Samuel Selvaraj*

Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli - 620 024, Tamilnadu, India. Dr. Samuel Selvaraj - Email: selvarajsamuel@gmail.com; Mobile: +919894390363; *Corresponding author

Received May 13, 2016; Revised May30, 2016; Accepted May 30, 2016; Published June 15, 2016

Abstract
The combination of physicochemical properties and energetic parameters derived from protein-ligand complexes play a vital role in determining the biological activity of a molecule. In the present work, protein-ligand interaction energy along with logP values was used to predict the experimental log (IC_{50}) values of 25 different kinase-inhibitors using multiple regressions which gave a correlation coefficient of 0.93. The regression equation obtained was tested on 93 kinase-inhibitor complexes and an average deviation of 0.92 from the experimental log IC_{50} values was shown. The same set of descriptors was used to predict binding affinities for a test set of five individual kinase families, with correlation values > 0.9. We show that the protein-ligand interaction energies and partition coefficient values form the major deterministic factors for binding affinity of the ligand for its receptor.

Keywords: Inhibition constant prediction, Protein-ligand interaction, energetic and solvent descriptors, Kinase inhibitors

Background:
Protein kinases are a large family of homologous proteins with more than 500 members in the human proteome **[1].** Kinase-mediated protein phosphorylation is a crucial component of the signal transduction pathways which plays a central role in diverse biological processes such as cell growth, metabolism, differentiation, and apoptosis **[2].** A number of diseases, including cancer, diabetes, inflammation, immune and neurodegenerative disorders are linked to perturbation of protein kinase–mediated cell signaling pathways **[3].** Since all members of the kinase families utilize ATP, kinase inhibitors are designed to bind with the ATP to prevent substrate phosphorylation **[1].**

Over 20 small-molecule protein kinase inhibitors have been currently approved and more than 150 kinase inhibitors are undergoing clinical trials **[4].** Yet, issues such as target specificity, **[5-8]** resistance development **[9-10]** hinge region binding and activation state dependence of kinase inhibitors need to be addressed **[2-4]**.

The three-dimensional structures of proteins with bound ligand are available in the Protein Data Bank **[11]** along with their experimental binding affinity information. Binding affinity data such as Ki, Kd, IC_{50} etc. obtained from experimental studies are also available in databases such as BindingDB **[12],** Binding MOAD **[13],** PDBbind **[14]** etc. Availability of valuable resources regarding kinase inhibitors made computational biologists to develop statistical models to accurately predict the binding affinity of complexes.

Structure-based virtual screening methods use docking programs to explore the possible binding modes of a ligand within the target binding site, and scoring functions to estimate the affinity of the ligand for the binding site **[15, 16].** While docking methods at present are in general successful in predicting the correct

binding conformations of ligand molecules, they do not perform well in correctly predicting the binding affinity for the predicted ligand conformations **[2].** Hence, it is essential to predict the binding affinity of a given ligand to its target known as the 'scoring problem' **[17]**.

Table 1: Experimentally observed and predicted IC_{50} values for 25 kinase-inhibitor complexes.

PDB_ID	Ligand_ID	Experimental	Experimental	Predicted log(IC_{50} (nM))	
		IC_{50} (nM)	Log (IC_{50} (nM))	Back-check	Jack-knife test
2I6A	5I5	22.8	1.36	1.33 (0.03)	1.30 (0.06)
2OO8	RAJ	1	0.00	0.26 (-0.26)	0.48 (-0.48)
4AT3	LTI	4	0.60	0.28 (0.32)	0.13 (0.47)
3SXF	BK5	5	0.70	0.69 (0.01)	0.69 (0.01)
2C1A	I5S	170	2.23	1.43 (0.80)	1.96 (0.27)
3MB6	01I	100	2.00	1.71 (0.29)	1.54 (0.46)
1Y6B	AAX	38	1.58	0.85 (0.73)	1.90 (-0.32)
2A4L	RRC	400	2.60	2.82 (-0.22)	2.90 (-0.30)
2YAK	OSV	2	0.30	0.24 (0.07)	0.17 (0.13)
4GK2	L66	40	1.60	1.55 (0.05)	1.53 (0.07)
3POZ	03P	23	1.36	1.65 (-0.28)	1.91 (-0.55)
4F64	0S8	63	1.80	2.00 (-0.20)	2.15 (-0.35)
3BZ3	YAM	1.5	0.18	0.27 (-0.09)	0.33 (-0.15)
1Q3D	STU	15	1.18	1.28 (-0.11)	1.34 (-0.16)
3C1X	CKK	45	1.65	1.01 (0.64)	0.68 (0.97)
3D94	D94	19	1.28	0.96 (0.32)	0.77 (0.51)
4BKZ	1WS	27	1.43	1.87 (-0.43)	2.20 (-0.77)
3HRB	I39	21	1.32	1.80 (-0.48)	1.93 (-0.61)
4BFV	ZVV	140	2.15	1.45 (0.70)	0.54 (1.61)
3LJ3	WYE	43	1.63	2.12 (-0.49)	2.74 (-1.10)
2VGO	AD5	500	2.70	1.93 (0.77)	1.39 (1.31)
4HDC	13Y	1.2	0.08	0.37 (-0.29)	0.77 (-0.69)
1RW8	580	1320	3.12	2.96 (0.16)	2.84 (0.28)
3KRR	DQX	0.48	-0.32	-0.26 (-0.06)	-0.16 (-0.16)
3MVH	WFE	0.5	-0.30	0.05 (-0.35)	0.19 (-0.49)
			Average deviation	**0.41**	**0.63**

Note: The deviation between the predicted and experimental IC_{50} values is given in parenthesis.

