CHEMMEDCHEM

DOI: 10.1002/cmdc.200500033

Comparative Analysis of the Surface Interaction Properties of the Binding Sites of CDK2, CDK4, and ERK2

Matthew D. Kelly[b] and Ricardo L. Mancera*[a]

Recently developed hydrogen-bonding and hydrophobic analysis algorithms were used to investigate the interaction properties of the ATP binding sites of CDK2, CDK4, and ERK2. We were able to prioritise those hydrogen-bonding groups that are observed to bind the native ATP ligand, as well as to identify other important groups found to bind inhibitors of these enzymes. However, as the hydrogen-bonding groups in the ATP binding sites of these enzymes are fairly well-conserved, we have confirmed that inhibitor selectivity may be predominantly due to differences in either the hydrophobic or steric properties of their binding sites. In particular, the hydrophobic properties of regions outside the specificity surface were observed to provide a rationale for the differences in specificity between various inhibitors to these enzymes. Our method was thus able to identify variations in hydrophobicity. The greater hydrophobicity of certain regions of CDK4 over analogous regions in CDK2 was detectable; likewise, it was possible to distinguish variations in hydrophobicity for regions of CDK2 against those in ERK2, despite the fact that these regions are largely composed of similar residue types.

Introduction

The cyclin-dependent kinases (CDKs) are a family of Ser/Thr kinases that act as regulators of cell progression through consecutive phases of the cell cycle.[1] CDKs themselves are regulated through binding to cyclin protein partners which results in an active heterodimeric complex. The catalytic activity of CDKs is further increased upon phosphorylation of a conserved threonine residue (Thr 161 in CDK1).[2] To date, 13 CDK family members have been identified.[3] CDK family members share similar sizes of between 30 and 40 kDa with approximately 40% sequence identity, and the catalytic core region is structurally conserved across all members.

The precise regulation of CDK activity is essential for the stepwise execution of the many processes required for cell growth and division, including DNA replication and chromosome separation.[4] The cellular mechanisms involved in regulating cell cycle, such as the CDK pathways, are often found to be altered in tumours.[5] Such revelations have led to the investigation of CDK inhibitors as possible cancer therapeutics.[6]

The design of inhibitors specific to a particular protein kinase was originally thought of as an impossible task owing to the high degree of homology shared by the ATP binding domains of these enzymes. This belief was strengthened by the discovery of a number of ATP-competitive protein inhibitors such as staurosporine[7] and flavopiridol[8], which demonstrated little selectivity and inhibited a wide range of kinases. However, the discovery of a potent and relatively specific ATP-competitive inhibitor (Iressa) of the epidermal growth factor class of tyrosine kinases highlighted the plausibility of designing selective inhibitors for these classes of proteins.[9]

The ATP binding site of CDK2 is located in a cleft between the ~80-residue C-helical N-terminal domain and the ~120-residue α-helical C-terminal domain. As mentioned above, CDKs do not become fully active until bound to a cyclin partner and phosphorylated. However, the ATP binding site is still able to bind ATP in the inactive monomeric conformation of CDK2.[10] The crystal structure of monomeric CDK2 with ATP bound reveals important interactions within the active site cleft.[11] In addition to numerous hydrogen bonds, favourable hydrophobic and van der Waals interactions are formed between the enzyme and the adenine ring of ATP.

Whereas crystal structures are available for both CDK2 and ERK2 (extracellular signal-regulated kinase 2), efforts to crystallise CDK4 have been unsuccessful. However, the structure of a CDK4-mimic CDK2 protein has been made available.[12] This protein was generated by replacing the binding site of CDK2 with that of CDK4 using site-directed mutagenesis. The use of this model has led to the generation of a selective CDK4 inhibitor.[12]

Herein we report a comparison of the hydrogen-bonding and hydrophobic properties of the ATP binding sites of CDK2, CDK4, and ERK2 using recently developed computational methods to assess the surface interaction properties of binding

[a] *R. L. Mancera*
Western Australian Biomedical Research Institute
School of Pharmacy and School of Biomedical Sciences
Curtin University of Technology
GPO Box U1987, Perth WA 6845 (Australia)
Fax: (+61) 8 9266 7485
E-mail: R.Mancera@wabri.org.au

[b] *M. D. Kelly*
Department of Pharmacology, University of Cambridge
Tennis Court Road, Cambridge CB2 1EW (UK)

sites.[13,14] Our analyses provide a rationalisation of the specificity and selectivity of known kinase inhibitors consistent with previous structural observations and molecular modelling studies, and gives prominence to the differences in hydrophobic properties that the binding sites of these enzymes have.

Computational Methods

Analyses of the ATP binding cavity of CDK2 were performed on the ATP-bound X-ray crystal structure of CDK2 (PDB code: 1hck). For the comparative analysis of the binding sites of CDK2, CDK4, and ERK2, the atomic coordinates were taken from PDB entries 1aq1, 1gij, and 4erk, respectively. Prior to our analyses, all hydrogen atoms were added to these structures by using the Biopolymer module in Insight 2000 (Accelrys). Water molecules found in the crystal structures were removed prior to our analyses. The choice of the inactive form of CDK2 was made for the sake of consistency with the structure of mimic-CDK4 used. The main difference between the active and inactive form of CDK2 resides is in the conformation of Lys 33, which is oriented toward the binding cavity in the inactive form of the enzyme. As will be described below, this residue was identified as a hydrogen-bonding sitepoint but did not play a role in the interpretation of inhibitor selectivity.

We used a recently developed method to estimate the relative importance of potential hydrogen-bonding groups (sitepoints) in the binding site of a protein.[13] This method uses a *strength-weighted accessible-probability score* (SWAPS) to determine the relative importance of the sitepoints within a binding pocket. For each hydrogen-bonding sitepoint, the SWAPS is calculated as the product of the potential strength of the hydrogen bond that may be formed and the *solvent accessible probability score* (SAPS): SWAPS $= d_{rel} \times$ SAPS. For the hydrogen-bond strength component we used the *relative density* (d_{rel}) of the hydrogen-bonding group, as calculated in the IsoStar database of the Cambridge Crystallographic Data Centre (CCDC).[15] These d_{rel} values represent the propensity of a hydrogen-bonding group to form close (and therefore strong) contacts with a probe group (that is, a hydrogen-bond donor or acceptor). In the case of protein surface groups, d_{rel} values range from 0.3 for weak interactions with guanidino (arginine), indolyl (tryptophan), or charged imidazole (histidine) groups, to 4.9 for interactions with charged carboxylate groups (glutamate or aspartate). The SAPS value represents the hydrogen-bonding solvent accessibility of the sitepoint and is calculated by determining the accessibility of binned probability regions surrounding the atom. Each bin has an associated probability value that represents the likelihood of observing a complementary hydrogen-bonding atom in that position. The greater this value, the greater the contribution of that bin to the final SAPS value. The end result is a normalised value ranging between 0 and 1.0 that represents the accessibility of a sitepoint based on the availability of those regions surrounding the atom that are more favourable for the positioning of a complementary hydrogen-bonding atom. Additional geometric rules were introduced to filter out hydrogen-bonding groups with an unfeasible geometry for interacting with a ligand that results from their being directed away from the binding site. Full details of this method can be found elsewhere.[13]

