SELECTION OF BEST CRYSTAL STRUCTURE FOR INITIATING DOCKING-BASED VIRTUAL SCREENING STUDIES OF CDK2 INHIBITORS: A CROSS-DOCKING AND DUD SET VALIDATION APPROACH

Joshi A.[a], Bhojwani H. [a], and Joshi U. [a]*

(Received 03 October 2018) (Accepted 02 May 2019)

ABSTRACT

A total of 95 crystal structures of CDK2 were selected after considering criteria such as resolution and absence of missing residues in the active site; and subjected to cross-docking. 14 out of 95 crystal structures exhibited docking accuracy for greater than 70% of ligands at RMSD cut off 2Å in the cross-docking studies. These 14 crystal structures were selected for the second part of the study, which included validation using DUD sets and enrichment calculations. 8 out of 14 crystal structures possessed the enrichment factor of >10 at 1% of the ranked database. ROC-AUC, AUAC, RIE, and BEDROC were calculated for these 8 crystal structures. 2WXV produced maximum BEDROC (0.768, at α=8) and RIE (11.22). 2WXV as a single initial crystal structure in the virtual screening protocol is likely to produce more accurate results than any other single crystal structure.

Keywords: CDK2, CDK2 inhibitors, Virtual screening, Cross-docking, DUD set validation

INTRODUCTION

Virtual Screening (VS) assists in the prediction of biological activities of a large set of compounds before synthesizing them. This enables prioritization of synthesis and saves on time and money aspects, at the same time helps in rationalizing the drug discovery process. When the three-dimensional structure of the protein under consideration is available, molecular docking is the method of choice for virtual screening. However, the choice of protein structure affects the success of a docking based virtual screening very significantly[1].

A crystal structure depicts the protein conformation which is optimally adapted for interaction with one specific ligand[2] and therefore, may not accommodate chemically diverse ligands. Thus, use of any one of the available crystal structures for virtual screening, without consideration of its suitability, could lead to significant bias in the virtual screening. The obvious solution to this problem is to incorporate protein flexibility into the docking program so that chemically diverse ligands can also be accommodated. One approach of incorporating some level of protein flexibility into virtual screening is "soft docking"[3] in which some steric clashes[4] between protein and ligand are relaxed, thereby allowing the overlap of protein and ligand surfaces[3]. Another way is to find a promiscuous protein structure which is able to accommodate a wide range of diverse ligands in the best possible way[5]. Generation of acceptable alternate conformations of the receptor using the predetermined rotamer library is also used to address the issue of receptor flexibility[6]. An ensemble of protein structures[7] has also been used. Each of these methods, however, is associated with a set of limitations[8], which make the choice of a method to address the protein flexibility issue more difficult.

Cyclin-dependent kinase 2 (CDK2) represents one such flexible kinase. It is an enzyme involved in the regulation of the cell cycle. Specific cyclin/CDK2 complex is involved in the cell cycle at two different phases[9]. It complexes to its regulatory protein Cyclin E, undergoes a series of conformational changes followed by phosphorylation to yield a fully active complex[10]. This complex helps transition from G1 to S phase while binding of CDK2 to Cyclin A helps progression through the S phase[9,10]. Excessive production of CDKs or insufficient production of endogenous cyclin-dependent kinase inhibitors is reported to disrupt the normal regulation of cell cycle leading to cancer[9]. The effects of CDK2 inhibitors on the cell cycle and their potential value in the treatment of cancer have been extensively studied[11]. Therefore, development of small molecule inhibitors, which target CDK2 for the treatment of cancer, has gained considerable interest.

[a] Department of Pharmaceutical Chemistry, Prin. K. M. K-undnani College of Pharmacy, Cuffe Parade, Mumbai - 400 005, Maharashtra, India

* For Correspondence E-mail: urmilajjoshi@hotmail.com, urmila.joshi1365@gmail.com

https://doi.org/10.53879/id.56.06.11592

Multiple crystal structures have been deposited in the protein data bank for CDK2 complexes with chemically diverse inhibitors. The inhibitors belong to the chemical classes such as purines[12,13], pyrimidines[14,15], oxindoles[16,17], indenopyrazoles[18,19], benzodipyrazoles[20], quinazolines[21], aminopyrazoles[22,23], aminothiazoles[24] , imidazopyridines[25,26] and flavonoids[27].

Several studies in the past have reported use of multiple protein structures available for CDK2, in an attempt to address the issue of receptor flexibility[28]. These include studies done by Barril (2005)[29], Thomas (2006)[30] and Duca (2008)[31, 32]. These studies were done to address many questions including effect of flexible docking and use of multiple cavities on the docking poses and binding affinities, effect of different softwares[29-32] and methodologies in comparing various crystal structures, and possibility of use of one single structure for docking and screening when multiple conformationally distinct structures are available. All these reports indicate that since a protein cavity is optimally adapted to accommodate a specific ligand, it may be biased for ligands of that chemical class and may not accommodate other ligands. Two of these studies[30-31] recommended GLIDE to be better software than GOLD. The results regarding use of multiple conformations of a receptor for docking-based studies were not consistent in all the previously published papers[29,30]. Docking into multiple protein conformations was reported to produce over-fitting artefacts rather than actual improvements in scoring accuracy[31, 32]. None of the papers recommend a single protein conformation as the most suitable for docking-based virtual screening.

Availability of a number of crystal structures of CDK2, co-crystallised with chemically distinct ligands have multiplied in the past 10 years since the publication of these studies. This prompted us to explore the crystal structures of CDK2 in an attempt to ascertain the single and most suitable structure for docking-based studies. As indicated in the previous studies[30-31], we decided to use GLIDE for docking as it is recommended to be the best software.