As a pioneering work, Bohm **[18]** (1994) developed a simple empirical function (LUDI) to estimate the binding constant for a protein-ligand complex of known structure. This empirical scoring function takes into account hydrogen bonds, ionic interactions, the lipophilic protein-ligand contact surface and the number of rotatable bonds in the ligand. Head *et al*. (1996) in their VALIDATE approach used electrostatic and steric interaction energies, octanol-water partition coefficient, polar and nonpolar contact surfaces, and a term to describe intramolecular flexibility **[19].** Following the approach of Bohm, **[18]** Eldridge *et al.* **[20]** (1997) included intramolecular flexibility in ChemScore and Wang *et al*. **[21]** (1998) classified hydrogen bonds and included the occurrence of interstitial water molecules in SCORE. Based on the statistical analysis of experimentally observed distributions

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

and frequencies of distance-dependent protein-ligand atom pair interactions, the potential of mean force (PMF) was obtained which has been used for affinity predictions of large datasets **[22].** Based on a larger set of 200 protein–ligand complexes, Wang *et al.* (2002) developed X-Score **[17],** consisting of four energy terms including van der Waals interactions, hydrogen bonds, hydrophobic effects and effective rotatable bonds. Docking programs such as FlexX **[23]** and Surflex **[24]** use empirical scoring functions by including different empirical energy terms. A large-scale validation of a quantum mechanics based scoring function to predict the binding affinity and binding mode of a diverse set of protein-ligand complexes containing different protein families including aspartic proteases, serine proteases, sugar binding proteins, amino acid binding proteins, and protein kinases was done by Raha and Merz (2004) **[25].** LigScore functions **[26]** have made use of three distinct terms, the van der Waals interaction, the polar attraction between the ligand and protein, and the desolvation penalty attributed to the binding of the polar ligand atoms to the protein to predict the experimental pK_i values of a diverse set of 118 protein–ligand complexes that span more than seven protein families. CLiBE, a database of computed ligand binding energy (based on molecular mechanics force field) for ligand-receptor three dimensional structures have been developed and a linear correlation between the computed ligand-receptor interaction energy and experimental binding affinity (Kcal/mol) has been observed **[27].** A Program for Energetic Analysis of Receptor-Ligand System (PEARLS) has been developed to compute free energy of protein-ligand complexes **[28].**

While the above methods use the known three dimensional structures to predict binding affinities, the Quantitative Structure-Activity Relationship (QSAR) methods serve as an alternative way of binding affinity predictions in the absence of 3D structure of target proteins or their complexes with ligands. These methods make use of physicochemical and structural properties (descriptors) of ligands to relate their biological activity using regression methods. Combined QSAR approaches in binding affinity predictions have been recently reported **[29, 30].**

In the present work, we have correlated the experimental IC_{50} values (in their logarithmic form) of 25 different kinase-inhibitor complexes with their protein- interaction energy and partition coefficient (logP) values for multiple regression analysis, which shows a good correlation with the experimental IC_{50} values. This shows that the protein-ligand interaction energies and logP values form the major factors that determine the ligand binding affinity of proteins. By incorporating these energetic as well as solvent terms, docking methodologies can be highly successful in predicting the binding affinity for the generated poses of their correct ligand binding modes.