We also used a recently developed method for the analysis of hydrophobic regions in the binding site of a protein.[14] This method considers not only atom type but also local surface properties such as shape, extent, and crowding when calculating a hydrophobicity score for a given atom on the binding site of a protein. It is known that these factors affect the degree of hydrophobicity of a solvent-exposed nonpolar surface.[16,17] For this purpose, a dot surface is mapped onto a protein cavity with an initial weight assigned to each dot based on the underlying atom type. For each dot on the surface, the contribution from neighbouring dots is calculated, and the final value is projected back onto the underlying protein atoms. This allows the final score of an atom to be not only dependent on its own type but also on that of neighbouring atoms, along with the extent and shape of the surrounding surface. This sensitivity to the local environment allows, for example, a hydrophobic atom surrounded by other hydrophobic atoms to score higher than the same hydrophobic atom surrounded by polar atoms. The method was parameterised by means of a genetic algorithm that optimised the weights of various atom types to increase the ability of the algorithm to identify those regions in a binding site that are more likely to form a strong interaction with a nonpolar group in a ligand. Unlike approaches that calculate hydrophobic molecular interaction fields such as GRID, this method is equally capable of accommodating changes in the shape and extent of an exposed nonpolar surface, so that concave and/or larger surfaces are determined to be more hydrophobic, consistent with current understanding of the hydrophobic effect.[17] Full details of this method can be found elsewhere.[14]

Results and Discussion

Hydrogen-bonding analysis of the ATP binding site of CDK2

Thirty potential hydrogen-bonding sitepoints were identified within the ATP binding cavity of CDK2 (Figure 1). These were defined as all nitrogen (H-bond donor) and oxygen (H-bond acceptor) atoms found within a cut-off distance of 5.0 Å from any ATP atom as found in its crystallographic binding mode (PDB code: 1hck). The 30 potential hydrogen-bonding sitepoints identified in the binding site were analysed with the hydrogen-bonding sitepoint prioritisation method[13] (Figure 2). This analysis demonstrates that the majority of the hydrogen-bonding sitepoints are localised around the triphosphate and ribose regions of ATP. The triphosphate region is dominated by hydrogen-bond donor sitepoints, whereas the ribose region is dominated by hydrogen-bond acceptor sitepoints. To facilitate the analysis of the hydrogen-bonding sitepoints in the ATP binding pocket, the binding site was divided into three regions based on the proximity to each of the three components of the ATP molecule: the adenine ring, the ribose ring, and the triphosphate group.

Six potential hydrogen-bonding sitepoints were identified within the adenine binding region of the ATP binding site of

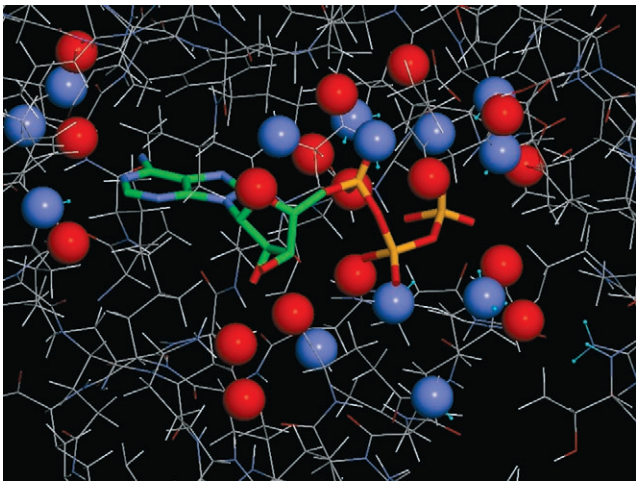


Figure 1. The 30 potential hydrogen-bonding sitepoints in the CDK2 active site within 5.0 Å of the bound ATP molecule (PDB code: 1hck). The potential sitepoints are represented as solid spheres with those in red representing oxygen (H-bond acceptor) atoms and those in blue representing nitrogen (H-bond donor) atoms.

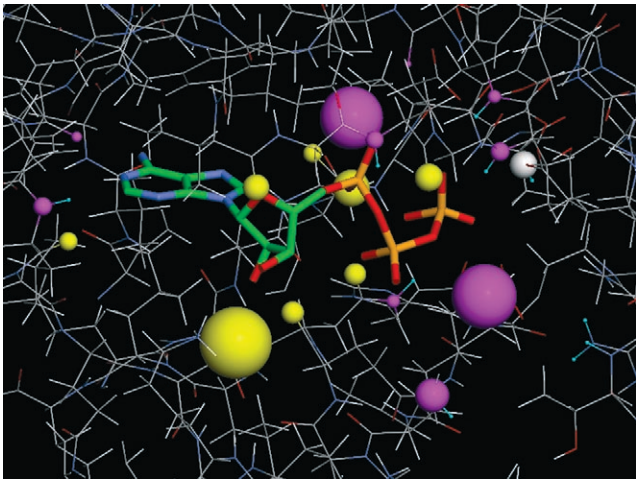


Figure 2. Analysis of the hydrogen-bonding groups in the ATP binding site. Hydrogen-bonding donor, acceptor, and amphiprotic sitepoints are coloured magenta, yellow, and white, respectively. The relative importance of each sitepoint as determined by its computed SWAPS, is directly proportional to the size of the CPK spheres representing them; larger spheres indicate sitepoints that are considered more important because of their potential hydrogen-bond strength and/or solvent accessibility.

CDK2, as listed in Table 1. Three of these sitepoints were filtered out by the H-bond prioritisation method: two were removed as a result of a complete lack of hydrogen-bonding solvent accessibility (that is, a SAPS value of zero) and the other was removed as it is oriented away from the binding cavity. This leaves three remaining potential sitepoints within the adenine region of the site. The highest ranking of these is the backbone N atom of Leu 83, which has almost full solvent accessibility (SAPS $>$ 0.995). In accordance with its assessment as the highest-ranking sitepoint in the region, this nitrogen atom forms a hydrogen bond with ATP. In addition, it is one of a triplet of core sitepoints, at least two of which are targeted by all

Table 1. Calculated scores of the hydrogen-bonding sitepoints found within 5.0 Å of the adenine ring of ATP.

