A Novel Computational Analysis of Ligand-Induced Conformational Changes in the ATP Binding Sites of Cyclin Dependent Kinases

Jyothi Subramanian, Somesh Sharma, and Chandrika B-Rao*

Cheminformatics, Nicholas Piramal Research Centre, 1 Nirlon Complex, Off Western Express Highway, Goregaon(E), Mumbai-400063, India

Protein kinases in general are known to be very flexible macromolecules. In this article, the conformational plasticity of the ATP binding site in cyclin dependent kinases is analyzed. Movement of the two lysine residues lining the ATP binding site are shown to play a major role in the conformational variability of the site. Linear models are developed to identify and quantify ligand properties that maximally influence the lysine side chain conformations. A few simple properties of the ligands are shown to account for more than 70% of the variation in the lysine conformations. The results are validated using test data and molecular simulation studies. Illustrative applications of the results of this analysis to finding the appropriate crystal structure for molecular docking and binding mode predictions of novel ligands are provided. This work provides a new approach to quantify ligand-induced conformational changes in the active sites of flexible proteins and to find the appropriate crystal structure for docking novel ligands.

Introduction

Cyclin dependent kinases (CDKs) play a crucial role in the regulation of cell division. CDK inhibitors, in particular inhibitors of CDKs 1, 2, and 4, are now being widely explored as antitumor agents.[1,2] With more than 50 crystal structures in complex with various chemical inhibitors, CDK2 is structurally, at present, the best-characterized member of this protein kinase family. The CDKs are folded into the bilobal structure typical of most protein kinases (Figure 1). The smaller *N*-terminal domain is predominantly made up of β-sheets, and the larger *C*-terminal domain consists mainly of α-helices. The ATP binding site is found in the deep hydrophobic cleft between the two lobes. The abundance of crystal structure information on CDK2 makes this target well suited for structure-based drug design.

It is well known that protein kinases are very flexible molecules. The conformational plasticity of protein kinases and the conformational changes that the kinase domains undergo on activation/ inactivation have been reviewed.[4] Apart from the conformational changes on activation, it has also been repeatedly observed that binding of ligands also induces considerable conformational changes in the active site residues.[5] To the best of our knowledge, there are no studies correlating these conformational changes to the properties of the bound inhibitor. In this study, we explore the geometry of the ATP binding site in CDKs and identify ligand properties that are correlated with the conformational changes in the binding site residues.

The vast majority of molecular docking programs currently in vogue take into account ligand flexibility.[6] However, docking methods that also take protein flexibility into account are still in their infancy and are computationally demanding.[7] Currently, a popular approach to deal with protein flexibility during docking is to dock the ligand of interest to an ensemble of protein crystal structures bound to a diverse set of ligands. Studies[8,9] illustrating this method for CDK2 differ in the number of structures necessary to obtain a reasonably good approximation of receptor flexibility. It is also not clear as to how the
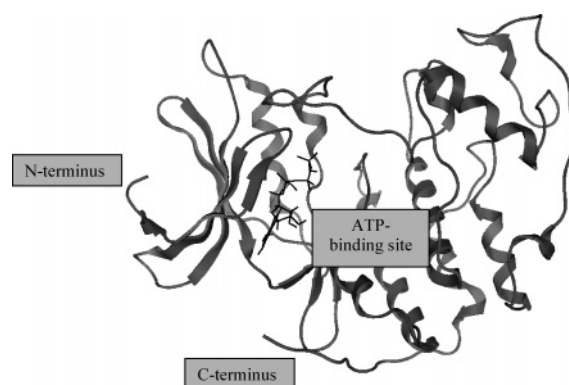
Figure 1. Structure of cyclin dependent kinase-2 bound to ATP (PDB ID: 1HCK[3]). ATP is depicted as a stick model.

results of these multiple dockings can be combined to arrive at the likely binding mode for a query ligand.[8] This method also does not provide the researcher with any idea of the protein conformation that is likely to be induced by a query ligand, particularly if it is novel.

In this study, we illustrate, using CDKs, how the correlation between ligand properties and binding site flexibility can be used to rationally identify the best protein crystal structure for ligand docking and binding mode prediction.

Materials and Methods

Protein and Ligand Structures. The protein−ligand crystal complexes used in this study were obtained from the Protein Data Bank (PDB) (http://www.rcsb.org/pdb).[10] To ensure quality data, only crystal structures with high resolutions were downloaded. Fifteen CDK2 crystal structures with a resolution of at least 2 Å were chosen for this study. To examine the general applicability of the methods to the CDK family, the best-resolved CDK5 crystal structure available, the lone CDK6 crystal structure, and a mutant CDK2 structure, where the ATP binding site of CDK2 had been mutated so that it resembled the CDK4 ATP binding site, were also included in this study. The sequence alignment of these CDKs is shown in Table 1. Only residues within 4 Å radius of any bound ligand is shown in this alignment. The set of protein structures used in this study
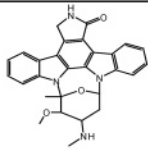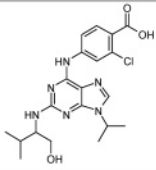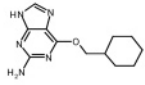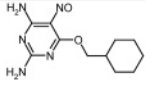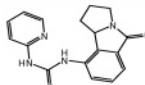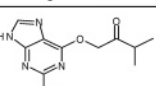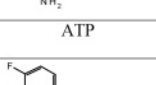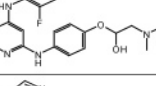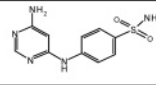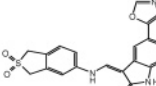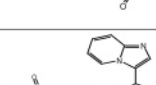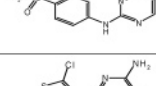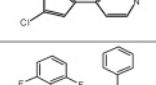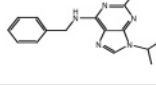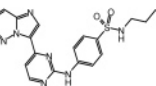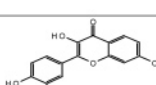
* Corresponding author. Phone: 91-22-30818714. Fax: 91-22-30818036. E-mail: chandrika@nicholaspiramal.co.in.