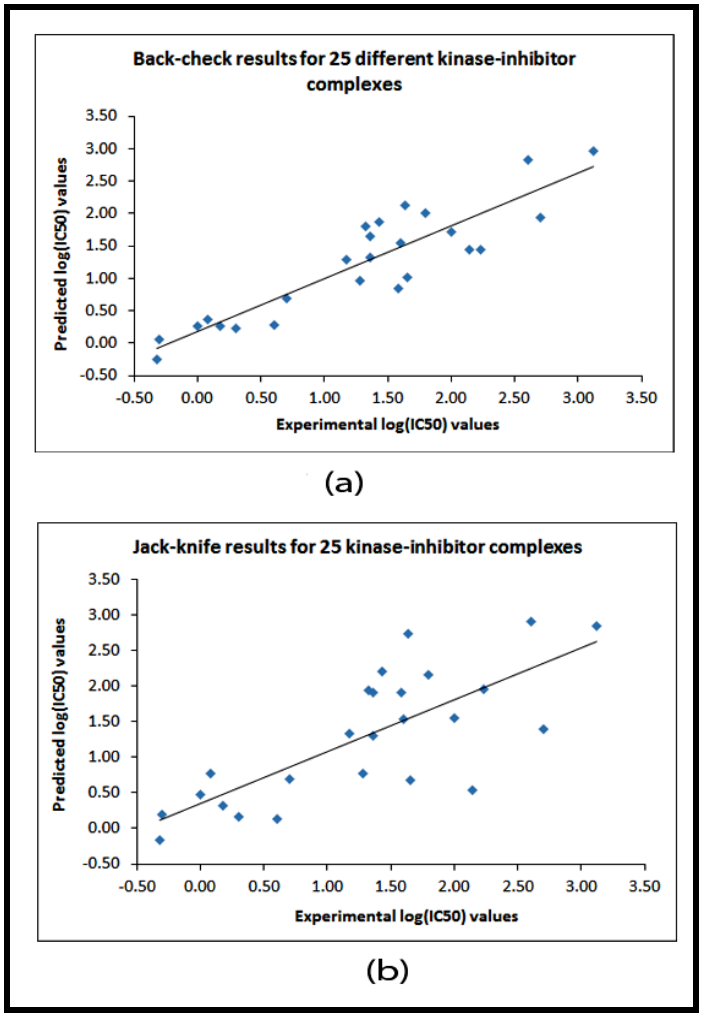

Figure 1: (a) Scatter plot of experimental and back-check predicted IC_{50} values in 25 different kinase-inhibitor complexes, (b) Scatter plot of experimental and jack-knife predicted IC_{50} values in 25 different kinase-inhibitor complexes.

Methodology

Information resources

Twenty five different protein kinase-inhibitor complexes solved by x-ray crystallography method were obtained from Protein Data Bank **[11].** The complexes in the dataset have resolution less than 2.5 Å with known IC_{50} values were treated as training set. The number of non-hydrogen atoms of the ligands and energetic profile comprising of i) total ligand-receptor interaction energy, ii) van der Waals energy, iii) electrostatic energy, iv) hydrogen bond energy, v) solvation free energy, vi) conformational entropy and vii) ligand-water-receptor binding energy were obtained from the PEARLS server for each of the kinase-inhibitor complexes. The PEARLS server uses the AMBER force field **[31]** for computing the above energetic contributions **[28].** LogP (octanol/water partition coefficient) values of the ligands were

calculated from the Molinspiration server **[32]** by providing SMILES code of the ligand as input.

Training set construction and validation

Multiple regression analysis was carried out to establish a relationship between the above-mentioned descriptors and experimental log (IC_{50}). A back-check test was carried out for predicting the binding affinity by re-substituting the values in the regression equation obtained. For the jack-knife test, coefficients of multiple regressions were determined using (n-1) data (omitting one protein-ligand complex at a time) and then predicting binding affinity of the omitted protein-ligand complex.

Test set information

The regression equation obtained from the training set was tested on i) a set of 93 kinase-inhibitor complexes with IC_{50} values, and ii) a set of 9 approved kinase inhibitors **[2].**

To further assess the predominant role of the chosen descriptors in binding affinity predictions, the experimental log (IC_{50}) values were regressed with the same set of nine descriptors in five independent protein-kinase families comprising 17 cyclic AMP-dependent kinase-inhibitors, 12 casein kinase-inhibitors, 15 hepatocyte growth factor receptor kinase-inhibitors, 12 cyclin-dependent kinase-inhibitors and 16 mitogen-activated kinase-inhibitors. For each of the five kinase families, five different regression equations were obtained which were then validated by back-check analysis. The dataset information of all the kinase-inhibitor complexes used in the present study, including PDB ID, protein name, ligand ID, x-ray resolution (Å), experimental IC_{50} values (nM) with their logarithmic form, and descriptor values are provided in the Appendix.

Discussion

The following multiple regression equation (1) between log (IC_{50}) values and nine energetic descriptors and log P with a correlation coefficient, r = 0.93 was obtained for the training set of 25 kinase-inhibitor complexes.

log (IC_{50}) = – 0.07 NHA + 44.71 IE – 44.61 vdW – 44.51 Elect – 44.47 H-bonds – 45.45 Solv – 43.19 entropy – 42.82 H-bonds (water-mediated) + 0.18 LogP + 3.58 (1)

where NHA denotes the number of non-hydrogen atoms of the ligand, IE, the total Ligand-receptor interaction energy (Kcal/mol), vdW, the van der Waals energy (Kcal/mol), Elect, the electrostatic Energy (Kcal/mol), H-bonds, the hydrogen bond energy (Kcal/mol), Solv, the solvation free energy (Kcal/mol), entropy, the conformational entropy (Kcal/mol), H-bonds (water-mediated), the ligand-water-receptor binding energy (Kcal/mol) and logP, (the octanol-water) Partition coefficient.