Sitepoint	SAPS	d_{rel}	SWAPS
O Phe 80		Not accessible for hydrogen bonding	
O Glu 81[a]	0.0628	1.4	0.0879
N Glu 81		Not accessible for hydrogen bonding	
N Phe 82		Infeasible geometry	
O Leu 83	0.4816	1.4	0.6742
N Leu 83[a]	0.9953	0.7	0.6967

[a] Sitepoints that form a hydrogen bond with ATP.

known ATP-competitive inhibitors of CDK2.[3] The final two members of this core triplet are the remaining sitepoints scored in this region, one of which is used to form hydrogen bonds by the adenine ring of ATP. Interestingly, it is the lowest scoring of the two remaining sitepoints (backbone O atom of Glu 81) that is used by ATP. However, the reason for its low SWAPS value is a very low hydrogen-bonding specific accessibility (SAPS $<$ 0.063), despite having a strength value greater than the highest-scoring sitepoint in this region (that is, a d_{rel} value of 1.4 in comparison with 0.7 for the backbone N atom of Leu 83). This would suggest that in practice, the low solvent accessibility of the backbone O atom of Glu 81 is overcome if a corresponding ligand group is suitably aligned by sufficient additional interactions between the protein and ligand. The remaining sitepoint (backbone O atom of Leu 83), while not used by ATP, is in fact used by the majority of known CDK2 inhibitors, including olomoucine,[18] roscovitine,[19] and purvalanol B.[20]

Twelve potential hydrogen-bonding sitepoints were identified in the ribose binding region of the ATP binding site of CDK2, as listed in Table 2. Five of these sitepoints were auto-

Table 2. Calculated scores of the hydrogen-bonding sitepoints found within 5.0 Å of the ribose ring of ATP.

Sitepoint	SAPS	d_{rel}	SWAPS
O Ile 10	0.8554	1.4	1.1976
N Gly 11		Infeasible geometry	
O Gly 11		Infeasible geometry	
O Glu 12	0.9268	1.4	1.2975
N Glu 12	0.9088	0.7	0.6362
N Gly 13		Infeasible geometry	
NZ Lys 33	0.5009	4.0	2.0036
OD1 Asp 86		Infeasible geometry	
OD2 Asp 86[a]	0.8554	4.9	4.1915
O Gln 131[a]	0.6743	1.4	0.9440
N Asn 132		Directed away from the ligand	
OD2 Asp 145	0.3845	4.9	1.8841

[a] Sitepoints that form a hydrogen bond with ATP.

matically filtered out by the H-bond prioritisation method as the result of an infeasible geometry for ligand binding. Of the remaining seven sitepoints, the highest ranking is one of the carboxylate O atoms of Asp 86. In fact, this particular sitepoint

is the highest-scoring sitepoint in the entire binding site and, as this would predict, it forms a hydrogen bond with ATP. Furthermore, this sitepoint is observed to form hydrogen bonds with a number of known CDK2 inhibitors including staurosporine,[21] purvalanol B,[20] and compound II.[12] One further hydrogen bond is formed in this region of the binding site with the backbone O atom of Gln 131; this sitepoint is also used by the inhibitors staurosporine and compound II. Five potential hydrogen-bonding sitepoints remain in this region of the binding site. Two of these, namely the side chain amino N atom of Lys 33 and one of the carboxylate O atoms of Asp 145, are instead targeted by the triphosphate group of ATP (Table 3). The three remaining sitepoints, which are not used by ATP (namely the backbone O atom of Ile 10 and Glu 12 and the backbone N atom of Glu 12), are positioned within the binding site in such a way that for ATP to form hydrogen-bonding interactions with them, it would have to sacrifice existing hydrogen bonds, including that with the overall top-ranked sitepoint of the side chain O atom of Asp 86. These three sitepoints, which are not used by the ATP molecule, have also not been observed to form hydrogen bonds with known CDK2 inhibitors.

Table 3. Calculated scores of the hydrogen-bonding sitepoints found within 5.0 Å of the triphosphate group of ATP.

Sitepoint	SAPS	d_{rel}	SWAPS
O Glu 12	0.9268	1.4	1.2975
O Gly 13		Infeasible geometry	
N Gly 13		Infeasible geometry	
N Thr 14[a]	0.9835	0.7	0.6885
OG1 Thr 14[a]	0.7881	1.6	1.2610
N Tyr 15[a]	0.4734	0.7	0.3314
O Gly 16	0.0079	1.4	0.0111
NZ Lys 33[a]	0.5009	4.0	2.0036
OD1 Asp 127		Infeasible geometry	
OD2 Asp 127[a]	0.4436	4.9	2.1736
NZ Lys 129[a]	0.5893	4.0	2.3572
O Gln 131	0.6743	1.4	0.9440
NE2 Gln 131	0.9942	1.7	1.6901
N Asn 132		Infeasible geometry	
OD1 Asn 132[a]	0.4496	2.0	0.8992
ND2 Asn 132	0.1931	0.7	0.3283
OD1 Asp 145		Infeasible geometry	
OD2 Asp 145[a]	0.3845	4.9	1.8841

[a] Sitepoints that form a hydrogen bond with ATP.

Eighteen new potential hydrogen-bonding sitepoints were identified in the triphosphate binding region of the ATP binding site of CDK2, as listed in Table 3 (note that Table 3 also contains some of the sitepoints identified in the ribose binding region, which are shared between the two regions). Five of these sitepoints were filtered out by the H-bond prioritisation method owing to an infeasible orientation for ligand binding. The backbone O atom of Gly 16 may also be removed, as it forms an intramolecular hydrogen bond with the backbone N atom of Gly 13. The backbone O atom sitepoints of Glu 12 and Gln 131 also lie in the ribose region of the binding site, with the latter forming a hydrogen bond with this part of ATP and the former being in competition with the highest-ranking sitepoint, as mentioned above. Of the remaining ten sitepoints, eight form hydrogen bonds with the triphosphate group of ATP. These eight include the four highest-ranking sitepoints in this group, two of which (the side chain N atom of Lys 33 and one of the carboxylate O atoms of Asp 145) also form hydrogen bonds with the inhibitors indirubin-5-sulfonate[22] and hymenialdisine,[23] respectively. The last two remaining sitepoints that do not form hydrogen bonds with ATP (the side chain amido N atoms of Gln 131 and Asn 132) are positioned in such a way that a ligand would have to sacrifice hydrogen bonds with higher-scoring sitepoints to interact with them. These two sitepoints not used by ATP have also not been observed to form hydrogen bonds with known CDK2 inhibitors.

It can be observed that in each of the three regions of the ATP binding site of CDK2, the highest-ranking sitepoint forms a hydrogen bond; this includes the overall highest-ranking sitepoint for the entire binding site. In fact, of the 19 sitepoints not filtered out by the H-bond prioritisation method, the five top-ranking sitepoints are all used by ATP to form hydrogen bonds. It is important to realise that the ranking of the potential sitepoints, coupled with the filtering out of 11 of these sitepoints owing to an infeasible orientation or absence of hydrogen bonding solvent accessibility (SAPS value of zero), has vastly decreased the combinatorics of sitepoint selection for this binding site.