Table 1. Alignment of the CDK Sequences Used in This Study[a]

residue #	10	11	12	13	14	31	32	33	80	81	82	83	84	85	86	89	131	132	133	134	144	145	146	147	148
CDK2	I	G	E	G	T	A	L	**K**	F	E	F	**L**	H	Q	D	**K**	**Q**	N	L	L	A	D	F	G	L
CDK2 mutant (1GII)	I	G	E	G	T	A	L	K	F	E	H	V	H	Q	D	T	Q	N	L	L	A	D	F	G	L
CDK5 (1UNL)	I	G	E	G	T	A	L	K	F	E	F	C	D	Q	D	K	Q	N	L	L	A	N	F	G	L
CDK6 (1XO2)	I	G	E	G	A	A	L	K	F	E	H	V	D	Q	D	T	Q	N	I	L	A	D	F	G	L

[a] Only residues around 4 Å radius of any bound ligand are shown. The residues studied in detail in this article are indicated in bold.

Table 2. Proteins and Ligands Used in This Study with Crystal Structure Resolutions and Literature References

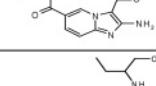| PDB Code	Resolution (Å)	Activation State	Ligand	Reference	PDB Code	Resolution (Å)	Activation State	Ligand	Reference
1AQ1	2.00	Inactive	*(chemical structure)*	16	1JSV	1.96	Inactive	*(chemical structure)*	22
1CKP	2.05	Inactive	*(chemical structure)*	17	1KE7	2.00	Inactive	*(chemical structure)*	23
1E1V	1.95	Inactive	*(chemical structure)*	18	1OIT	1.60	Inactive	*(chemical structure)*	24
1E1X	1.85	Inactive	*(chemical structure)*	18	1PXI	1.95	Inactive	*(chemical structure)*	4
1GII (CDK2 mutant)	2.00	Inactive	*(chemical structure)*	19	1PYE	2.00	Inactive	*(chemical structure)*	25
1GZ8	1.30	Inactive	*(chemical structure)*	20	1UNL (CDK5)	2.20	Active	*(chemical structure)*	26
1HCK	1.60	Inactive	ATP	3	1URW	1.60	Inactive	*(chemical structure)*	27
1H00	1.90	Inactive	*(chemical structure)*	21	1XO2 (CDK6)	2.90	Active	*(chemical structure)*	28
1H0V	1.90	Inactive	*(chemical structure)*	20	2BHE	1.90	Inactive	*(chemical structure)*	29

is listed in Table 2 along with their crystal structure resolutions, the state of the activation loop (active/inactive), and the original references. This table also gives the chemical structures of the ligands to which these proteins are bound.

Protein and Ligand Preparation. The protein and ligand structures were prepared as per standard procedures[11] using MOE 2005-06 software.[12] If the CDK crystal structure contained multiple chains, only one of the chains involved in ligand binding was retained. Solvent and other small molecules were removed. Hydrogens were added to the protein structures. The structures were then energy-minimized using the CHARMM force field[13] after fixing all nonhydrogen atoms to their crystallographic positions. The minimization was carried out to an RMS gradient of 0.01 kcal/mol. From each protein−ligand complex, the ligands were then extracted and any structural errors were corrected. The structures were minimized using the MMFF94x force field[14] to an RMS gradient of 0.01 kcal/mol. These ligand structures were stored in an MOE 2005-06 database.

Binding Site Geometry. The 18 CDK structures (Table 2) were brought to a common frame of reference by aligning all of them to the 1AQ1 protein structure, which was arbitrarily taken as the standard. The alignment was done on the basis of all protein residues using MOE 2005-06 software. A sample of five CDK2 structures that showed the most diversity in the ATP binding region (namely PDB structures 1AQ1, 1H00, 1JSV, 1PXI, and 2BHE) were taken for the alignment. The aligned backbone of the ATP binding cleft in the CDKs and the important residues in this region are shown in Figure 2.

Most of the CDK inhibitors form a hydrogen bond with the backbone of Glu81/Leu83 and additional H bonds with at least one of Gln131, Lys33, or Lys89. The inter-residue distances
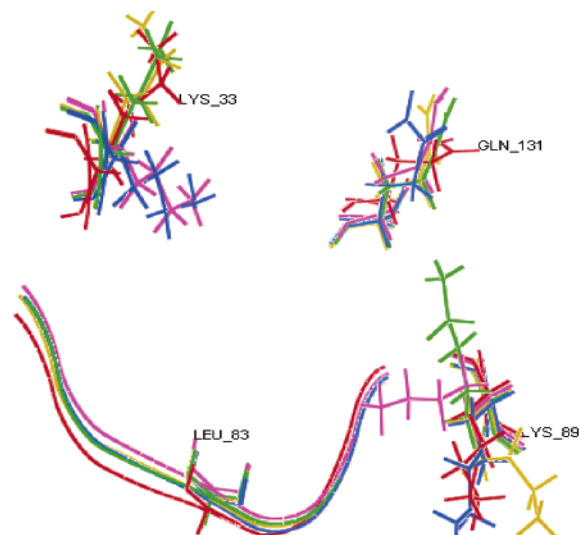
Figure 2. Alignment of the ATP binding cleft in a sample of five CDK2s bound to different small molecule inhibitors. Only the protein backbone and side chains for those residues that form hydrogen bonds with the inhibitors are shown.

were calculated as the distances between heavy atoms of these residues (Figure 3). The heavy atoms considered were the backbone NH of Leu83, terminal side chain NH of Lys33, terminal side chain NH of Lys89, and the backbone O of Gln131. The area of the binding site is defined to be the area of the quadrilateral (ABCD) formed by the line segments joining residues Lys33−Leu83, Leu83−Lys89, Lys89−Gln131, and Gln131−Lys33 (Figure 4). In this article, the geometry of the binding site is defined by the distance-area vector, (Lys33−Leu83, Lys89−Gln131, Area)[37] (see the Results section).

Descriptor Calculations. To avoid any bias in descriptor selection, all 2D descriptors as well as all 3D descriptors that do not depend on the frame of reference of the molecule were calculated for all of the 18 ligands using the descriptor calculation module of MOE 2005-06 software. This set of 241 descriptors[15] includes physical property descriptors, subdivided surface area descriptors, atom and bond count descriptors, Kier and Hall connectivity descriptors, Kappa shape indices, adjacency and distance matrix descriptors, pharmacophore feature descriptors, partial charge descriptors, potential energy descriptors, surface area, volume and shape descriptors, and conformation dependent charge descriptors. To avoid redundancy in this descriptor set, of any two descriptors with more than 90% correlation , one was removed. This resulted in a set of 123 nonredundant descriptors.