The experimentally observed and predicted IC_{50} values for 25 kinase-inhibitor complexes in back-check and jack-knife predictions are provided in **Table 1**. The average deviation of the predicted log (IC_{50}) values from the experimental log (IC_{50}) values was 0.41 for back-check and 0.63 for jack-knife predictions respectively. The relationship between the experimental log (IC_{50}) values with back-check and jack-knife predictions are provided as scatter plots in **Figure 1a and 1b.**

Table 2: Prediction of experimental log (IC_{50}) values in approved kinase inhibitors.

S. NO.	PDB ID	Protein name	Ligand ID	Ligand name	Experimental log (IC_{50} (nM))	Predicted log (IC_{50} (nM))
1	1M17	Epidermal growth factor receptor	AQ4	Erlotinib	- 0.70 – 3.16 (1.23)	0.87
2	1IEP	Proto-oncogene tyrosine-protein kinase abl	STI	Imatinib	0.04 – 3.93 (1.99)	1.79
3	2ITY	Epidermal growth factor receptor	IRE	Gefitinib	0.00 – 3.44 (1.72)	1.46
4	2J2I	Proto-oncogene serine/threonine-protein kinase pim-1	LY4	Ruboxistaurin	2.30	1.61
5	2GQG	Proto-oncogene tyrosine-protein kinase abl1	1N1	Dasatinib	- 0.70 – 2.85 (1.07)	-0.97
6	2JAV	Serine/threonine-protein kinase nek2	5Z5	Sunitinib	3.90	1.02
7	1UWH	B-raf proto-oncogene serine/threonine-protein kinase	BAX	Sorafenib	1.04 – 3.86 (2.45)	1.12
8	1XKK	Epidermal growth factor receptor	FMM	Lapatinib	0.46 – 2.64 (1.55)	1.88
9	2F2U	Rho-associated protein kinase 2	M77	Fasudil	2.26 – 4.07 (3.16)	2.45

Note: The mean value between the logarithm of minimum and maximum experimental IC_{50} values are given in parenthesis.

Table 3: Experimentally observed and predicted IC_{50} values for 17 cyclic AMP-dependent protein kinase-inhibitor complexes.

PDB_ID	LIG_ID	Experimental IC_{50} (nM)	Experimental log (IC_{50} (nM))	Back-check Prediction
1STC	STU	51	1.71	1.45 (0.26)
1SVE	I01	5	0.70	0.99 (-0.29)
1XH4	R69	30	1.48	1.58 (-0.10)
1YDS	IQS	5300	3.72	3.13 (0.59)
2C1A	I5S	170	2.23	2.23 (0.00)
2F7X	4EA	38	1.58	1.64 (-0.06)
2GNI	M77	7605	3.88	3.59 (0.29)
2JDS	L20	27	1.43	1.45 (-0.01)
2OH0	2PY	18	1.26	1.21 (0.04)
2OJF	4PY	110	2.04	2.15 (-0.11)
2UW6	GVO	280	2.45	3.07 (-0.62)
2UZT	SS3	14	1.15	1.11 (0.04)
3L9L	L9L	167	2.22	1.78 (0.45)
3MVJ	XFE	3200	3.51	3.77 (-0.27)
3OW3	SMY	742	2.87	2.88 (-0.01)
3ZO2	15I	80	1.90	2.10 (-0.19)
4C35	NU3	560	2.75	2.76 (-0.01)
			Average deviation	**0.28**

Note: The deviation between the predicted and experimental IC_{50} values is given in parenthesis.

Table 4: Experimentally observed and predicted IC_{50} values for 12 casein kinase-inhibitor complexes.

PDB_ID	LIG_ID	Experimental IC_{50} (nM)	Experimental log(IC_{50} (nM))	Back-check Prediction
2QC6	G12	100	0.27	0.24 (0.03)
2ZJW	REF	40	0.38	0.37 (0.01)
3AMY	AGI	800	0.36	0.38 (-0.02)
3BQC	EMO	2000	0.18	0.20 (-0.03)
3MB6	01I	100	0.24	0.25 (0.00)
3PE1	3NG	1	0.20	0.21 (-0.01)
3PWD	CZ0	220	0.34	0.31 (0.03)
3R0T	FU9	0.91	0.24	0.24 (0.01)
3RPS	4B0	320	0.36	0.37 (-0.01)
3U4U	LNH	3100	0.34	0.34 (0.01)
4ANM	WUL	28	0.23	0.22 (0.01)
4DGM	AGI	1200	0.22	0.23 (-0.02)
			Average deviation	**0.02**

Note: The deviation between the predicted and experimental IC_{50} values is given in parenthesis.