In addition to decreasing the combinatorics of sitepoint selection, the H-bond prioritisation method has also identified further hydrogen-bonding sitepoints not used by ATP that provide potential targets for inhibitor design. For example, the unused sitepoint (backbone O atom on Leu 83) in the adenine binding region of the binding site (Table 1), which scores higher than one of the sitepoints used by ATP, is observed to be involved in the binding of numerous CDK2 inhibitors.[3]

A consensus principal component analysis of molecular interaction fields computed with GRID has been used in the past to describe the three-dimensional electrostatic, steric, and hydrophobic properties of protein kinases.[24] The above-described hydrogen bonding analysis is consistent with contour maps indicating areas of favourable polar interactions in the binding site of CDK2.[24]

Hydrophobic analysis of the ATP binding site of CDK2

The hydrogen-bonding properties of the ATP binding site of CDK2 are well known, and our above analysis has served as a good example to validate our H-bond prioritisation method. On the hand, the hydrophobic properties of this binding site have not been properly rationalised despite their potential role in the specificity and selectivity of designed inhibitors. Molecular field analyses of the hydrophobic properties of protein kinases have clearly revealed the presence of well-defined hydrophobic areas in the binding site.[24] However, such methods use hydrophobic probes with the GRID program which are unable to capture subtle differences in the hydrophobicity that arise from changes in the shape, extent, and crowding of protein surfaces. We thus applied our recently developed method

toward analysis of the hydrophobic properties of binding sites.[15] As expected, this analysis identified the adenine binding region as the most hydrophobic part of the ATP binding site (Figure 3), with the least hydrophobic region being the tri-

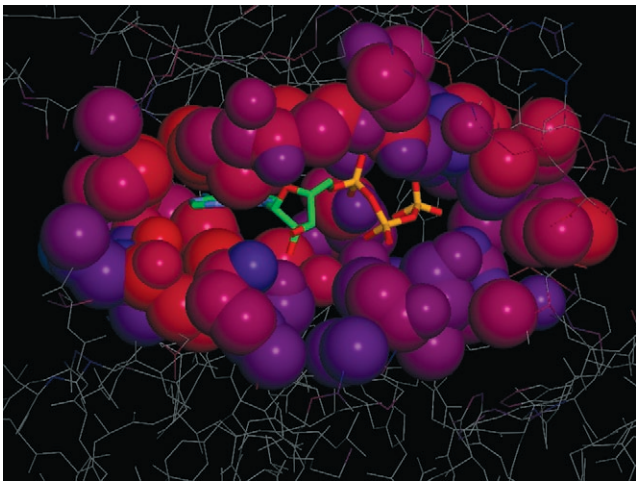


Figure 3. Analysis of the hydrophobic properties of the ATP binding site of CDK2. Binding-site atoms are coloured according to their calculated hydrophobic score, with atoms in red as the most hydrophobic and atoms in blue as the least. ATP is shown in stick representation in its crystallographically bound conformation (PDB code: 1hck).

phosphate binding region. These observations confirm that the relatively nonpolar adenine ring of ATP binds in the most hydrophobic region of the binding site, whereas the highly charged triphosphate group binds to the least hydrophobic region.

In addition to calculating what effectively constitutes a hydrophobic profile of the binding site, this analysis also provides an indication of the type of ligand groups that would enable favourable interactions between an inhibitor and additional regions within the binding cavity. For example, the hydrophobic adenine binding region extends into an area unused by ATP. This region may be targeted in the design of inhibitors through the addition of nonpolar groups to improve binding affinity or, if this region were unique to the CDK2 binding site, it may be targeted to improve the selectivity of inhibitors.[25] We comment on this further below.

Comparative analysis of selected cyclin-dependent kinases

A comparative analysis of the hydrogen-bonding and hydrophobic properties of a selection of different ATP binding sites was performed, namely on CDK2, CDK4, and ERK2; any differences between them was investigated further.

In the first part of our analysis, any differences between the hydrogen-bonding sitepoints in the ATP binding site of the three kinases were investigated (Tables 4 and 5). Table 4 presents the set of nine sitepoints that are used by a range of CDK2 inhibitors (see Table 6 for the complete list of inhibitors studied), along with the respective sitepoints in the binding sites of CDK4 and ERK2. All except one of these sitepoints are

Table 4. The set of nine sitepoints A–I used by a range of inhibitors of CDK2 along with the corresponding sitepoints in the binding sites of CDK4 and ERK2.

Code	CDK2	CDK4	ERK2
A	O Glu 81	O Glu 81	O Asp 104
B	N Leu 83	N Val 83	N Met 109
C	O Leu 83	O Val 83	O Met 106
D	OD2 Asp 86	OD2 Asp 86	OD2 Asp 109
E	NZ Lys 33	NZ Lys 33	NZ Lys 52
F	O Gln 131	O Gln 131	O Ser 151
G	OD1 Asn 132	ND2 Asn 132	ND2 Asn 152
H	OD2 Asp 145	OD2 Asp 145	OD2 Asp 165
I	N Asp 145	N Asp 145	N Asp 165

conserved across the three binding sites with structurally equivalent groups. The only sitepoint that is not conserved is the side chain amide O atom of Asn 132, which is replaced by the side chain amide N atom of the equivalent asparagine residue in the pockets of CDK4 and ERK2 (the rotameric conformation of the terminal amide group of Asn 132 in these enzymes is consistent with intraprotein hydrogen bonds). However, of the inhibitors studied, this particular sitepoint is only used by hymenialdisine; as this inhibitor demonstrates no significant selectivity between the three binding sites investigated, this sitepoint is unlikely to play a role in determining ligand selectivity.

The SWAPS values calculated for the set of nine sitepoints in the binding site of CDK2, along with those for their equivalent sitepoints in the binding sites of CDK4 and ERK2, are shown in Table 5. The first clear difference between the SWAPS values

Table 5. The d_{rel}, SAPS, and SWAPS values calculated for each of the sets of nine hydrogen-bonding sitepoints used by inhibitors of CDK2, along with the scores for the corresponding sitepoints in CDK4 and ERK2.[a]

	CDK2			CDK4			ERK2		
Code	SAPS	d_{rel}	SWAPS	SAPS	d_{rel}	SWAPS	SAPS	d_{rel}	SWAPS
A	0.1655	1.4	0.2317	0.0769	1.4	0.1074	0.1135	1.4	0.1589
B	0.9998	0.7	0.6999	0.9219	0.7	0.6453	0.4649	0.7	0.3254
C	0.0534	1.4	0.0748	0.5890	1.4	0.8246	0.3516	1.4	0.4922
D	0.9153	4.9	4.4850	0.4889	4.9	2.3956	0.8018	4.9	3.9288
E	0.5035	4.0	2.0140	0.5000	4.0	2.0000	0.7094	4.0	2.8379
F	0.6483	1.4	0.9076	0.5946	1.4	0.8324	0.5294	1.4	0.7412
G	0.1560	2.0	0.3120	0.0580	1.7	0.0986	0.0242	1.7	0.0411
H	0.6326	4.9	3.0997	0.1275	4.9	0.6248	0.9299	4.9	4.5565
I	0.9978	0.7	0.6985	0.9925	0.7	0.6947	0.7136	0.7	0.4995

[a] The sitepoints are represented by the letters A–I, as defined in Table 4.