METHODOLOGY

Structure-based Docking

Protein Preparation

There are 382 crystal structures of CDK2 deposited in RCSB PDB, out of which we selected 95 crystal structures as shown in Table I, including monomeric CDK2, CDK2 bound to Cyclin A, and inhibitors; and resolution between 1.50 to 2.7 Å. The selection was based on the criterion of resolution and absence of missing residues in the active site. This was followed by the addition of hydrogen atoms. The hydrogen atoms were then optimized using OPLS 2005 force field and the protein-inhibitor complex was minimized using the impact refinement module. A docking grid was computed for each of the 95 protein structures using co-crystallized ligand as the reference[33, 34].

Ligands and Ligand Preparation

For 95 co-crystallized ligand molecules, cyclin A/CDK2 inhibition data (IC_{50}) was available in the literature. These co-crystallized ligands had IC_{50} values ranging from 0.3 nM to 1000 µM. The ligand molecules were subjected to a conformational search (Macromodel v 10.0) and the lowest energy conformation was used for the docking experiments[33,34].

Molecular Docking

Docking was carried out using extra precision (XP) mode (since it is recommended in earlier studies[30, 32]) of GLIDE (v 5.9) using default parameters implemented in Maestro 9.6, Schrodinger. Every ligand was docked into all 94 non-native crystal structures. The G-score value was calculated by taking into consideration the favourable van der Waals, coulombic, lipophilic and hydrogen-bonding interactions. Penalties for steric and buried polar clashes were also considered. The root-mean-square deviation (RMSD) for each ligand between the docked conformation and the one observed in the crystal structure were calculated[33, 35,36].

RMSD = sqrt ([Σ distance $atom_{ref\ i}$ - $atom_{docked\ i}$]^2/*N*), where *N* is the number of atoms]

Validation using a decoy set

Decoys are molecules that are presumed to be inactive (that is, these are not likely to bind to the target) and are used to validate the performance of a virtual screening workflow. The decoy set of 2074 molecules was used from the public domain database of decoys, *Directory of Useful Decoys* (DUD). The decoys were used after calculating the chemical similarity between the decoys and the actives. A set of 51 chemically diverse known inhibitors of CDK2 and other molecules belonging to the chemical class of co-crystallized ligands was included as actives in the database. This database was used for comparative evaluation of the performance of the crystal structures.

Screening of database using molecular docking

Screening of the database consisting of active and decoy molecules was done using of GLIDE 5.9 with default parameters implemented in Maestro 9.6.

Criteria for performance evaluation

A virtual screening method retrieves 'n' molecules called as 'hits 'from a database consisting of 'N' entries. The 'hit list' consists of compounds that possess activity against a target of interest and are called as 'actives' or 'true positives (TP) and compounds called as 'decoys' or 'false positives'(FP) which lack activity. Active molecules that are not identified by the VS method as 'hits' are defined false negatives (FN), whereas the decoys that are correctly rejected represent the true negatives (TN).

Enrichment Factor (EF)

This descriptor takes into account the improvement of the rate at which hits are retrieved by a VS protocol as compared to a random selection and is calculated by using Equation 1. It measures the performance of a method in consideration with respect to a randomly performing method by calculating the fraction of found known actives in the top x% percentage of the ranked list, compared to the ratio between actives and decoys in the entire database[37,38].

$$EF^{x\%} = \frac{Actives^{x\%}_{Selected} / N^{x\%}_{Selected}}{Actives_{total} / N_{total}} \quad (1)$$

Sensitivity (Se)

It is defined as the ratio of the retrieved true positive compounds (TP) to all active compounds in the database, which is the sum of TP and FN (Equation 2). Sensitivity values can range from 0 to 1, where Se = 0 means that no actives were retrieved from the database and Se = 1 means that the search returned all active compounds[37,38].

$$Se = \frac{TP}{TP + FN} \quad (2)$$

Specificity (Sp)

It is defined as the ratio of the number of true negative compounds (TN) divided by the sum of TN and the number of retrieved false positive compounds (FP) (Equation 3). Specificity ranges from 0 to 1 and denotes the percentage of truly inactive compounds. Sp = 0 indicates that all inactive molecules are selected by error as actives and Sp = 1 represents that all inactive compounds have been correctly rejected[37,38].

$$Sp = \frac{TN}{TN + FP} \quad (3)$$

Receiver Operating Characteristic Curve

The ROC curve method describes the sensitivity (Se) for any possible change of selected molecules as a function of (1-Sp) and considers all Se and Sp pairs for each score threshold. A ROC curve is plotted by setting the score of the active molecule as the first threshold. Afterwards, the number of decoy within this cut off is counted and the corresponding Se and Sp pair is calculated. This calculation is repeated for the active molecule with the second highest score and so forth, until the scores of all actives are considered as selection thresholds. The calculation is done using the post-docking processing script in Maestro 9.6[37,38].