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

a) Test set of diverse protein kinases (Test set I)
The regression equation (1) obtained was tested on 93 kinase-inhibitor complexes (results are provided as **Table 7** in the Supporting Information file) and the relationship between the experimental and predicted log (IC_{50}) values is presented as a scatter plot **(Figure 2).** An average deviation of 0.92 from the original log (IC_{50}) values was observed for the 93 kinase-inhibitor complexes. The difference between the experimental and calculated log (IC_{50}) values was found to be less than ±1 log unit for 64 out of 93 kinase-inhibitor complexes.

b) Approved kinase inhibitors as test set (Test set II)
To further test the predictability of our regression equation (1), we have tested it for nine approved kinase inhibitors[2]. The experimental values of those inhibitors were found to have minimum and maximum range of values. Hence, the mean value between the logarithm of minimum and maximum values were calculated and compared with the predicted values. The predicted values were almost closer (the deviation was less than 1) to the experimental log (IC_{50}) values in 7 out of 9 kinase-inhibitor complexes **(Table 2).**

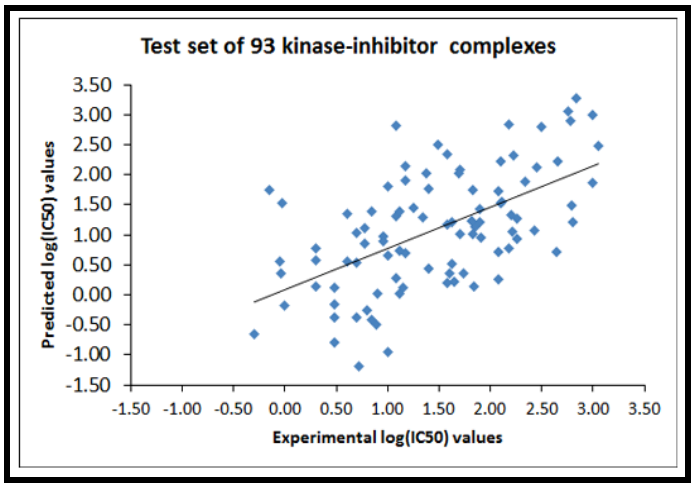

Figure 2: Scatter plot of experimental and predicted IC_{50} values in the test set of 93 kinase-inhibitor complexes.

c) Kinase classes
In order to further validate the use of the same set of descriptors in predicting binding affinity, the experimental IC_{50} values of five kinase families were regressed. The regression equation obtained for individual kinase families and the result of back-check predictions are discussed as follows:

i) Cyclic AMP-dependent protein kinase-inhibitor complexes
For a data-set of 17 cyclic AMP-dependent protein kinase-inhibitor complexes, a correlation value of 0.95 was obtained using the regression equation (2)

log (IC_{50}) = -0.06 NHA – 76.09 IE + 76.08 vdW + 75.98 Elect + 76.26 H-bonds + 76.62 Solv + 75.21 entropy + 77.07 H-bonds (water-mediated) – 0.13 LogP + 5.21 (2)

Using the regression equation (Eq. 2), log (IC_{50}) values for 17 cyclic AMP-dependent protein kinase-inhibitor complexes were predicted. The experimental as well as predicted log (IC_{50}) values are presented **(Table 3)** and plotted **(Figure 3a).** The average deviation for the back-check test was 0.28 from the experimental values.

ii) Casein kinase-inhibitor complexes
12 casein kinase-inhibitor complexes were taken for the multiple regression analysis which has shown a good correlation of r = 0.97 for the regression equation (3)

log (IC_{50}) = 0.01 NHA – 1.40 IE + 1.45 vdW + 1.39 Elect + 1.36 H-bonds + 1.43 Solv + 1.35 entropy + 1.69 H-bonds (water-mediated) + 0.01 LogP + 0.47 (3)

The set of 12 casein kinase-inhibitor complexes with their experimental and predicted values has been provided **(Table 4).** The scatter plot shows the relationship between the experimental and predicted IC_{50} values **(Figure 3b),** the average deviation being 0.02 for back-check predictions.

iii) Hepatocyte growth factor receptor kinase-inhibitor complexes
A set of 15 hepatocyte growth factor receptor kinase-inhibitors has shown a correlation coefficient value of 0.90 when subjected to regression with multiple descriptors, the equation (4) being

log (IC_{50}) = 0.05 NHA – 232.16 IE + 231.87 vdW + 231.02 Elect + 232.03 H-bonds + 232.68 Solv + 224.75 entropy +236.13 H-bonds (water-mediated) – 0.27 LogP + 0.24 (4)

The experimental and predicted log (IC_{50}) values are provided in **Table 5.** The correlation between experimental and calculated values for the 15 hepatocyte growth factor receptor kinase-inhibitors is shown in **Figure 3c**. An average deviation of 0.31 was observed.

iv) Cyclin-dependent kinase-inhibitor complexes
A very good correlation of r = 0.94 was obtained for 12 cyclin-dependent kinase-inhibitor dataset using the regression equation (5)

log (IC_{50}) = – 0.35 NHA + 94.15 IE – 94.18 vdW – 96.24 Elect – 92.97 H-bonds – 92.06 Solv –96.77 entropy – 95.47 H-bonds (water-mediated) + 1.94 LogP + 6.27 (5)

The predicted results of 12 cyclin-dependent kinase-inhibitor complexes are tabulated **(Table 6).** The average deviation value from the experimental value was found to be 0.49. The results are plotted **(Figure 3d)**.