for a particular sitepoint and its equivalent sitepoints in the other binding sites is the low score calculated for the backbone O atom of Leu 83 (sitepoint C) in CDK2. This sitepoint is part of the core triplet of sitepoints in the ATP binding site and is used by the majority of inhibitors of CDK2, in addition to those demonstrating selectivity for CDK4 over CDK2 (that is, compound II). As such, it is unlikely that the targeting of this sitepoint would lead to the generation of a ligand selective be-

tween these three binding sites. Another difference in the SWAPS values are the higher scores calculated for the side chain amide O atom of Asn 132 and the carboxylate O atom of Asp 145 (sitepoints G and H, respectively) in CDK2, as compared with the equivalent sitepoints in CDK4. Of the inhibitors studied, these two sitepoints are only used by the inhibitor hymenialdisine, which forms hydrogen bonds with both of them. Whereas the near tenfold increase in selectivity of hymenialdisine for CDK2 over CDK4 (Table 6) correlates with this difference in SWAPS values, this inhibitor displays even greater selectivity against ERK2, for which sitepoint H (a carboxylate O atom of Asp 165) has an even higher SWAPS value than in CDK2. As such, it appears unlikely that the targeting of these sitepoints alone would lead to the generation of a subtype-selective inhibitor.

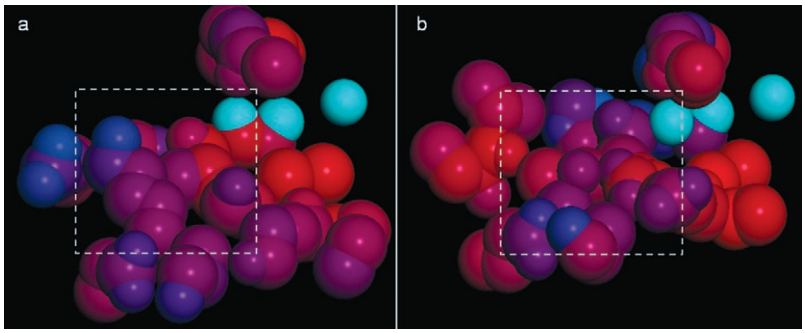


Figure 4. Comparison of the hydrophobic profiles of the specificity surface (dashed box) in a) CDK2 and b) CDK4. Protein atoms are coloured according to their predicted hydrophobic score, with the most hydrophobic atoms in red and the least, in blue. The three core hydrogen-bonding sitepoints of the adenine binding region (O Leu 83, N Leu 83, and O Glu 81 in CDK2) are coloured cyan to indicate alignment. The atomic coordinates for CDK2 were taken from PDB entry 1aq1 and those for CDK4 were taken from PDB entry 1gij.

Table 6. IC_{50} values of a range of inhibitors of CDK2, along with the corresponding values for CDK4 and ERK2.[a]

Inhibitor	IC_{50} [μM] CDK2	IC_{50} [μM] CDK4	IC_{50} [μM] ERK2
olomoucine	7.0	>1000	50.0
staurosporine	0.007	>10	0.020
(*R*)-roscovitine	0.7	>100	14.0
purvalanol B	0.006	>10	3.33[b]
hymenialdisine	0.07	0.6	2.0
indirubin-5-sulfonate	0.04	0.3	>100
compound I[c]	0.096	0.051	ND[d]
compound II[c]	25	0.21	ND[d]

[a] Values were taken from Knockaert et al.[3] [b] Value for the ERK1 subtype. [c] Values taken from Ikuta et al.[12] [d] Not determined.

Given the overall conserved nature of the hydrogen-bonding sitepoints in the ATP binding sites of CDK2, CDK4, and ERK2, it seems likely that inhibitor selectivity may be predominantly the result of differences in either the hydrophobic or steric properties of the binding sites, as previous molecular field analyses of protein kinases have revealed.[24] To explore this further, the hydrophobic properties of each of the ATP binding sites of these enzymes were analysed with our recently developed method and compared, to find a rationale for selectivity. Our analysis identified three main regions that display a significant variation in their hydrophobic profiles across these binding sites.

The first region identified is a hydrophobic surface that extends away from the adenine binding site and is not used by ATP (Figure 4). This region has previously been described as the specificity surface based on its restricted conservation across multiple protein kinase binding sites;[26] for instance, in ERK2, the ATP binding site does not extend into this region. Based on this, the targeting of the specificity surface with nonpolar aromatic groups has led to large gains in both binding affinity and specificity.[19] Our hydrophobic analysis of this specificity surface in CDK2 reveals a highly hydrophobic region near the core sitepoints, which becomes less hydrophobic as the surface extends away (Figure 4a). In contrast, the corresponding surface in CDK4 is initially slightly less hydrophobic near the core sitepoints, but becomes progressively more hydrophobic as the surface extends away (Figure 4b). In addition to this difference in hydrophobicity, the specificity surface in CDK2 contains a bulky lysine residue (Lys 89) in place of a smaller threonine residue (Thr 89) in CDK4. This residue substitution may impose steric restrictions on any inhibitor groups targeting this area.

A second region (termed Region 2) that exhibits differences between the hydrophobic properties of the two sites is highlighted in a view of the specificity surface from a slightly different perspective (Figure 5). The region indicated by the dashed box in Figure 5 reveals a surface with greater hydrophobicity in the CDK4 binding site than that of the CDK2 binding site, despite being composed of the same residue types (Asp 86, Lys 129, Gln 131, Asn 132, and Leu 132 in both CDK2 and CDK4). This suggests that the observed difference in hydrophobicity arises from the surface characteristics of this region, a property explicitly considered by our algorithm. This increased hydrophobicity in CDK4 can be explained by a combination of this region lying on a slightly more concave surface in CDK4 along with the closer proximity of the largely nonpolar residues Ile 10 and Gly 11 from the opposing surface. This latter difference would result in an increased nonpolar crowding effect on this region and a consequent increase in hydrophobicity, an effect that our method was designed to capture explicitly. Local changes in the conformation of this region in the active form of CDK2 and/or CDK4 may have a more noticeable effect given that both proteins share the same amino acid composition in this part of the binding site.

There is a third region (termed Region 3) of the binding site that demonstrates a substantial difference in hydrophobic properties between the proteins. This region is local to Phe 80 in CDK2 (Figure 6). The analysis of this region in CDK2 (Fig-

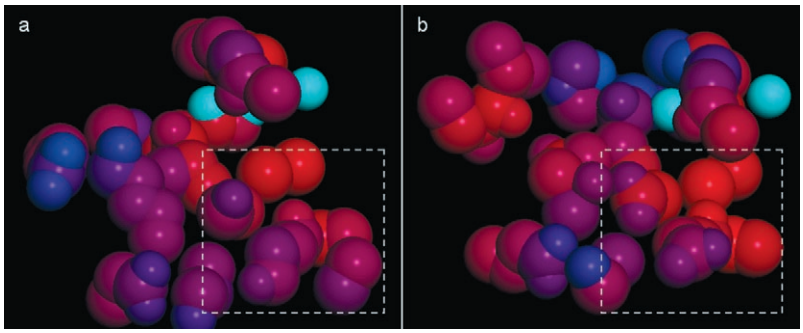


Figure 5. Hydrophobic analysis of the region local to the specificity surface (dashed box) in a) CDK2 and b) CDK4, shown from a different perspective to highlight additional differences between the two regions. Protein atoms are coloured according to their predicted hydrophobic scores, with the most hydrophobic atoms in red and the least hydrophobic in blue. The three core hydrogen-bonding sitepoints of the adenine binding region (O Leu 83, N Leu 83, and O Glu 81 in CDK2) are coloured cyan to indicate alignment. The atomic coordinates for CDK2 were taken from PDB entry 1aq1 and those for CDK4 were taken from PDB entry 1gij.