Results

Mobility of the Active Site Side Chains. It is evident from Figure 2 that there is a significant amount of conformational variability in the case of the side chains, Lys33 and Lys89, whereas, in the backbone, the positions of Leu83 and Gln131 are more or less invariant. The distance variation between the residues lining the active site among the 18 CDKs (Figure 3) corroborates this observation.

It can be seen from Figure 2 that the conformations of Lys33 can be grouped into two distinct types, trans and gauche. The Lys33 conformation in the structures bound to fragment-like ligands (molecular weight less than 300) is gauche (Lys33−Leu83 distance plot from CDK structures 1PXI to 1E1X in Figure 3). On the other hand, bulkier ligands tend to push this side chain to the trans conformation (Lys33−Leu83 distance plot from CDK structures 1HCK onward in Figure 3). In fact,

the Lys33−Leu83 distance has a fairly strong positive correlation of 0.64 with the molecular weight of the ligands.

It can also be seen from Figure 3 that the Leu83−Lys89, Gln131−Lys33 and Leu83−Gln131 distances are less variable (standard deviations of 1.33, 1.82, and 0.42 Å respectively) than the other distances (Lys33−Leu83, Lys89−Gln131, and Lys33−Lys89 with standard deviations of 2.36, 2.54, and 3.73 Å respectively). In view of this, defining the frame of reference by the positions of Leu83 and Gln131, the movement of Lys33 is measured with respect to Leu83 and that of Lys89 is measured with respect to Gln131. The area of the binding site is seen to vary from 52.77 Å^2 in 1GZ8 (area of quadrilateral A′BC′D) to 117.64 Å^2 in 2BHE (area of quadrilateral ABCD), which is more than a 2-fold difference (Figure 4). The Area provides an intuitive feel for the shape and size of the binding site.

Correlation between Ligand Structure and Side Chain Mobility. To find the properties of the ligands that maximally influence the lysine side chain movements, the correlation coefficients of the side chain distances (Lys33−Leu83 and Lys89−Gln131) and Area with the set of 123 nonredundant ligand descriptors were computed. In the set of CDK2 structures, 1H00 is a CDK2 structure bound to a CDK4 inhibitor. This structure shows a ligand receptor clash with the side chain of Lys89. Hence, this structure and ligand were excluded when computing correlations. All descriptors that had less than 50% correlation with side chain distances and Area were then discarded. Descriptors having more than 50% correlation to the Lys33−Leu83 distance, the Lys89−Gln131 distance, and the Area are shown in Table 3.

To find the optimum linear combination of descriptors that explain most of the variation in the inter-residue distances, multiple linear regression models connecting the inter-residue distances and Area with the descriptors were fitted. The models were derived using stepwise selection of variables. The best linear models are given in Table 4. The relative importance of each descriptor, calculated by taking the ratio of the absolute value of the normalized coefficient of the descriptor to the absolute value of the largest normalized coefficient, is also shown in Table 4. The cross-validated root-mean-square deviations between the actual values and the predictions for the Lys33−Leu83 and Lys89−Gln131 distances and the Area of the active site are 1.30 Å, 1.82 Å, and 11.54 Å^2, respectively.

To analyze the sensitivity of the linear models to CDK4, CDK5, and CDK6 structures in the training set, we computed the linear models after removing the non-CDK2 structures (1GII, IUNL, and 1XO2) from the training set. The best linear models in this case are given in Table 5. The cross-validated root-mean-square deviations between the actual values and the predictions for the Lys33−Leu83 and Lys89−Gln131 distances and the Area of the active site in this case are 1.34 Å, 1.38 Å, and 12.53 Å^2, respectively.

Validation of the Linear Models on a New Set of CDK2 Structures. For validation of the linear models, six CDK2 structures (with resolution less than 2.15 Å) that were not used in the training set were downloaded from the PDB. The set of protein structures used in the validation is listed in Table 6 along with their crystal structure resolutions, the state of the activation loop (active/inactive), and the original references. This Table also gives the chemical structures of the ligands to which these proteins are bound.

These proteins and ligands were prepared as explained in Materials and Methods. Relevant descriptors (Table 4) were computed for the new ligands. The Lys33−Leu83 and Lys89−Gln131 distances and the Area of the active site were predicted
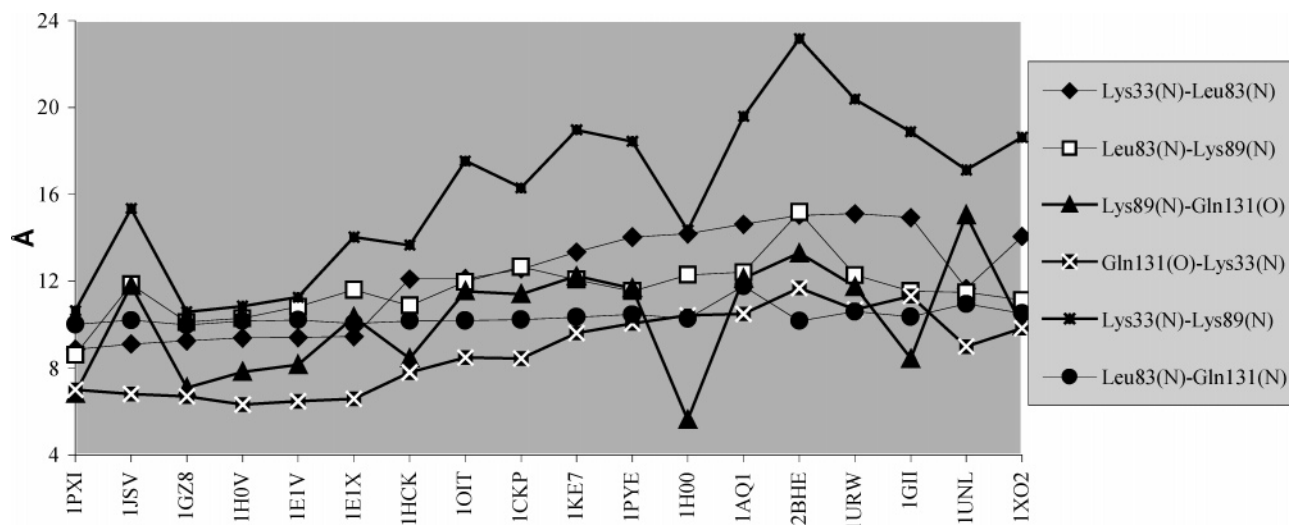
Figure 3. Inter-residue distance variations among the residues lining the ATP binding cleft. The CDK2 proteins are arranged in increasing order of the Lys33(N)−Leu83(N) distance. For 1UNL, 1XO2 and 1GII, see ref 37.
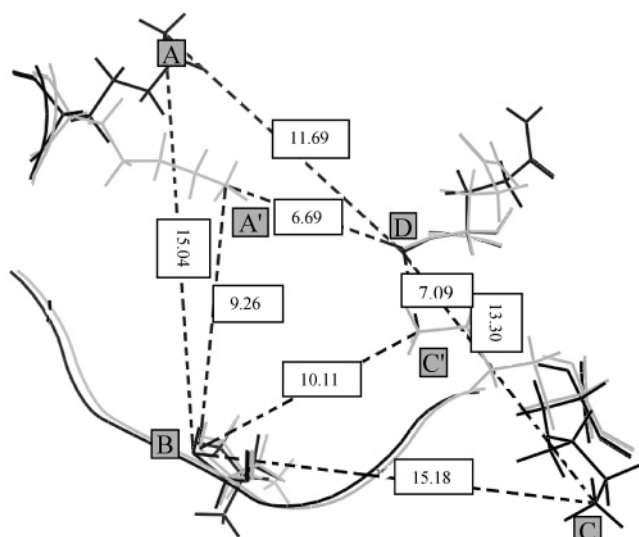
Figure 4. Range of variation in the Lys33 and Lys89 side chain conformations. 1GZ8 (in gray) has the smallest active site area (A′BC′D) of 52.77 Å² and 2BHE (in black) has the largest active site area (ABCD) of 117.64 Å². The distances (in Å) between pairs of atoms are also shown.