Table 5: Experimentally observed and predicted IC_{50} values for 15 hepatocyte growth factor receptor kinase-inhibitor complexes.

PDB_ID	LIG_ID	Experimental IC_{50} (nM)	Experimental log(IC_{50} (nM))	Back-check Prediction
2WD1	ZZY	82	1.91	1.56 (0.36)
2WKM	PFY	9	0.95	1.26 (-0.30)
3C1X	CKK	45	1.65	1.56 (0.09)
3CCN	LKG	120	2.08	1.69 (0.39)
3CD8	L5G	9	0.95	1.11 (-0.16)
3CTH	319	35	1.54	1.84 (-0.29)
3I5N	B2D	17	1.23	1.69 (-0.46)
3F66	IHX	900	2.95	2.88 (0.07)
3L8V	L8V	8	0.90	0.72 (0.18)
3QTI	3QT	14	1.15	1.18 (-0.04)
3RHK	M97	520	2.72	2.60 (0.11)
4DEG	0JJ	6	0.78	1.43 (-0.65)
4DEH	0JK	612	2.79	2.71 (0.08)
4EEV	L1X	42	1.62	1.09 (0.53)
4GG7	0J8	6.5	0.81	0.72 (0.09)
			Average deviation	**0.31**

Table 6: Experimentally observed and predicted IC_{50} values for 12 cyclin-dependent kinase-inhibitor complexes.

PDB_ID	LIG_ID	PDB_ID	Experimental IC_{50} (nM)	Experimental log(IC_{50} (nM))	Back-check Prediction
1AQ1	STU	1AQ1	7	0.85	0.71 (0.13)
1DI8	DTQ	1DI8	1000	3.00	3.07 (-0.07)
1E1X	NW1	1E1X	2200	3.34	3.89 (-0.54)
1H01	FAL	1H01	22000	4.34	4.74 (-0.40)
1W0X	OLO	1W0X	7	0.85	1.59 (-0.74)
3S2P	PMU	3S2P	68	1.83	1.36 (0.48)
3TIY	TIY	3TIY	17000	4.23	4.36 (-0.13)
3TNW	F18	3TNW	20000	4.30	3.52 (0.78)
3ULI	1N3	3ULI	70	1.85	1.48 (0.37)
3UNJ	0BX	3UNJ	11000	4.04	3.41 (0.63)
3WBL	PDY	3WBL	23000	4.36	4.22 (0.14)
4BGH	3I6	4BGH	4	0.60	1.24 (-0.63)
				Average deviation	**0.49**

Note: The deviation between the predicted and experimental IC_{50} values is given in parenthesis.

Table 7: Experimentally observed and predicted IC_{50} values for 16 mitogen-activated protein kinase-inhibitor complexes

PDB_ID	LIG_ID	Experimental IC_{50} (nM)	Experimental log (IC_{50} (nM))	Back-check Prediction
1W82	L10	196	2.29	2.40 (-0.11)
1WBN	L09	350	2.54	2.20 (0.34)
1ZYJ	BI5	1500	3.18	2.84 (0.33)
3FLZ	FLZ	106	2.03	2.25 (-0.22)
3FMH	533	11	1.04	1.25 (-0.21)
3HL7	I47	110	2.04	1.56 (0.48)
3HP2	P36	680	2.83	3.07 (-0.24)
3HRB	I39	21	1.32	1.99 (-0.67)
3IPH	G11	316.23	2.50	2.40 (0.10)
3L8X	N4D	10	1.00	0.96 (0.04)
3MVM	39P	3.9	0.59	0.63 (-0.03)
3NWW	3NW	7	0.85	0.96 (-0.12)
3S4Q	NK0	4	0.60	0.44 (0.16)
3UVP	048	35	1.54	1.26 (0.28)
3ZSG	T75	7.1	0.85	1.15 (-0.30)
4EWQ	GG5	600	2.78	2.62 (0.16)
			Average deviation	0.29

Note: The deviation between the predicted and experimental IC_{50} values is given in parenthesis.

v) Mitogen-activated protein kinase-inhibitor complexes

The multiple regression analysis of 16 mitogen-activated protein kinase-inhibitors gave a correlation of r = 0.94 using the regression equation (6)

$$\log (IC_{50}) = - 0.13 \text{ NHA} + 16.62 \text{ IE} - 16.05 \text{ vdW} - 17.09 \text{ Elect} - 16.66 \text{ H-bonds} - 16.62 \text{ Solv} - 13.78 \text{ entropy} - 18.19 \text{ H-bonds (water-mediated)} + 0.40 \text{ LogP} + 8.27 \quad (6)$$

The observed and computed values for a dataset of 16 mitogen-activated protein kinase-inhibitors are presented **(Table 7)** showing an average deviation value of 0.29. The predicted IC_{50} values were plotted against the experimental values **(Figure 3e).**

In QSAR studies, usually a large number of physicochemical and structural properties (descriptors) of chemical compounds are calculated and the best combinations of descriptors that correlate maximally with the biological activity are chosen. QSAR models have been developed for predicting kinase selectivity profiles to provide understanding of structure selectivity relationships for kinase inhibitor design **[33-35].** The development of widely accepted 'universal' set of descriptors applicable for diverse datasets has also been a focus of QSAR-based analysis **[36-38].** In the present work, a uniform set of descriptors have been used across different kinase-inhibitor complexes for binding affinity prediction.