Area under ROC Curve

Another way of interpreting the results of ROC curves was by calculating the area under the ROC curve. The area under the curve (AUC) was calculated as the sum of all rectangles formed by the Se and 1-Sp values for the different thresholds. This calculation was also done using the post-docking processing script in Maestro 9.6 and represented by Equation 4 as given below where threshold Si is the score of the ith active molecule[37,38].

$$AUC = \sum_{i} [(Se_{i+1}) (Sp_{i+1} - Sp_i)] \quad (4)$$

Robust Initial Enhancement (RIE)

The RIE is a descriptor that does not suffer from large value variations if only a small number of actives are investigated. The rank for the i[th] active molecule was related to the number of scored compounds investigated a. To get a weight of approximately 1 for the active molecule which was located at the beginning of the list and to retrieve decreased weights for increasing ranks of the actives, an exponential function described in Equation 5 was utilized[39-41].

$$S = \sum_{i=1}^{actives} \exp(-rank(i)/a) \quad (5)$$

The sum of all weights for all active molecules S was then related to the mean sum <S>, which was derived from calculations where the active molecules got randomly selected ranks. This led to the calculation of final RIE descriptor as represented in Equation 6.

$$RIE = \frac{S}{\langle S \rangle} \quad (6)$$

Boltzmann-enhanced discrimination of ROC (BEDROC)

The aim is to derive a new descriptor that addresses the early recognition problem like the RIE descriptor, but also possesses the advantages of AUC, such as values limited by 0 and 1 or a measurement of the performance above all thresholds. The BEDROC descriptor is a generalized AUC descriptor that includes a decreasing exponential weighting function that focuses on active molecules ranked at the beginning of the ordered list. Equation 7 displays the formula of the BEDROC descriptor[39-41].

$$BEDROC = RIE \times \frac{R_\alpha \sinh(\alpha/2)}{\cosh(\alpha/2) - \cosh(\alpha/2 - aR_a)} + \frac{1}{1 - e^{a(1-R\alpha)}} \approx \frac{RIE}{\alpha} + \frac{1}{1 - e^{a}}, \quad \text{if } \alpha R_\alpha \ll 1 \text{ and } a \neq 0 \tag{7}$$

RESULTS AND DISCUSSION

Selection of crystal structures and the cross-docking approach

A total of 382 crystal structures of CDK2 were deposited in protein data bank. Of these, 320 crystal structures were present in complex with chemically diverse co-crystallized ligands. Some of these structures were monomeric structures were some are bound to cyclins. CDK-2 is reported to bind to different cyclins including cyclin A and cyclin E [9, 10]. Since different cyclins are likely to produce different conformational changes in the CDK2, we focussed our attention on conformational changes induced by cyclin A, and considered the structures for monomeric CDK2 as well as CDK2 bound to Cyclin A. We primarily looked at high resolution ($\leq 2.7A^0$) structures. The second criterion used was an absence of missing residues in the active site of CDK2. Thus, the complexes with missing atoms/residues in the active site were discarded, giving us 95 crystal structures. The third criterion was exploring the possibility of selecting one CDK2 structure from a group of crystal structures bearing same ligand or chemically similar ligands. This attempt was based on the assumption that chemically similar ligands bring about similar conformational changes in the flexible kinase domain. However, we observed that these 95 ligands exhibited a significant diversity in their structures, as confirmed by calculation of Tanimoto coefficients. Therefore, we could not eliminate any crystal structures based on this criterion and continued the study with 95 crystal structures as shown in Table I. These structures

Table I: PDB IDs of 95 crystal structures used in this study

Sr. No.	PDB ID	Sr. No.	PDB ID	Sr. No.	PDB ID	Sr. No.	PDB ID	Sr. No.	PDB ID
1.	1AQ1	20.	1CKP	39.	1DI8	58.	1DM2	77.	1E1V
2.	1FVT	21.	1G5S	40.	1GII	59.	1H1S	78.	1KE5
3.	1KE6	22.	1KE7	41.	1KE8	60.	1KE9	79.	1OI9
4.	1OIQ	23.	1OIT	42.	1P2A	61.	1PF8	80.	1PKD
5.	1PXJ	24.	1PXL	43.	1PXM	62.	1PXN	81.	1PXP
6.	1PYE	25.	1R78	44.	1VYW	63.	1VYZ	82.	1W0X
7.	1Y8Y	26.	1Y91	45.	1YKR	64.	2A4L	83.	2B52
8.	2B53	27.	2B54	46.	2B55	65.	2BKZ	84.	2BPM
9.	2BTR	28.	2BTS	47.	2C5N	66.	2C5Y	85.	2C68
10.	2C69	29.	2C6I	48.	2C6K	67.	2C6L	86.	2C6M
11.	2C6O	30.	2C6T	49.	2DS1	68.	2DUV	87.	2IW6
12.	2J9M	31.	2R3G	50.	2R3J	69.	2R3K	88.	2R3M
13.	2R3N	32.	2R3O	51.	2R64	70.	2UZE	89.	2UZL
14.	2UZN	33.	2VTH	52.	2VTI	71.	2VTJ	90.	2VTL
15.	2VTM	34.	2VTN	53.	2VTO	72.	2VTP	91.	2VTQ
16.	2VTR	35.	2VTS	54.	2VTT	73.	2VU3	92.	2W05
17.	2WIH	36.	2WPA	55.	2WXV	74.	3BHT	93.	3BHU
18.	3DDQ	37.	3EZR	56.	3F5X	75.	3LE6	94.	3NS9
19.	3S2P	38.	3ULI	57.	3WBL	76.	4ERW	95.	4KD1

Table II: Enrichment parameters for decoy set screening using 8 PDB Ids

PDB	ROC-AUC	AUC	RIE	BEDROC			EF (1%)	EF (5%)	No. of different chemical classes retrieved in top 1%
				160	20	8			
1OIT	0.85	0.84	3.33	0.595	0.535	0.617	24	11	4
1P2A	0.80	0.80	5.76	0.341	0.363	0.475	12	7.1	4
1YKR	0.84	0.84	7.59	0.475	0.478	0.587	24	9.4	7
1Y91	0.97	0.96	4.76	0.320	0.300	0.322	14	6.3	6
2BTS	0.84	0.83	7.68	0.519	0.483	0.583	22	9	8
2C6O	0.80	0.79	6.95	0.517	0.438	0.530	20	8.6	7
2W05	0.85	0.84	7.80	0.541	0.491	0.597	20	5.9	6
2WXV	0.92	0.91	11.22	0.784	0.707	0.768	36	14	8