using the linear equations given in Table 4. The root-mean-square errors between the actual values and the predictions for the Lys33−Leu83 and Lys89−Gln131 distances and the Area of the active site for this test set are 2.25 Å, 1.49 Å, and 16.38 Å², respectively. The detailed results of the validation are shown in Table 7.

Validation of the Importance of SlogP_VSA7 and PEOE_VSA-0 to Lys89−Gln131 Distance Using a Molecular Simulation Study. This study was carried out to see the dynamics of the Lys89 side chain following binding to two different ligands. Two protein structures, 1E1V and 1JSV, that are bound to fragment-like ligands were considered. However, their potencies to CDK2 vary, with the 1JSV ligand being more potent. The values for SlogP_VSA7 and PEOE_VSA_HYD for the 1E1V ligand are 0.00 and 194.52, respectively. For the 1JSV ligand, the values are 105.00 and 199.43, respectively. From the values of these descriptors, we can infer that the K89−Q131 distance for the 1E1V receptor would be less than that for the 1JSV receptor. This is indeed observed in the crystal

structures; the conformation of the Lys89 side chain in 1E1V is gauche, and the conformation in 1JSV is trans.

To see whether the 1JSV ligand can change the conformation of Lys89 in 1E1V from gauche to trans, the 1JSV ligand was first aligned to the 1E1V receptor. This caused a steric clash of the Lys89 side chain of the 1E1V protein with the 1JSV ligand (Figure 5a). Different rotamers for Lys89 in 1E1V were explored (using MOE 2005-06 software) to see which rotamer is best suited for the new receptor−ligand environment. This was accomplished by computing the change in free energy of solvation between the rotamer−protein complex, and the sum of energies of free protein (without rotamer) and free rotamer. The scoring function here also took into account the changes in the types of surface area contacts. This change in the free energy was added to the torsional strain energy of the rotamer, and the sum was denoted by ΔG. The rotamers were ranked in ΔG ascending order.

Lys89 of the 1E1V receptor was then mutated with the rotamer having the least value of ΔG. To relieve any resultant strain in the protein, energy minimization of the active site was carried out using MOE 2005-06 software. As seen from Figure 5b, the resulting side chain conformation of Lys89 closely resembles the native conformation found in 1JSV. Thus, our simple rotamer exploration of the Lys89 side chain, with the aim of minimizing the steric clash caused by manual placement of the 1JSV ligand in the 1E1V receptor results in a conformation of the 1E1V receptor, which matches the conformation predicted by our model. This provides additional proof that a combination of ligand hydrophobicity characteristics, SlogP_VSA7 and PEOE_VSA_HYD, influences the conformation of Lys89 (gauche or trans).

Application to Ligand Docking. The results of this work are applicable to docking and binding mode predictions. For example, assume that the 1R78 crystal structure is unknown and that we are interested in predicting the most suitable crystal structure to study the binding of the 1R78 ligand. The predicted Lys33−Leu83 distance for this ligand is 12.31 Å, the predicted Lys89−Gln131 distance is 10.92 Å, and the predicted area of the active site is 93.30 Å² (Table 7). We now compute the Euclidean distance between the vector (12.31, 10.92, 93.30) and the actual distance−area vectors for all of the proteins in the training set. These Euclidean distances show a wide variation, starting from 1.89 in the case of 1PYE and going up to 40.83 in the case of 1GZ8 (Table 8). 1PYE would be the most suitable

Table 3. Descriptors with More than 50% Correlation to the Lys33–Leu83 Distance, the Lys89–Q131 Distance, and the Area of the Active Site

correlation	descriptors correlated with the Lys33–Leu83 distance	descriptors correlated with the Lys89–Gln131 distance	descriptors correlated with the area of the active site
(0.75, 1]	E_vdw, SlogP_VSA7	−	SlogP_VSA7, E_vdw, PEOE_VSA-0, a_aro
(0.6, 0.75]	PEOE_VSA-0, PEOE_VSA_NEG, chi0v_C, weinerPol, a_aro Weight, SlogP_VSA4, ASA_H	a_aro, SlogP_VSA7, chi0v_C, ASA_H, PEOE_VSA_HYD, PEOE_VSA_NEG	chi0v_C, weinerPol, ASA_H, PEOE_VSA_NEG, PEOE_VSA+5, SMR_VSA5
(0.5, 0.6]	PEOE_VSA+5, std_dim2, ASA, PEOE_VSA+1, diameter, ASA-, ASA+, b_double	E_vdw, SMR_VSA5, ASA, PEOE_VSA-0	SlogP_VSA4, Q_VSA_FHYD, PEOE_VSA+1, weight, E_stb, E, PEOE_VSA_HYD, ASA, std_dim2, ASA-
[−0.6, −0.5)	Q_VSA_FPPOS, Q_RPC+	logS, Q_VSA_FPPOS	balabanJ, logS, Q_RPC-, PEOE_VSA-3
[−0.75, −0.6)	vsa_don, balabanJ, Q_RPC-	−	Q_VSA_FPPOS, vsa_don
[−1, −0.75)	−	−	−