Han *et al*. (2006) suggested that the calculated interaction energies highly depend on van der Waals contacts, electrostatic interactions, hydrogen bonds, metal-receptor binding, and solvation **[28],** which agrees with the observations of our present analysis. These observations are reflected in the interaction energy contributions of our present analysis. The protein-ligand interaction energies obtained using PEARLS server has been used in other studies on inhibitor discovery such as HIV-1 protease **[39]** and ribonuclease A inhibitors **[40]** to predict the binding affinity values using regression analysis. Log P, remains the main deterministic factor for the ligand's affinity for the protein active site with reference to the surrounding solvent environment **[41].**

In the present study, we have used the various energetic components as independent variables along with logP values, to predict the experimental binding affinity. This set of descriptors developed from a small set of 25 kinase-inhibitor complexes were able to predict IC_{50} values for 93 test set complexes spanning 4 orders of magnitude of IC_{50} values. The same set of descriptors was also found to be suitable for family specific regression models as well.

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

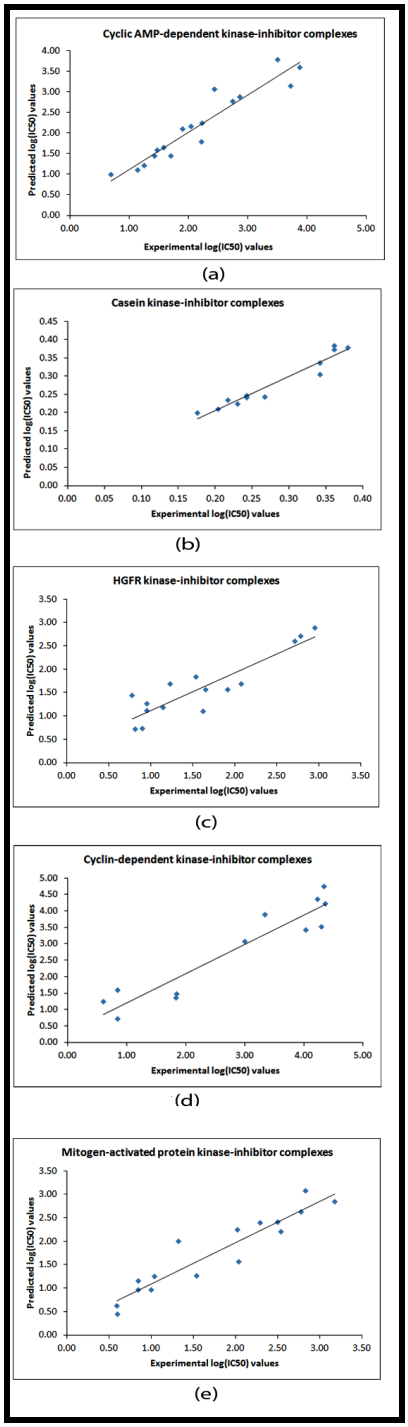

Figure 3: Scatter plot of experimental and back-check predicted IC_{50} values in (a) 17 cyclic AMP-dependent kinase-inhibitor complexes; (b) 12 caesin kinase-inhibitor complexes; (c) 15 Hepatocyte growth factor receptor kinase-inhibitor complexes; (d)12 cyclin-dependent kinase-inhibitor complexes; (e) 16 mitogen-activated protein kinase-inhibitor complexes.

As docking methods improve to reproduce conformations observed through x-ray crystallographic and NMR determined structures, it will be possible to use our present approach to predict the IC_{50} values for various protein targets, more significantly for specific protein families. Alternatively, if IC_{50} values for kinase-inhibitor complexes are known, the method can also be used to predict the pose of a given ligand as well.

Conclusion:

Despite intensive research over more than two decades, accurate prediction of the binding affinities of large set of diverse protein□ligand complexes remains one of the most important open problems in computational molecular biology **[42].** The issues currently being addressed are the scoring of modelled protein conformations, and including the binding free energy due to presence of water molecules **[43].** In the present work, we have addressed these issues by using energetic and solvent descriptors to predict the binding affinity of kinase-inhibitor complexes using multiple regression analysis. A high correlation value of 0.9 between the predicted and experimental binding affinity was obtained for a test set of kinase-inhibitor complexes. The method was validated by predicting a test of 93 kinase-inhibitor complexes covering five kinase families which has shown a good predictive ability. Our methodology can provide valuable insights for the prediction accuracy of molecular docking strategies. Further studies will be required to validate the general applicability of these set of descriptors to predict the binding affinity for a diverse set of enzyme-inhibitor complexes.