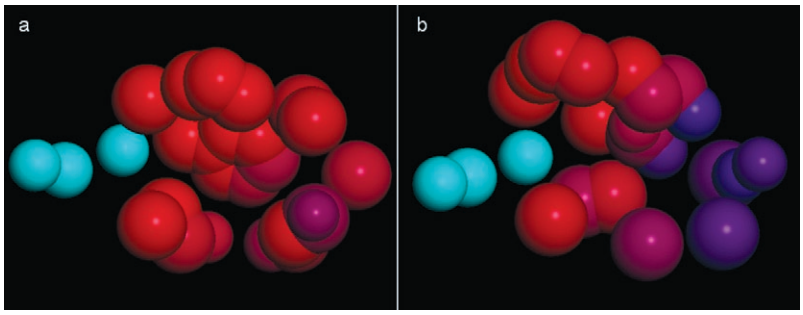


Figure 6. Comparison of the hydrophobic analysis of the region local to Phe 80 in a) CDK2 and b) the corresponding region in ERK2. Protein atoms are coloured according to their predicted hydrophobic scores, with the most hydrophobic atoms coloured red and the least hydrophobic in blue. The three core hydrogen-bonding sitepoints of the adenine binding region (O Leu 83, N Leu 83, and O Glu 81 in CDK2) are coloured cyan to indicate alignment. The atomic coordinates for CDK2 were taken from PDB entry 1aq1 and those for ERK2 were taken from PDB entry 4erk.

ure 6 a) and CDK4 predict a highly hydrophobic surface relative to that for the corresponding region in ERK2 (Figure 6 b). This difference most likely arises from variations in surface shape and polarity due to the presence of a glutamine residue (Gln 103) in ERK2 as opposed to the more bulky and nonpolar phenylalanine residue (Phe 103) present in CDK2 and CDK4.

Rationalisation of inhibitor selectivity

Comparative analyses of the crystal structures of a number of key inhibitors of CDKs have helped in the search for a rationalisation of their specificity and selectivity and to guide the design of more potent inhibitors.[26] Our methods for analysing the hydrogen-bonding and hydrophobic properties of the binding sites of proteins are aimed in this case at providing a different light under which to look at the key structural differences across a few selected kinase proteins, looking for consistencies with previous analyses, and for a more detailed analysis of the interaction properties of their binding sites.

Following the separate analysis of the above three discriminatory regions, the relationship between inhibitor selectivity and their targeting was investigated. By using a set of inhibitors for the cyclin-dependent kinases (listed in Table 6), which demonstrate a range of selectivities across the three protein kinases studied and for which crystallographic binding modes are known, any relationships between their binding characteristics and selectivity could be elucidated.

The first group of inhibitors all demonstrate selectivity for CDK2 over CDK4, whilst also binding to ERK2 (although with less affinity than for CDK2). This group of inhibitors consist of olomoucine, staurosporine, (*R*)-roscovitine, and purvalanol B (Figure 7). In contrast, the second group of inhibitors all bind CDK4, whilst demonstrating a range of selectivities against CDK2 and ERK2. This group of inhibitors consist of hymenialdisine, indirubin-5-sulfonate, compound I and a CDK4-selective inhibitor (compound II, Figure 8).

As Figures 7 and 8 illustrate, both sets of ligands target the hydrophobic selectivity surface with nonpolar groups. Whereas the targeting of this region with nonpolar ligand groups has been shown to produce ligands with improved binding (for example, the addition of a phenyl group to the CDK2 O^6-cyclohexylmethylguanine inhibitor NU2058 increases its binding affinity ~10-fold[27]), it does not improve selectivity for CDK2 over CDK4 (Table 7). This inability to improve selectivity would be expected given the relative conservation of this largely hydrophobic region between the CDK2 and CDK4 binding sites (although there are some steric differences, as noted above). These observations are consistent with a five-point pharmacophore model for kinase frequent hitters (promiscuous inhibitors),[28] which include two acceptor and two donor points that match the three core sitepoints and an additional hydrogen-bonding group (probably Lys 33), and an aromatic point that matches this hydrophobic specificity region.

One difference between these two sets of ligands is found in the ribose binding region (Region 2). The first group of inhibitors, which do not bind CDK4, occupy the region local to Gln 131 with polar groups (that is, the hydroxy groups in olomoucine, roscovitine, and purvalanol B and an amine in staurosporine; Figure 7). In contrast, the second group of inhibitors, which do bind CDK4, do not occupy the region local to Gln 131 to the same extent and instead present less polar groups to this region (that is, the tricyclic region of compound

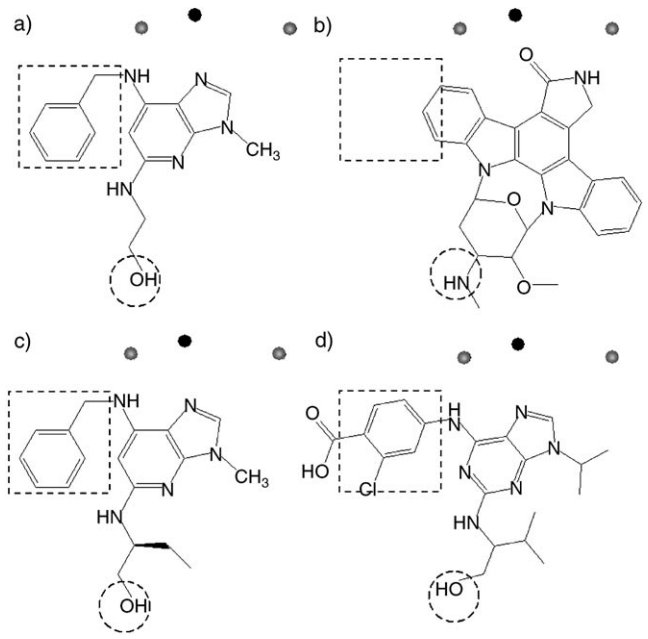


Figure 7. Chemical structures of the CDK2 inhibitors a) olomoucine, b) staurosporine, c) (*R*)-roscovitine, and d) purvalanol B. The three core sitepoints (O Glu 81, N Leu 83, and O Leu 83 in CDK2) are highlighted (grey spheres) to indicate the relative orientation of the inhibitors. In addition, the inhibitor structures that bind in regions 1 (specificity surface, dashed box) and 2 (dashed circle), as identified in the hydrophobic analysis, are shown.