Table 4. Best Linear Models for Inter-Residue Distances and the Area of the Active Site Based on Correlated Descriptors[a]

best linear model[b]	$R^2(q^2)$	relative importance
K33−L83 = 9.357 + 0.083*E_vdw − 0.086*vsa_don + 0.023*PEOE_VSA-0	0.78 (0.69)	E_vdw (1.00), vsa_don (0.67), PEOE_VSA-0 (0.60)
K89−Q131 = 5.953 + 0.017*SlogP_VSA7 + 0.012*PEOE_VSA_HYD	0.54 (0.38)	SlogP_VSA7 (1.00), PEOE_VSA_HYD (0.82)
Area = 41.893 + 0.126*SlogP_VSA7 + 0.559*E_vdw + 0.182*PEOE_VSA-0	0.80 (0.68)	SlogP_VSA7 (1.00) E_vdw (0.90) PEOE_VSA_0 (0.64)

[a] Stepwise selection of descriptors was done to remove redundant descriptors. [b] E_vdw represents the van der Waals component of the potential energy; vsa_don represents the sum of the van der Waals surface area of hydrogen-bond donors; PEOE_VSA-0 represents the van der Waals surface area of atoms with charges in [-0.05, 0.00); SlogP_VSA7 represents the van der Waals surface area of atoms with logP in (0.25, 0.30], and PEOE_VSA_HYD represents the total hydrophobic van der Waals surface area.

Table 5. Best Linear Models for Inter-Residue Distances and the Area of the Active Site after Removing the Mutant CDK2, CDK5, and CDK6 Structures

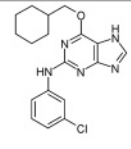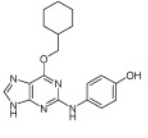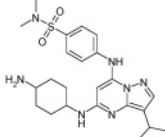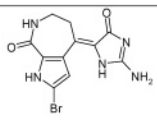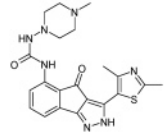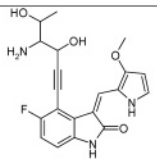| best linear model[b]	$R^2(q^2)$	relative importance
K33−L83 = 10.042 + 0.084*E_vdw − 0.111*vsa_don + 0.018*PEOE_VSA-0	0.81 (0.69)	E_vdw (1.00), vsa_don (0.69), PEOE_VSA-0 (0.46)
K89−Q131 = 7.851 + 0.016*SlogP_VSA7 + 0.026*PEOE_VSA-0	0.69 (0.58)	SlogP_VSA7 (1.00), PEOE_VSA-0 (0.70)
Area = 43.403 + 0.179*SlogP_VSA7 + 0.583*E_vdw	0.77 (0.66)	SlogP_VSA7 (1.00), E_vdw (0.66),

protein for docking the 1R78 ligand because the predicted ligand-induced modifications in the protein binding site are closest to the 1PYE structure. The alignment of the ATP binding sites of 1PYE and the actual 1R78 protein crystal structures is illustrated in Figure 6 and shows good overlap of the two lysine side chains.

Discussion

Proteins in general and kinases in particular are known to be highly flexible macromolecules.[4] The conformational changes occurring upon the activation−deactivation of kinases are well understood. However, although it has long been known that inhibitor binding also leads to significant conformational changes in the residues lining the active site, the quantification of these changes had not been attempted so far.

In this article, we have shown that the movements of the two lysine residues, Lys33 and Lys89, lining the ATP binding site, are major determinants of the geometry of the binding site. The importance of these lysine residues has also been demonstrated through docking experiments.[9] The authors have observed that it is these lysine side chain conformations in the ATP binding pocket that have a major influence on the accuracy of the binding mode predictions for CDK2. Even though other residues such as Asp145 and Phe80 project into the binding site, the positions of these side chains were not found to vary significantly among different CDK2 crystal structures. These authors have also shown that the conformational differences in the flexible T-loop is less relevant to small molecule docking because this loop falls outside the defined binding site into which the ligands are docked.[9] Taken together, these results justify

Table 6. Proteins and Ligands Used in the Test Set with Crystal Structure Resolutions and Literature References

PDB Code	Resolution (Å)	Activation State	Ligand	Reference
1H1R	2.00	Active		30
1OI9	2.10	Active		31
1Y91	2.15	Inactive		32
1DM2	2.10	Inactive		33
2B52	1.88	Inactive		34
1R78	2.00	Inactive		35

Table 7. Predictions for the Test Set Data.

ligand	K33−L83 distance (Å)	predicted K33−L83 distance (Å)	K89−Q131 distance (Å)	predicted K89−Q131 distance (Å)	area (Å²)	predicted area (Å²)
1H1R	13.33	13.46	8.41	10.41	85.15	86.64
1OI9	13.32	13.16	11.52	10.95	103.49	84.75
1Y91	13.65	13.15	12.3	12.74	95.46	100.01
1DM2	13.96	9.10	11.07	8.47	94.25	64.92
2B52	14.16	12.68	10.18	11.25	96.68	82.48
1R78	14.38	12.31	11.88	10.92	106.48	93.30

our concentrating only on the ligand-induced variability in the lysine side chains as a measure of the variability in the binding site.

From a large set of ligand descriptors, we have identified the crucial descriptors that are well correlated with the lysine side chain conformational changes in the ATP binding site. It is interesting to note that out of 29 descriptors that show more than 50% correlation, 5 descriptors are adequate to capture the major features of ligand-induced conformational changes in the protein binding site. Our best linear regression equations (Table 4) suggest that E_vdw, vsa_don, and PEOE_VSA-0 have the most influence in the movement of the Lys33 side chain. E_vdw is also correlated to the bulkiness of the molecule. Hence, we can infer that bulky molecules and negatively charged groups in the molecule tend to move the Lys33 side chain to trans, whereas bulky hydrogen-bond donor groups tend to keep the Lys33 side chain in gauche. In the case of Lys89, we see that

(a)