Acknowledgement:

SS thanks the University Grants Commission, New Delhi for the award of Emeritus Fellowship (Grant No: F.6-6/ 2014-15/ EMERITUS-2014-15-GEN-4545/ (SA-II)).

References:

[1] Patricelli MP *et al. Biochemistry* 2007 **46:**350 [PMID:17209545]
[2] Ghose AK *et al. J Med Chem.* 2008 **51:**5149 [PMID:18710211]
[3] Noble ME *et al. Science* 2004 **303:** 1800 [PMID:15031492]
[4] Cohen P & Alessi DR, *ACS Chem Biol.* 2013 **8:** 96 [PMID:23276252]
[5] Feng L *et al. J Am Chem Soc.* 2011 **133:** 5976 [PMID:21446733]
[6] Meggers E *Angew Chem Int Ed.* 2011 **50:** 2442 [PMID:21328511]
[7] Dar AC & Shokat KM, *Annu Rev Biochem.* 2011 **80:**769 [PMID:21548788]
[8] Endicott JA *et al. Annu Rev Biochem.* 2012 **81:**587 [PMID:22482904]
[9] Tamborini E *et al. Gastroenterology* 2004 **127:**294 [PMID:15236194]
[10] Cools J *et al. Drug Resist Updat.* 2005 **8:**119 [PMID:15869901]
[11] Berman HM *et al. Nucleic Acids Res.* 2000 **28:**235 [PMID:10592235]
[12] Liu T *et al. Nucleic Acids Res.* 2007 **35:**D198 [PMID:17145705]
[13] Hu L *et al. Proteins Struct Funct Bioinf.* 2005 **60:**333 [PMID:15971202]

[14] Wang R *et al. J Med Chem*. 2004 **47:**2977 [PMID:15163179]
[15] Leach AR *et al. J Med Chem*. 2006 **49:**5851 [PMID:17004700]
[16] Huang SY & Zou X, *Int J Mol Sci*. 2010 **11:**3016 [PMID:21152288]
[17] Wang R *et al. J Comput-Aided Mol Des.* 2002 **16:**11 [PMID:12197663]
[18] Böhm HJ *J Comput-Aided Mol Des*. 1994 **8:**243 [PMID:7964925]
[19] Head RD *et al. J Am Chem Soc*. 1996 **118:**3959
[20] Eldridge MD *et al. J Comput-Aided Mol Des.* 1997 **11:**425 [PMID: 9385547]
[21] Wang R *et al. J Mol Model*. 1998 **4:**379
[22] Muegge I & Martin YC, *J Med Chem.* 1999 **42:**791 [PMID:10072678]
[23] Rarey M *et al. J Mol Biol*. 1996 **261:**470 [PMID:8780787]
[24] Jain AN *J Med Chem*. 2003 **46:**499 [PMID:12570372]
[25] Raha K & Merz KM, *J Am Chem Soc.* 2004 **126:**1020 [PMID:14746460]
[26] Krammer A *et al. J Mol Graphics Modell.* 2005 **23:**395 [PMID:15781182]
[27] Chen X *et al.Comput Chem.* 2002 **26:**661 [PMID:12385480]
[28] Han LY *et al. J Chem Inf Model.* 2006 **46:**445 [PMID:16426079]
[29] Verma J *et al. Curr Top Med Chem.* 2010 **10:**95 [PMID:19929826]
[30] Martin E *et al. J Chem Inf Model.* 2011 **51:**1942 [PMID:21667971]
[31] Weiner SJ *et al. J Am Chem Soc.* 1984 **106:**765
[32] www.molinspiration.com/
[33] Anderson DR *et al. J Med Chem.* 2007 **50:**2647 [PMID:17480064]
[34] Sheridan RP *et al. J Chem Inf Model*. 2009 **49:**1974 [PMID:19639957]
[35] Sciabola S *et al. J Chem Inf Model*. 2008 **48:**1851 [PMID:18717582]
[36] Labute P *J Mol Graphics Modell.* 2000 **18:**464 [PMID:11143563]
[37] Dominy BN & Shakhnovich EI, *J Med Chem.* 2004 **47:**4538 [PMID:15317465]
[38] Burden FR *et al.J Chem Inf Model.* 2009 **49:**710 [PMID: 19434903]
[39] Vijjulatha M & Kanth SS, *Cent Eur J Chem.* 2007 **5:**1064
[40] Samanta A *et al. Bioorg Med Chem.* 2011 **19:**2478 [PMID:21420869]
[41] Leo AJ *Chem Rev.* 1993 **93:**1281
[42] Ballester PJ *et al. J Chem Inf Model.* 2014 **54:**944 [PMID:24528282]
[43] Moitessier N *et al. Br J Pharmacol.* 2008 **153:**7 [PMID:18037925]

Edited by P Kangueane
Citation: Usha & Selvaraj, Bioinformation 12(3): 172-181 (2016)