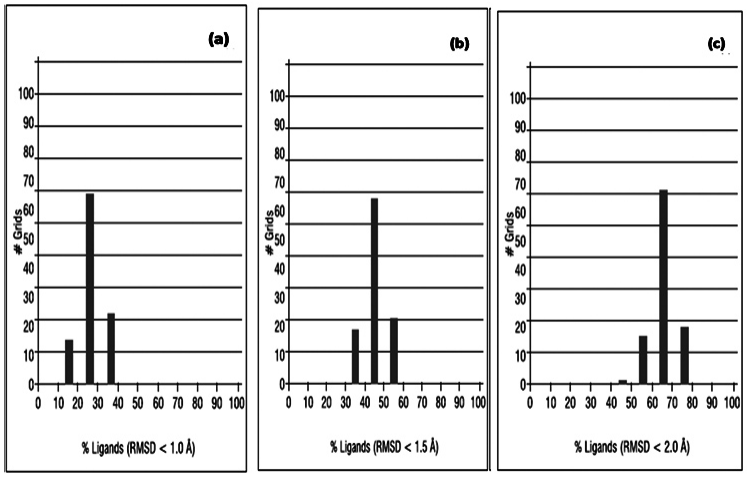

Fig. 1: Centered histogram of a number of grids having a given percentage of ligands docked with RMSD of (a) 1.0 Å or better (b) 1.5 Å or better (c) 2.0 Å or better

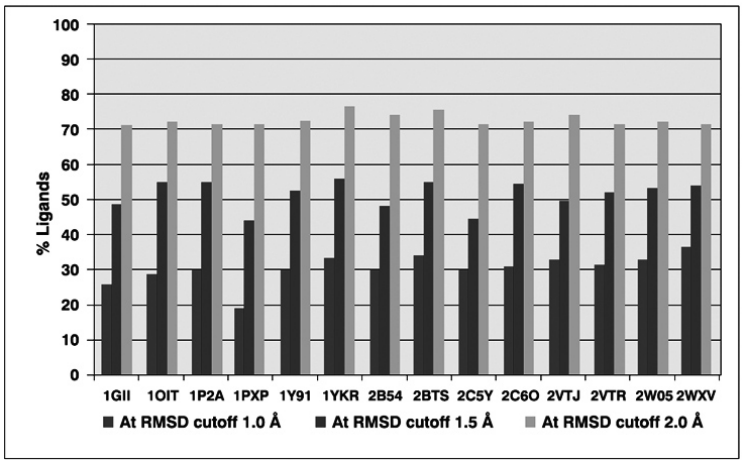

Fig. 2: % of Ligands retrieved for 14 PDB IDs at 1.0 Å, 1.5 Å and 2.0 Å RMSD cut off

represented three different conformational states including the monomeric CDK2, the cyclin A-CDK2 complex and cyclin A-CDK2 complex with phosphorylated Thr 160.

In order to identify the crystal structure of CDK2 for use in docking-based virtual screening, we decided to use cross-docking as our first tool. Since a particular protein conformation is optimally adapted to suit the ligand present in its binding cavity, the cross-docking experiments can investigate how well a given ligand is docked across all protein conformations. We initiated the study by docking each of the 94 ligands into each of 94 non-native protein structures, and the pose thus obtained for each ligand was characterized by its docking accuracy which is represented in terms of RMSD. Docking accuracy is the degree of agreement between the orientation and conformation of a ligand observed in the crystal structure and the pose derived from docking experiments. RMSD values were calculated for the docked pose of each of 94 ligands with respect to their pose attained in the crystal structure as mentioned earlier. The acceptable RMSD value was set to lesser than or equal to 2 Aº. 14 out of 95 crystal structures exhibited docking accuracy for greater than 70% of ligands at this cut off. The remaining crystal structures exhibited docking accuracy values lesser than 65% at RMSD cut off value of $2A^0$ (Fig. 1 and 2).

The 14 top performing crystal structures consist of 2 cyclinA-CDK2 complexes and 12 monomeric structures of CDK2. Additionally, with the exception of 1OIT, all other top performing crystal structures are deposited after the paper published by Barril[29] and Thomas[30]. We could not

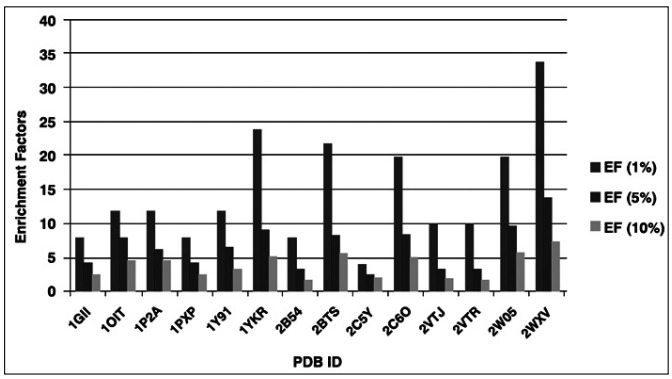

Fig. 3: Enrichment Factors (EF) at percentage of screened ranked dataset for Structure-based Molecular Docking

compare Duca's work[31, 32] in a similar way as that paper reports use of in-house crystal structures, details not being available.