(b)
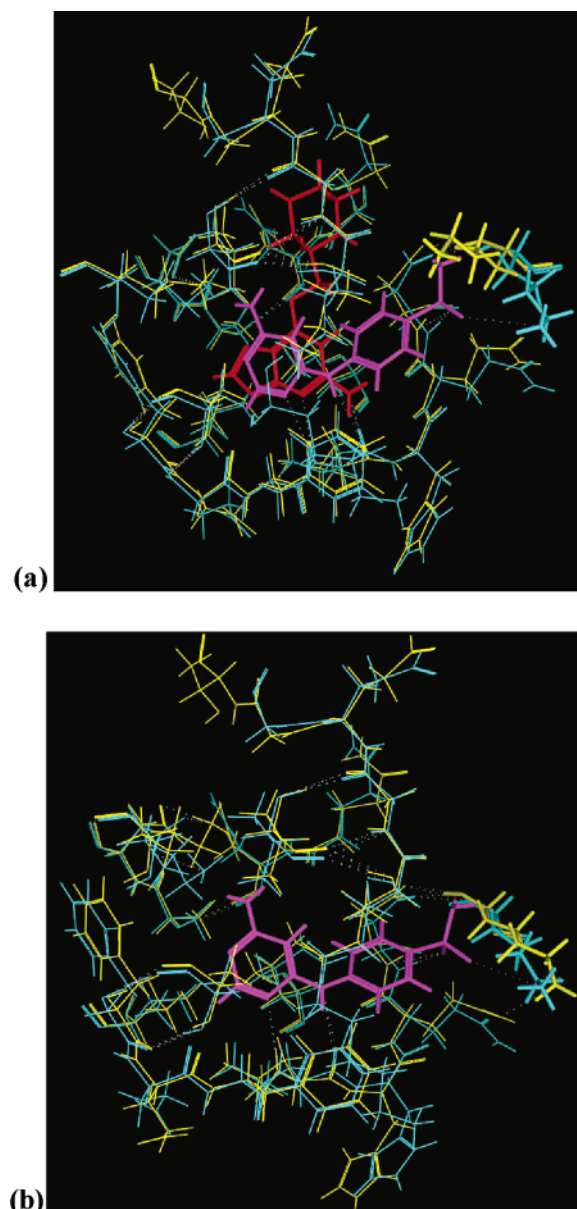
Figure 5. (a) ATP binding sites of CDK2 structures 1E1V and 1JSV are superimposed along with their ligands. Protein 1E1V is in yellow, and its ligand is in red. Protein 1JSV is in cyan, and its ligand is in magenta. The Lys89 side chains of the two proteins are in stick model. We observe a clash of the 1E1V Lys89 side chain with 1JSV ligand. (b) CDK2 structure of 1E1V, after rotamer exploration with 1JSV ligand and minimization, superimposed on the structure of 1JSV. The color coding is the same as that in a. The Lys89 side chain of 1E1V takes on a conformation close to the Lys89 conformation of 1JSV.

a combination of SlogP_VSA7 and PEOE_VSA_HYD have the most influence on the movement of the side chain. The more the number of hydrophobic groups in the molecule, the more likely the trans conformation in the Lys89 side chain. The descriptors SlogP_VSA7, E_vdw and PEOE_VSA-0, a combination of hydrophobicity, charge, and bulk, make a positive contribution to the area of the active site. Interestingly, the most important descriptors for the Area are a combination of the descriptors for the Lys33 and Lys89 movements.

The linear models built using exclusively the CDK2 ligands (Table 5) show R^2 and q^2 values comparable to the models built using the full training set for the Lys33−Leu83 distance and the Area of the active site. But the Lys89−Q131 model shows a marked improvement (R^2 and q^2 of 0.69 and 0.58, respectively, compared to R^2 and q^2 of 0.54 and 0.38, respectively, for the

Table 8. Euclidean Distances (Å) of the Distance−Area Vector Predicted for 1R78 Ligand with the Actual Distance−Area Vectors for the Training Set Proteins in Table 2[a]

protein	Euclidean distance	protein	Euclidean distance
1PYE	1.89	1AQ1	19.74
1KE7	2.82	1E1X	23.02
1XO2	3.56	1HCK	24.59
1OIT	5.45	2BHE	24.60
1GII	5.50	1E1V	35.95
1UNL	8.44	1H0V	38.27
1CKP	11.16	1PXI	40.36
1URW	12.09	1GZ8	40.83
1JSV	15.79		

[a] The active site distances and area of the 1PYE protein structure are closest to the active site distances and area predicted for the 1R78 ligand.
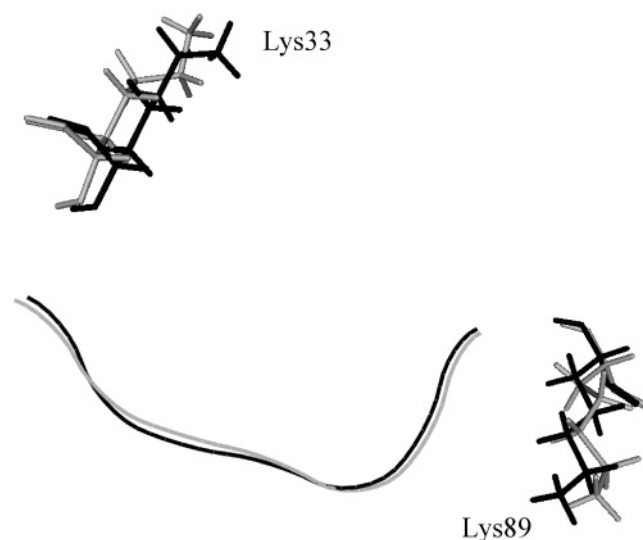
Figure 6. Alignment of the ATP binding sites of 1PYE (in black) and 1R78 (in gray). Only the protein backbone and the backbone and side chains of the two lysine residues are shown.

full training set) for the more homogeneous training set. This may be due to the fact that it is Lys89 that is changed to a Thr residue in the CDK2 mutant and CDK6 structures, whereas Lys33 remains intact. Interestingly, the Lys89 side chain shows more positional flexibility than the Lys33 side chain. Lys89 (or its corresponding residue) shows almost 180° movement across the CDK family for different ligands. This reiterates the importance of this side chain in designing specific ligands for different CDKs.

The linear models are validated using two independent methods; one using recently published data on new protein−ligand complexes and the other using molecular simulations. Comparisons of the theoretical predictions and reported experimental data as well as results of molecular simulations are in good agreement with each other.