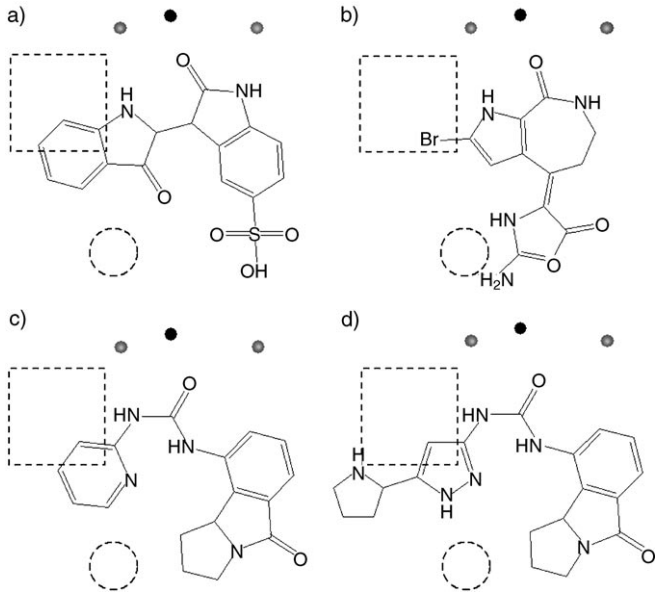
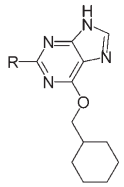
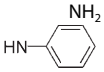
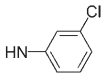
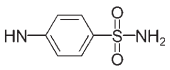


Figure 8. Chemical structures of the CDK2/4 inhibitors a) indirubin-5-sulfonate, b) hymenialdisine, and c) compound I, along with the CDK4-selective inhibitor d) compound II. The three core sitepoints (O Glu 81, N Leu 83 and O Leu 83 in CDK2), are highlighted (grey spheres) to indicate the relative orientation of the inhibitors. In addition, the inhibitor structures that bind in regions 1 (specificity surface, dashed box) and 2 (dashed circle), as identified in the hydrophobic analysis, are shown.

II; Figure 8). This difference between the binding modes of the two sets of ligands is in agreement with the difference in the hydrophobic properties calculated for this region of the ATP binding site (Figure 5), which in CDK2 is predicted as being far less hydrophobic (more polar) than in CDK4.

Table 7. Effects of substituting different groups on the CDK2 O^6-cyclohexylmethylguanine inhibitor NU2058 to target the hydrophobic specificity region in the ATP binding pocket.[a]



Inhibitor	R	Inhibition of CDK activity [μM] CDK2	CDK4
NU2058	NH_2	12 ± 3[b]	>100[b]
NU6094		1.0 ± 0.03	16 ± 5
NU6086		2.3 ± 0.3	>100
NU6102		0.006 ± 0.0005	1.6 ± 1.0

[a] Data reported as IC_{50} values, taken from Davies et al.[27]. [b] Inhibition reported as K_i; the authors demonstrated that the K_i and IC_{50} values described in their study are proportional.

The CDK4-selective inhibitor (compound II) in the second set of ligands is an example of a ligand that does not bind to CDK2 (Figure 8c). This lack of affinity is most likely due to steric restrictions in the region to the outside of the specificity region, imposed by the presence of a lysine residue (Lys 89) in the CDK2 binding site as opposed to the smaller threonine residue (Thr 89) in the CDK4 binding site. This steric restriction prevents the bulky nonpolar indol-1-yl extension on this inhibitor from binding in this region. The importance of this region in the binding site to selectivity against CDK2 was demonstrated by the 1000-fold increase in selectivity conferred by the addition of such a bulky nonpolar extension;[12] the removal of this group from the CDK4-selective inhibitor results in a ligand that binds to both CDK2 and CDK4. Our hydrophobic analysis of this region in the binding site (Figure 4) predicts it as being more hydrophobic in CDK4 than in CDK2, which suggests that it is important for the bulky extension of the ligand to be nonpolar, as it is in compound II. On the other hand, a smaller and more polar extension would be expected to favour CDK2 binding. This prediction is confirmed in the binding of the CDK2-inhibitor NU6102 (Table 7), which has a polar sulfonamide group at this position and demonstrates ~1000-fold selectivity for CDK2 over CDK4.

In the case of inhibitor selectivity for CDK2 versus ERK2, the region of the binding site showing the most variation was that around Phe 80 in CDK2 (Region 3). As predicted by our hydrophobic analysis of this region (Figure 6), targeting it with a nonpolar group would be expected to produce a ligand that binds preferentially to CDK2 over ERK2. This behaviour was indeed demonstrated by the binding profiles of the inhibitors indirubin-5-sulfonate and hymenialdisine, which show selectivi-

ty for CDK2 over ERK2; both compounds project nonpolar rings into this region. In contrast, the ERK2-binding p38 inhibitor SB220025, which has been co-crystallised with ERK2 and shown to bind to the ATP binding site (PDB code: 3erk),[29] targets this region with the more polar fluorine atom (Figure 9).

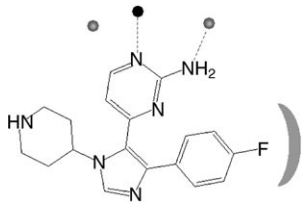


Figure 9. Structure of the p38 inhibitor SB22025 as bound to ERK2 (PDB code: 3erk). Hydrogen bonds formed between the inhibitor and two of the three core sitepoints (O Asp 104, N Met 106, and O Met 106) are highlighted (grey spheres) to indicate orientation. Region 3 (the area local to Phe 80 in CDK2) is highlighted by the grey crescent.

Similar conclusions about the importance of Phe 80 were drawn after a molecular field analysis of the differences between the binding sites of CDK2 and GSK3β.[30]

Conclusion

The ATP binding sites of the cyclin-dependent kinases CDK2 and CDK4 and of the extracellular signal-regulated kinase ERK2 have been analysed with recently developed hydrogen-bonding prioritisation and hydrophobic analysis methods. The hydrogen-bonding prioritisation method, designed to rank hydrogen-bonding sitepoints within a binding site, was able to prioritise those sitepoints observed to bind the native ATP ligand, in addition to suggesting further high-scoring sitepoints found to bind inhibitors of these enzymes. However, it is clear that the hydrogen-bonding sitepoints in the ATP binding sites of CDK2, CDK4, and ERK2 are fairly well conserved, making it likely that inhibitor selectivity may be predominantly the result of differences in either the hydrophobic or steric properties of the binding sites.