The 14 crystal structures that exhibited a good docking accuracy were subjected to decoy set validation in the second part of the study for selection of crystal structure/s for virtual screening. The results of screening by docking were analyzed using enrichment factors (EF) and the results are depicted in Fig. 3. The cross docking studies indicated ability of the crystal structures to accommodate chemically diverse ligands whereas enrichment studies were considered necessary to check the ability of these crystal structures to retrieve actives from a larger dataset.

Enrichment Calculations using DUD set validation approach

Out of the 14 crystal structures, 8 crystal structures, viz. PDB id 1OIT, 1P2A, 1Y91, 1YKR, 2BTS, 2C6O, 2W05 and 2WXV possessed the enrichment factor of >10 at 1% of the ranked database. The highest values of enrichment factor were obtained in docking based screening using crystal structure with PDB id 2WXV. Crystal structures with PDB accession ids 1OIT and 1YKR performed similarly at 1% (EF=24). 2BTS performed slightly better than 2C6O and 2WO5. 1P2A performed the worst amongst all the crystal structures. The enrichment factor values at 5% of ranked database indicated that 2WXV performed the best at this level too and, was followed by 1OIT. Interestingly, the crystal structures with PDB id 1YKR, 1Y91, and 2W05 which performed better at 1% in comparison to 1P2A, failed to do so at 5% of the ranked dataset indicating in greater retrieval of the decoys at this level. Furthermore, out of 9 chemical classes, 2BTS and 2WXV could retrieve the highest number while 1OIT and 1P2A retrieved least for 1% of ranked dataset. We selected 8 crystal structures which were associated with enrichment factor value above 10 at 1%.

The ROC curve was the second classic enrichment metric used in this study owing to their ability to provide visual as well as a numerical summary of the screening behavior[37,38]. We plotted ROC curves for the selected 8 crystal structures. The ROC curves are shown in Fig. 4.

An ideal ROC curve will start from zero and extend vertically on X-axis till it reaches the value of 1 and then extend rightward indicating complete sensitivity and specificity. The ROC curve for crystal structure with PDB id 1Y91 depicts performance, which is poor in comparison to a crystal structure that will perform random as it crosses the line for random performance. The ROC curve for PDB id 1Y91 begins at the origin zero and extends upward on X-axis till a 0.18 indicating retrieval of actives only. Thereafter, the curve begins to shift slightly right and up indicating that decoys are also being retrieved. At x=0.3, the curve moves only rightwards indicating that only decoys are being retrieved. The next three points on the curve indicate retrieval of actives followed by retrieval of decoys. The last active was retrieved at x=0.4 and the curve terminated abruptly indicating that all 51 actives were not retrieved. A visual inspection of the eight ROC curves also indicated that 2WXV performed better than the other crystal structures.

Since the protein 1Y91 the last active was retrieved at x=0.65, y=0.4 for 1Y91, we investigated this particular case in detail. The kinase domain of CDK-2 is well known for its conformational flexibility like any other kinase domain. Different ligands which bind to the ATP binding site of the kinase domain of CDK-2 are reported to induce different conformational changes[42] in the kinase domain, resulting in either narrowing or widening of the ATP cleft. Binding of 1Y91 ligand to its native protein causes narrowing of the ATP binding cleft whereas widening is seen in the other 7 shortlisted proteins due to the native ligands. Narrowing of the binding site allows the 1Y91 ligand to adopt a U-shaped conformation, thereby enabling it to form a salt-bridge with Asp145, located deeper inside the cleft. Our observation indicated that ligands containing a basic primary amino group only are able to form the salt-bridge with Asp145 in addition to a hydrogen bond with Asn132. We observed that the ligands devoid of a basic primary amino group were not able to form the H-bond.

Narrowing of the pocket also restricts the conformational space of the ligands attempting to bind to the protein 1Y91. Most of the active ligands contain two or more heterocyclic