Currently, the most popular approach for incorporating protein flexibility during docking is to dock the ligand of interest to an ensemble of protein crystal structures bound to a diverse set of ligands. But, independent experiments with CDK2[8,9] have found different sets of CDK2 structures to be most suitable to approximate the binding mode of ligands. These observations suggest a lack of robustness in the guidelines obtained using the cross-docking approach with different training sets and software packages. Also, docking methods that iteratively dock ligands and model receptor conformational changes are computationally very expensive for practical use in drug discovery settings.[36]

The quantitative approach of fitting multiple linear regression models to important geometrical features of the active site developed by us provides a rational method to identify the best protein binding site conformation for ligand docking and binding mode predictions. We have also illustrated this point with an example (Table 8) and demonstrated the goodness of the prediction (Figure 6). This statistical model-fitting approach can be extended to other receptor−ligand systems involving ligand-induced conformational changes in the receptor. One can also envisage using the models to define the range of variations in side chain positions to be explored during molecular dynamic simulations for incorporating protein flexibility in docking.

As with any statistical model fitting, the linear models developed here are likely to be sensitive to the choice of protein structures in the training set. Hence, we have taken care to choose only the best-resolved crystal structures available for model building. The models would have greater reliability with more and better quality input data. The advantage of this method is that given 3D structure data on a few ligand-bound protein complexes, one can predict the binding site geometry for any other ligand. This would provide more reliable inputs to in silico screening of molecules prior to synthesis or testing in a biological assay.

In summary, currently available docking programs show poor reliability in dealing with proteins showing vast conformational flexibility in the binding site. The present work provides a rational and simple approach for estimating ligand-induced conformational changes in the protein and for using this information in docking studies for lead evaluation and optimization. It, therefore, opens up opportunities for using molecular modeling techniques in some hitherto difficult situations.

References

(1) Fischer, P. M.; Endicott, J.; Meijer, L. Cyclin-Dependent Kinase Inhibitors. *Prog. Cell Cycle Res.* **2003**, *5*, 235−248.
(2) Malumbres, M.; Barbacid, M. Mammalian Cyclin Dependent Kinases. *Trends Biochem. Sci.* **2005**, *130*, 630−640.
(3) Schulze-Gahmen, U.; De Bondt, H. L.; Kim, S. H. High-Resolution Crystal Structures of Human Cyclin-Dependent Kinase 2 with and without ATP: Bound Waters and Natural Ligand as Guides for Inhibitor Design. *J. Med. Chem.* **1996**, *39*, 4540−4546.
(4) Huse, M.; Kuriyan, J. The Conformational Plasticity of Protein Kinases. *Cell* **2002**, *109*, 275−282.
(5) Wu, S. Y.; McNae, I.; Kontopidis, G.; McClue, S. J.; McInnes, C.; Stewart, K. J.; Wang, S.; Zheleva, D. I.; Marriage, H.; Lane, D. P.; Taylor, P.; Fischer, P. M.; Walkinshaw, M. D. Discovery of a Novel Family of CDK Inhibitors with the Program LIDAEUS: Structural Basis for the Ligand Induced Disordering of the Activation Loop. *Structure* **2003**, *11*, 399−410.
(6) Kroval E. M.; Steindl T.; Langer T. Recent Advances in Docking and Scoring. *Curr. Comput. Aided Drug Des.* **2005**, *1*, 93−102.
(7) Carlson, H. A. Protein Flexibility and Drug Design: How to Hit a Moving Target. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447−452.
(8) Barril, X.; David Morley S. Unveiling the Full Potential of Flexible Receptor Docking Using Multiple Crystallographic Structures. *J. Med. Chem.* **2005**, *48*, 4432−4433.
(9) Thomas, M. P.; McInnes, C.; Fischer, P. M. Protein Structures in Virtual Screening: A Case Study with CDK2. *J. Med. Chem.* **2006**, *49*, 92−104.
(10) Berman, H. M.; et al. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(11) Kontoyianni M., McClellan L. M.; Sokol G. S. Evaluation of Docking Performance: Comparative Data on Docking Algorithms. *J. Med. Chem.* **2004**, *47*, 558−565.
(12) Chemical Computing Group. www.chemcomp.com.

(13) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(14) Halgren, T. A. The Merck Force Field. *J. Comput. Chem.* **1996**, *17*, 490−512, 520−641.

(15) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: New York, 2000.

(16) Lawrie, A. M.; Noble, M. E.; Tunnah, P.; Brown, N. R.; Johnson, L. N.; Endicott, J. A. Protein Kinase Inhibition by Staurosporine Revealed in Details of the Molecular Interaction with CDK2. *Nat. Struct. Biol.* **1997**, *4*, 796−801.

(17) Gray, N. S.; Wodicka, L.; Thunnissen, A. M.; Norman, T. C.; Kwon, S.; Espinoza, F. H.; Morgan, D. O.; Barnes, G.; LeClerc, S.; Meijer, L.; Kim, S. H.; Lockhart, D. J.; Schultz, P. G. Exploiting Chemical Libraries, Structure, and Genomics in the Search for Kinase Inhibitors. *Science* **1998**, *281*, 533−538.

(18) Arris, C. E.; Boyle, F. T.; Calvert, A. H.; Curtin, N. J.; Endicott, J. A.; Garman, E. F.; Gibson, A. E.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Lawrie, A. M.; Newell, D. R.; Noble, M. E.; Sausville, E. A.; Schultz, R.; Yu, W. Identification of Novel Purine and Pyrimidine Cyclin-Dependent Kinase Inhibitors with Distinct Molecular Interactions and Tumor Cell Growth Inhibition Profiles. *J. Med. Chem.* **2000**, *43*, 2797−2804.

(19) Ikuta, M.; Kamata, K.; Fukasawa, K.; Honma, T.; Machida, T.; Hirai, H.; Suzuki-Takahashi, I.; Hayama, T.; Nishimura, S. Crystallographic Approach to Identification of Cyclin-Dependent Kinase 4 (CDK4)-Specific Inhibitors by Using CDK4 Mimic CDK2 Protein. *J. Biol. Chem.* **2001**, *276*, 27548−27554.

(20) Gibson, A. E.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Davies, N. J.; Curtin T. G.; Endicott, J. A.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E.; Tucker, J. A.; Whitfield, H. J. Probing the ATP Ribose-Binding Domain of Cyclin-Dependent Kinases 1 and 2 with O(6)-Substituted Guanine Derivatives. *J. Med. Chem.* **2002**, *45*, 3381.

(21) Beattie, J. F.; Breault, G. A.; Ellston, R. P. A.; Green, S.; Jewsbury, P. J.; Midgley, C. J.; Naven, R. T.; Minshull, C. A.; Pauptit, R. A.; Tucker, J. A.; Pease, J. E. Cyclin-Dependent Kinase 4 Inhibitors as a Treatment for Cancer. Part 1: Identification and Optimisation of Substituted 4,6-Bis Anilino Pyrimidines. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2955.

(22) Clare, P. M.; Poorman, R. A.; Kelley, L. C.; Watenpaugh, K. D.; Bannow, C. A.; Leach, K. L. The Cyclin-Dependent Kinases CDK2 and CDK5 Act by a Random, Anticooperative Kinetic Mechanism. *J. Biol. Chem.* **2001**, *276*, 48292−48299.

(23) Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe Jr., R. T.; Harris, P. A.; Hassell, A.; Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana, V.; Rocque, W. J.; Rusnak, D.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Oxindole-Based Inhibitors of Cyclin-Dependent Kinase 2 (CDK2): Design, Synthesis, Enzymatic Activities, and X-ray Crystallographic Analysis. *J. Med. Chem.* **2001**, *44*, 4339−4358.

(24) Anderson, M.; Beattie, J.; Breault, G.; Breed, J.; Byth, K.; Culshaw, J.; Ellston, R.; Green, S.; Minshull, C.; Norman, R.; Pauptit, R.; Stanway, J.; Thomas, A.; Jewsbury, P. Imidazo[1,2-A]pyridines: A Potent and Selective Class of Cyclin-Dependent Kinase Inhibitors Identified through Structure-Based Hybridisation. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3021.

(25) Hamdouchi, C.; Keyser, H.; Collins, E.; Jaramillo, C.; De Diego, J. E.; Spencer, C. D.; Dempsey, J. A.; Anderson, B. D.; Leggett, T.; Stamm, N. B.; Schultz, R. M.; Watkins, S. A.; Cocke, K.; Lemke, S.; Burke, T. F.; Beckmann, R. P.; Dixon, J. T.; Gurganus, T. M.; Rankl, N. B.; Houck, K. A.; Zhang, F.; Vieth, M.; Espinosa, J.; Timm, D. E.; Campbell, R. M.; Patel, B. K.; Brooks, H. B. The Discovery of a New Structural Class of Cyclin-Dependent Kinase Inhibitors, Aminoimidazo[1,2-a]pyridines. *Mol. Cancer Ther.* **2004**, *3*, 1−9.

(26) Mapelli, M.; Massimilinao, L.; Crovace, C.; Seeliger, M. A.; Tsai, L.-H.; Meijer, L.; Musacchio, A. Mechanism of Cdk5/P25 Binding by Cdk Inhibitors. *J. Med. Chem.* **2005**, *48*, 671.

(27) Byth, K.; Cooper, N.; Culshaw, J.; Heaton, D.; Oakes, S.; Minshull, C.; Norman, R.; Pauptit, R.; Tucker, J.; Breed, J.; Pannifer, A.; Rowsell, S.; Stanway, J.; Valentine, A.; Thomas, A. Imidazo[1,2-B]pyridazines: A Potent and Selective Class of Cyclin-Dependent Kinase Inhibitors. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2249.

(28) Lu, H. S.; Chang, D. J.; Baratte, B.; Meijer, L.; Schulze-Gahmen, U. Crystal Structure of a Human Cyclin-Dependent Kinase 6 Complex with a Flavonol Inhibitor, Fisetin. *J. Med. Chem.* **2005**, *48*, 737−747.

(29) Schaefer, M.; Jautelat, R.; Brumby, T.; Briem, H.; Eisenbrand, G.; Schwahn, S.; Krueger, M.; Luecking, U.; Prien, O.; Siemeister, G. From the Insoluble Dye Indirubin towards Highly Active, Soluble Cdk2-Inhibitors. *ChemBioChem* **2005**, *6*, 531.

(30) Davies, T. G.; Bentley, J.; Arris, C. E.; Boyle, F. T.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Hardcastle, I. R.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Tucker, J. A.; Wang, L.; Whitfield, H. J.; Structure-Based Design of a Potent Purine-Based Cyclin-Dependent Kinase Inhibitor, **2002**, *Nat. Struct. Biol. 9*, 745.

(31) Hardcastle, I. R.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Chen, Y.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Jewsbury, P.; Menyerol, J.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Pratt, D. J.; Wang, L.-Z.; Whitfield, H. J. N2-Substituted O6-Cyclohexylmethylguanine Derivatives: Potent Inhibitors of Cyclin-Dependent Kinases 1 and 2. *J. Med. Chem.* **2004**, *47*, 3710.

(32) 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.; Jackson, P. S.; Lockie, A. M.; Murray, J. B.; Nunns, C. L.; Powles, J.; Robertson, A.; Surgenor, A. E.; Torrance, C. J. Structure-Guided Design of Pyrazolo[1,5-a]pyrimidines as Inhibitors of Human Cyclin-Dependent Kinase 2. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 863−867.

(33) Meijer, L.; Thunnissen, A. M.; White, A. W.; Garnier, M.; Nikolic, M.; Tsai, L. H.; Walter, J.; Cleverley, K. E.; Salinas, P. C.; Wu, Y. Z.; Biernat, J.; Mandelkow, E. M.; Kim, S. H.; Pettit, G. R. Inhibition of Cyclin-Dependent Kinases, GSK-3beta and CK1 by Hymenialdisine, a Marine Sponge Constituent. *Chem. Biol.* **2000**, *7*, 51−63.

(34) 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.; Chen, H.; Chang, C.-H.; Trainor, G. L.; Seitz, S. P. Synthesis and Evaluation of Indenopyrazoles as Cyclin-Dependent Kinase Inhibitors. Part 4: Heterocycles at C3. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 343−346.

(35) Luk, K.-C.; Simcox, M. E.; Schutt, A.; Rowan, K.; Thompson, T.; Chen, Y.; Kammlott, U.; DePinto, W.; Dunten, P.; Dermatakis, A. A New Series of Potent Oxindole Inhibitors of CDK2. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 913−917.

(36) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, *49*, 534−553.

(37) In the case of PDB entry 1UNL (CDK5), the residue corresponding to Leu83 is Cys83, and the residue corresponding to Gln131 is Gln130. In the case of PDB entry 1XO2 (CDK6), the residues corresponding to Leu83, Lys89, and Gln131 are Val101, Thr107, and Gln149, respectively. For PDB entry 1GII (ATP binding pocket of CDK2 replaced with CDK4), the residues corresponding to Leu83, Lys89, and Gln131 are Val83, Thr89, and Gln131, respectively. To compute inter-residue distances for these structures, the corresponding atoms were taken. Most importantly, the side chain OH group of the corresponding Thr residue was used in place of the terminal side chain NH of Lys89 in the case of the PDB entries 1XO2 and 1GII.