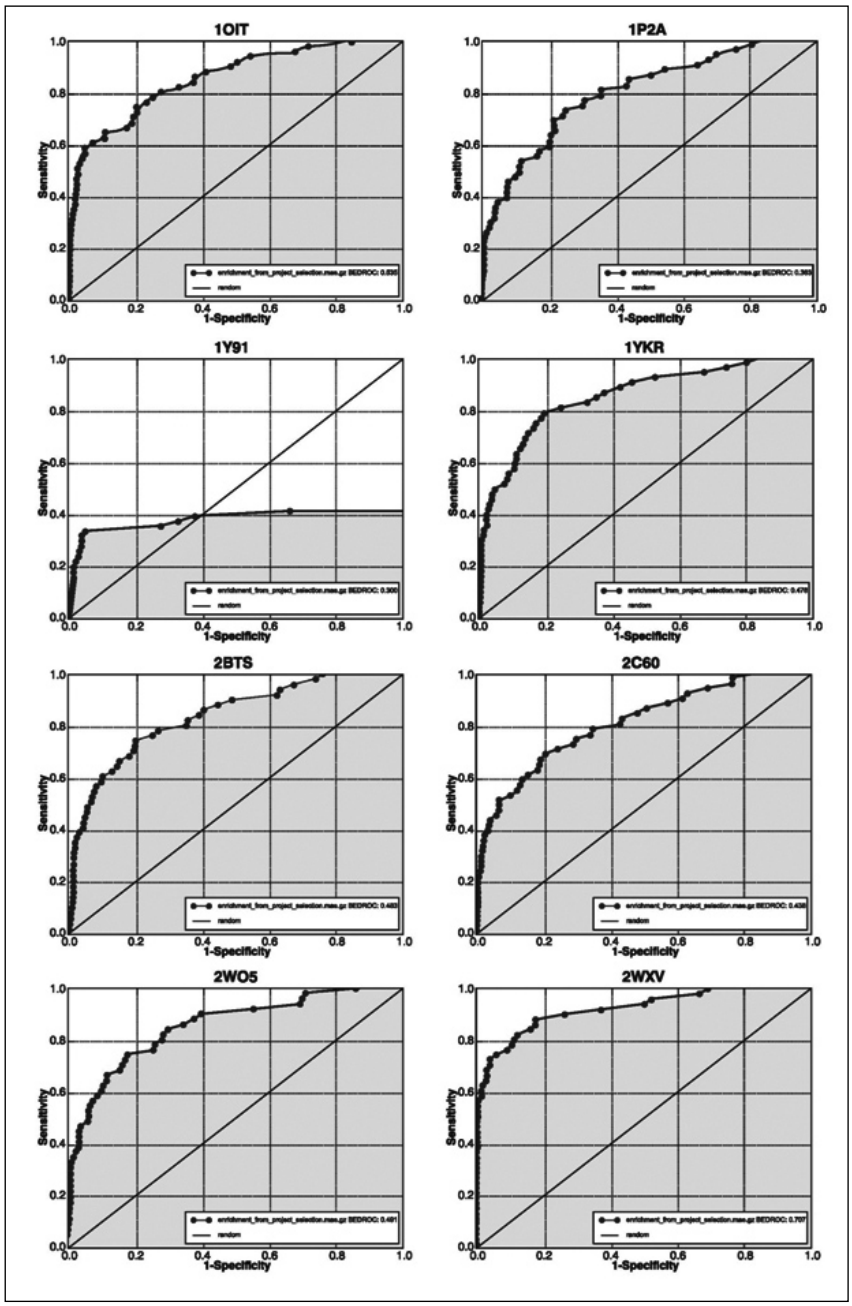

Fig. 4: ROC curves for screened ranked dataset for 8 PDB ids

rings in their structures. Presence of a flexible linker as well as optimum length of the molecules allows the ligands to adopt a suitable conformation so as to fit in the narrow pocket and form interactions with Asp145. Absence of these two structural features also resulted in inability of many actives to dock into the protein 1Y91.

The next enrichment parameter related to ROC and used in the present study was the area under curve (ROC-AUC). Table II shows AUC values of the selected 8 crystal structures. All the crystal structures showed an AUC value of >0.8 indicating a performance superior to random distribution for which AUC values lie between 05-1.0[37-38].

It may be noted that the in one of the earlier studies by Barril and others[29], the classical metric viz. enrichment factor was only considered for the screening performance evaluation. Further the enrichment studies have not been performed by Thomas[30] and Duca[31,32]. However, comparing the values of enrichment factors at the various percentages of the ranked database may indicate early recognition of actives, but it cannot judge the goodness of ranking. Additionally, the ROC curve and the AUC also do not take this into consideration. Thus, in addition to enrichment factor, ROC and AUC; advanced enrichment parameters such as Robust Initial Enhancement (RIE) and Boltzmann-Enhanced Discrimination of ROC (*BEDROC*) were calculated to address the issue of an early recognition that is not considered by classical enrichment metrics[37,38,40,41].

RIE quantitatively indicates the ability of a ranking method to recognize the actives earlier than the decoys. Table II indicates the RIE values of the 8 selected crystal structures of CDK-2. In the present study, the 8 selected crystal structures had an RIE value greater than 1 indicating a performance better than random. However, it was evident from the values mentioned in Table II; 2WXV outperformed other crystal structures with an RIE value 11.22.

The values of next advanced metric viz. BEDROC were calculated at three levels of α-values viz. 8, 20, and 160. The α-value is a value that contributes to θ% of the total score at z% of the rank. The α-value of 20 indicates that 80% of the maximum contribution of actives comes from the first 8% of the list, whereas α-value of 160 indicates that this 80% contribution comes from 1% of the list[39-41]. As indicated by BEDROC values at all three α-values, the performance of 2WXV is the superlative. The high BEDROC values indicated 2WXV could retrieve actives earlier than decoys when compared to other crystal structures. A close observation of the BEDROC values reveals that the crystal structure with PDB id 1Y91 performed worst

with lowest BEDROC value followed by crystal structure with PDB id 1Y91. The remaining five crystal structures had comparable BEDROC values.

CONCLUSION

Although the previous reports stated that no single structure can be best for virtual screening studies, availability of the new crystal structures opened new possibilities prompting us to undertake this work. There was a huge choice for selection of crystal structures of CDK2 present in complex with chemically diverse ligands for docking. 95 crystal structures were subjected to cross-validation studies and 14 were selected. Our cross-docking analysis also revealed that monomeric structures are also able to accommodate ligands of diverse chemical classes apart from the cyclin-CDK2 complexes. Additionally, classical and advanced enrichment parameters were used for further validation and selection of a single crystal structure. Taking into consideration the performance of the protein structures in cross docking and decoy set validation, we propose that 2WXV is able to dock a number of chemically diverse ligands accurately and can detect a greater number of actives at an earlier stage. Therefore, the use of 2WXV as single initial crystal structure in the virtual screening protocol is likely to produce more accurate results than any other single crystal structure.

ACKNOWLEDGEMENTS

University Grants Commission for sanction of Research Grant F. No. 41-745/2012 (SR)]

REFERENCES

1. Erickson, J. A., Jalaie, M., Robertson, D. H., Lewis, R. A., Vieth, M., Lessons in Molecular Recognition : The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy, **J. Med. Chem.**, 2004, 47, 45-55
2. Mengang, X. and Markus, A. L. Significant Enhancement of Docking Sensitivity using Implicit Ligand Sampling, **J Chem. Inf. Model.,** 2011, 51(3), 693–706.
3. Ferrari, A. M., Wei, B.Q., Costantino, L., Shoichet, B. K. Soft Docking and Multiple Receptor Conformations in Virtual Screening, **J Med Chem.,** 2006, *47* (21), 5076–5084.
4. Ramachandran, S., Kota, P., Ding, F., Dokholyan, N. V. Automated Minimization of Steric Clashes in Protein Structures, **Proteins**, 2011, 79(1), 261–270.
5. Barril, X. and Fradera, X. Incorporating protein flexibility into docking and structure-based drug design, **Expert Opin. Drug Discov.,** 2006, 1(4), 1-15.
6. Ding, F., Yin, S., Dokholyan, N. V. Rapid Flexible Docking Using a Stochastic Rotamer Library of Ligands, **J. Chem. Inf. Model,** 2010, 50 (9), 1623–1632.
7. Rao, C. B., Subramanian, J., Sharma, D. S. Managing protein flexibility in docking and its applications, **Drug Discovery Today**, 2009, 14 (7–8), 394-400.
8. Cavasotto, C. N., Orry, A. J. W., Abagyan, R. A. The Challenge of Considering Receptor Flexibility in Ligand Docking and Virtual Screening, **Curr. Comput. Aided Drug Des.,** 2005, 1, 423-440.
9. Malumbres, M., Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. **Nat. Rev. Cancer.,** 2009, *9* (3), 153–166.
10. Yin, M., Guo, B., Panadero, A., Frank, C., Wrzosek, C., Slocum, H. K., Rustum, Y. M. Cyclin E – Cdk2 Activation Is Associated with Cell Cycle Arrest and Inhibition of DNA Replication Induced by the Thymidylate Synthase Inhibitor Tomudex, **Exp. Cell Res.**, 1999, 247, 189–199.
11. Shapiro, G. I. Cyclin-Dependent Kinase Pathways as Targets for Cancer Treatment, **Clin. Oncol.,** 2006, 24 (11), 1770-1783.
12. Oh, C. H., Kim, H. K., Lee, S. C., Oh, C., Yang, B. S., Rhee, H. J., Cho, J. H. Synthesis and biological properties of C-2, C-8, N-9 substituted 6-(3-chloroanilino)-purine derivatives as cyclin-dependent kinase inhibitors. Part II, **Arch. Pharm. (Weinheim),** 2001, *334* (11), 345–350.
13. Otyepka, M., Krystof, V., Havlicek, L., Siglerova, V., Strnad, M., Koca, J. Docking-based development of purine-like inhibitors of cyclin-dependent kinase-2, **J. Med. Chem.**, 2000, *43* (13), 2506–2513.
14. Kim, D. C., Lee, Y. R., Yang, B.-S., Shin, K. J., Kim, D. J., Chung, B. Y., Yoo, K. H. Synthesis and biological evaluations of pyrazolo[3,4-d]pyrimidines as cyclin-dependent kinase 2 inhibitors, **Eur. J. Med. Chem.**, 2003, 38, 525–532.
15. Williamson, D. S., Parratt, M. J., Bower, J. F., Moore, J. D., Richardson, C. M., Dokurno, P., Cansfield, A. D., Francis, G. L., Hebdon, R. J., Howes, R., et al. Structure-guided design of pyrazolo[1,5-a]pyrimidines as inhibitors of human cyclin-dependent kinase 2, **Bioorg. Med. Chem. Lett.**, 2005, 15 (4), 863–867.
16. Dermatakis, A., Luk, K. C., DePinto, W. Synthesis of potent oxindole CDK2 inhibitors, **Bioorg. Med. Chem.,** 2003, 11 (8), 1873–1881.
17. Bramson, H. N., Holmes, W. D., Hunter, R. N., Lackey, K. E., Lovejoy, B., Luzzio, M. J., Montana, V., Rocque, W. J., Rusnak, D., Shewchuk, L., et al. Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): Design, synthesis, enzymatic activities, and X-ray crystallographic analysis, **J. Med. Chem.,** 2001, 44 (25), 4339–4358.
18. Nugiel, D. A., Etzkorn, A. M., Vidwans, A., Benfield, P. A., Boisclair, M., Burton, C. R., Cox, S., Czerniak, P. M., Doleniak, D., Seitz, S. P. Indenopyrazoles as novel cyclin dependent kinase (CDK) inhibitors [1], **J. Med. Chem.**, 2001, 44 (9), 1334–1336.
19. Yue, E. W., DiMeo, S. V., Higley, C. A., Markwalder, J. A., Burton, C. R., Benfield, P. A., Grafstrom, R. H., Cox, S., Muckelbauer, J. K., Smallwood, A. M., et al. Synthesis and evaluation of indenopyrazoles as cyclin-dependent kinase

inhibitors. Part 4: Heterocycles at C3**, Bioorg. Med. Chem. Lett.**, 2004, 14 (2), 343–346.

20. D'Alessio, R., Bargiotti, A., Metz, S., Brasca, M. G., Cameron, A., Ermoli, A., Marsiglio, A., Polucci, P., Roletto, F., Tibolla, M., et al. Benzodipyrazoles: A new class of potent CDK2 inhibitors, **Bioorg. Med. Chem. Lett.**, 2005, 15 (5), 1315–1319.
21. Ecki, T. M., JohnsonSiel, T. L., Liu, J., Muckelbauer, J. K., Grafstrom, R. H., Cox, S., Boylan, J., Burton, C. R., Chen, H., Smallwood, a, et al. Quinazolines as cyclin dependent kinase inhibitors, **Bioorg. Med. Chem. Lett.**, 2001, 11 (9), 1157–1160.
22. Pevarello, P., Brasca, M. G., Orsini, P., Traquandi, G., Longo, A., Nesi, M., Orzi, F., Piutti, C., Sansonna, P., Varasi, M., et al. 3-Aminopyrazole inhibitors of CDK2/cyclin A as antitumor agents. 2. Lead optimization, **J. Med. Chem.**, 2005, 48 (8), 2944–2956.
23. Pevarello, P., Brasca, M. G., Orsini, P., Traquandi, G., Longo, A., Nesi, M., Orzi, F., Piutti, C., Sansonna, P., Varasi, M., et al. 3-Aminopyrazole Inhibitors of CDK2 / Cyclin A as Antitumor Agents . 1 . Lead Finding, **J. Med. Chem.**, 2004, 47, 3367-3380.
24. Vulpetti, A., Casale, E., Roletto, F., Amici, R., Villa, M., Pevarello, P. Structure-based drug design to the discovery of new 2-aminothiazole CDK2 inhibitors, **J. Mol. Graph. Model.,** 2006, 24 (5), 341–348. (24) 17
25. Hamdouchi, C., Zhong, B., Mendoza, J., Collins, E., Jaramillo, C., De Diego, J. E., Robertson, D., Spencer, C. D., Anderson, B. D., Watkins, S. A., et al. Structure-based design of a new class of highly selective aminoimidazo[1,2-a] pyridine-based inhibitors of cyclin dependent kinases, **Bioorg. Med. Chem. Lett.**, 2005, 15 (7), 1943–1947.
26. Jaramillo, C., de Diego, J. E., Hamdouchi, C., Collins, E., Keyser, H., Sanchez-Martinez, C., del Prado, M., Norman, B., Brooks, H. B., Watkins, S. A., et al. Aminoimidazo[1,2-a] pyridines as a new structural class of cyclin-dependent kinase inhibitors. Part 1: Design, synthesis, and biological evaluation, **Bioorg. Med. Chem. Lett.,** 2004, 14 (24), 6095–6099.
27. Casagrande, F., Darbon, J.-M. Effects of structurally related flavonoids on cell cycle progression of human melanoma cells: regulation of cyclin-dependent kinases CDK2 and CDK1, **Biochem. Pharmacol.**, 2001, 61 (10), 1205–1215.
28. Cavasotto, C. N., Abagyan, R. A., Llc, M., Torrey, N. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. **J. Mol. Biol**. 2004, 337, 209–225.
29. Barril, X., Morley, S. D. Unveiling the Full Potential of Flexible Receptor Docking Using Multiple Crystallographic Structures, **J. Med. Chem**., 2005, 48, 4432-4443.
30. Thomas, M. P., Mcinnes, C., Fischer, P. M., Place, J. L. Protein Structures in Virtual Screening : A Case Study with CDK2, **J. Med. Chem**., 2006, 49, 92-104.
31. Duca, J. S., Madison, V. S., Voigt, J. H. Cross-docking of inhibitors into CDK2 structures. 1, **J. Chem. Inf. Model**., 2008, *48* (3), 659–668.
32. Duca, J. S., Madison, V. S., Voigt, J. H. Cross-docking of inhibitors into CDK2 structures. 2, **J. Chem. Inf. Model**., 2008, *48* (3), 669–678.
33. Akl, M.R., Foudah, A.I., Ebrahim, H.Y., Meyer, S.A., El Sayed, K.A. The marine-derived sipholenol A-4-O-3',4'-dichlorobenzoate inhibits breast cancer growth and motility in vitro and in vivo through the suppression of Brk and FAK signaling, **Mar. Drugs.**, 2014, 12(4), 2282-2304.
34. Van Den Driessche, G., Fourches, D.Adverse Drug Reactions Triggered by the Common HLA-B57:01 Variant: A Molecular Docking Study, **J. Chem. inform**., 2017, 9 (13), 1-17.
35. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, **J Med Chem.**, 2004, 47 (7), 1739-1749.
36. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, **J Med Chem.**, 2004, 47 (7), 1750-1759.
37. Braga, R. C., Andrade, C. H. Assessing the performance of 3D pharmacophore models in virtual screening: how good are they? **Curr. Top. Med. Chem.**, 2013, 13 (9), 1127–1138.
38. Kirchmair, J., Markt, P., Distinto, S., Wolber, G., Langer, T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection - What can we learn from earlier mistakes? **J. Comput. Aided. Mol. Des.**, 2008, 22s (3-4), 213–228.
39. Truchon, J. F., Bayly, C. I. Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem, **J. Chem. Inf. Model.**, 2007, 47 (2), 488–508.
40. Triballeau, N., Acher, F., Brabet, I., Pin, J. P., Bertrand, H. O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4, **J Med Chem.**, 2005, 48 (7), 2534–2547.
41. Empereur, C., Guillemain, H., Latouche, A., Zagury, J. F., Viallon, V., Montes, M. Predictiveness curves in virtual screening, **J. Chem. Inform**., 2015, 1–17.
42. Kontopidis, G., McInnes, C., Pandalaneni, S.R., McNae, I., Gibson, D., Mezna, M., Thomas, M., Wood, G., Wang, S., Walkinshaw, M.D. and Fischer, P.M., 2006. Differential binding of inhibitors to active and inactive CDK2 provides insights for drug design. **Chem. Biol.**, 13(2), 201-211.