The hydrophobic analysis predicted properties of a binding site that, as would be expected, is complementary to the native ATP ligand. Moreover, this hydrophobic analysis indicated regions in the binding pocket unoccupied by ATP which may provide additional targets for inhibitor design. Three key regions were identified by the hydrophobic analysis. The first region, which has previously been described as the specificity surface, is a largely hydrophobic surface present in both CDK2 and CDK4 binding sites. The targeting of this region with a nonpolar ligand group has been found to increase binding affinity. Owing to its presence in both CDK2 and CDK4 binding sites, the targeting of this region is unlikely to result in selectivity between these two proteins. However, the binding-site region to the outside of this specificity surface is more hydrophobic in the CDK4 binding site and is also able to accommodate a more bulky ligand group. This difference is reflected in the binding mode of the CDK4-selective inhibitor, compound II. The second region, local to this specificity surface, consisted of a surface in CDK4 more hydrophobic than in CDK2, which was reflected in the nature of the inhibitor groups occupying this region. The third region, local to Phe 80 in CDK2, presented a surface more hydrophobic in CDK2 than in ERK2. This difference was again reflected by the binding profiles of kinase-selective inhibitors.

Whereas the selectivity-conferring regions identified in this analysis have already been targeted by existing inhibitors, our hydrophobic analysis method enabled their rapid detection in addition to providing an indication of the type of ligand group that is likely to bind favourably. Importantly, our method was also able to identify variations in hydrophobicity that derived from surface-based effects alone, such as the greater hydrophobicity of certain regions in CDK4 in comparison with CDK2, despite being made up of largely similar residue types. This, in turn, allows the rationalisation of increased selectivity of kinase inhibitors on the basis of the hydrophobic properties of their binding sites.

Acknowledgements

M.D.K. gratefully acknowledges De Novo Pharmaceuticals for the award of a postgraduate scholarship.

Keywords: CDKs · drug design · hydrogen bonds · hydrophobic effect · inhibitors

[1] C. Norbury, P. Nurse, *Annu. Rev. Biochem.* **1992**, *61*, 441 – 470.
[2] P. Kaldis, *Cell. Mol. Life Sci.* **1999**, *55*, 284 – 296.
[3] M. Knockaert, P. Greengard, L. Meijer, *Trends Pharmacol. Sci.* **2002**, *23*, 417 – 425.
[4] D. M. Gitig, A. Koff, *Mol. Biotechnol.* **2001**, *19*, 179 – 188.
[5] A. Kamb, *Cold Spring Harbor Symp. Quant. Biol.* **1994**, 59, 39 – 47.
[6] T. M. Sielecki, J. F. Boylan, P. A. Benfield, G. L. Trainor, *J. Med. Chem.* **2000**, *43*, 1 – 18.
[7] S. Omura, Y. Iwai, A. Hirano, A. Nakagawa, J. Awaya, H. Tsuchya, Y. Takahashi, R. Masuma, *J. Antibiot.* **1977**, *30*, 275 – 282.
[8] H. H. Sedlacek, J. Czech, R. Naik, G. Kaur, P. Worland, M. Losiewicz, B. Parker, B. Carlson, A. Smith, A. Senderowicz, E. Sausville, *Int. J. Oncol.* **1996**, *9*, 1143 – 1168.
[9] J. Baselga, S. D. Averbuch, *Drugs* **2000**, *60*, 41 – 42.
[10] H. L. De Bondt, J. Rosenblatt, J. Jancarik, H. D. Jones, D. O. Morgan, S. H. Kim, *Nature* **1993**, *363*, 595 – 602.
[11] U. Schulze-Gahmen, H. L. De Bondt, S. H. Kim, *J. Med. Chem.* **1996**, *39*, 4540 – 4546.
[12] M. Ikuta, K. Kamata, K. Fukasawa, T. Honma, T. Machida, H. Hirai, I. Suzuki-Takahashi, T. Hayama, S. Nishimura, *J. Biol. Chem.* **2001**, *276*, 27548 – 27554.
[13] M. D. Kelly, R. L. Mancera, *J. Comput.-Aided Mol. Des.* **2003**, *17*, 401 – 414.
[14] M. D. Kelly, R. L. Mancera, *J. Med. Chem.* **2005**, *48*, 1069 – 1078.
[15] I. J. Bruno, J. C. Cole, J. P. M. Lommerse, R. S. Rowland, R. Taylor, M. L. Verdonk, *J. Comput.-Aided Mol. Des.* **1997**, *11*, 525 – 537.
[16] A. Nicholls, K. A. Sharp, B. Honig, *Proteins Struct. Funct. Genet.* **1991**, *11*, 281 – 296.
[17] K. Lum, D. Chandler, J. D. Weeks, *J. Phys. Chem. B* **1999**, *103*, 4570 – 4577.
[18] U. Schulze-Gahmen, J. Brandsen, H. D. Jones, D. O. Morgan, L. Meijer, J. Vesely, S. H. Kim, *Proteins Struct. Funct. Genet.* **1995**, *22*, 378 – 391.
[19] W. F. De Azevedo, S. Leclerc, L. Meijer, L. Havlicek, M. Strnad, S. H. Kim, *Eur. J. Biochem.* **1997**, *243*, 518 – 526.

[20] N. S. Gray, L. Wodicka, A. M. Thunnissen, T. C. Norman, S. Kwon, F. H. Espinoza, D. O. Morgan, G. Barnes, S. Leclerc, L. Meijer, S. H. Kim, D. J. Lockhart, P. G. Schultz, *Science* **1998**, *281*, 533 – 538.
[21] A. M. Lawrie, M. E. Noble, P. Tunnah, N. R. Brown, L. N. Johnson, J. A. Endicott, *Nat. Struct. Biol.* **1997**, *4*, 796 – 801.
[22] R. Hoessel, S. Leclerc, J. A. Endicott, M. E. Nobel, A. Lawrie, P. Tunnah, M. Leost, E. Damiens, D. Marie, D. Marko, E. Niederberger, W. Tang, G. Eisenbrand, L. Meijer, *Nat. Cell Biol.* **1999**, *1*, 60 – 67.
[23] L. Meijer, A. M. Thunnissen, A. W. White, M. Garnier, M. Nikolic, L. H. Tsai, J. Walter, K. E. Cleverley, P. C. Salinas, Y. Z. Wu, J. Biernat, E. M. Mandelkow, S. H. Kim, G . R. Pettit, *Chem. Biol.* **2000**, *7*, 51 – 63.
[24] T. Naumann, H. Matter, *J. Med. Chem.* **2002**, *45*, 2366 – 2378.
[25] T. G. Davies, J. Bentley, C. E. Arris, F. T. Boyle, N. J. Curtin, J. A. Endicott, A. E. Gibson, B. T. Golding, R. J. Griffin, I. R. Hardcastle, P. Jewsbury, L. N. Johnson, V. Mesguiche, D. R. Newell, M. E. Noble, J. A. Tucker, L. Wang, H. J. Whitfield, *Nat. Struct. Biol.* **2002**, *9*, 745 – 749.
[26] I. R. Hardcastle, B. T. Golding, R. J. Griffin, *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 325 – 348.
[27] T. G. Davies, D. J. Pratt, J. A. Endicott, L. N. Johnson, M. E. M. Noble, *Pharmacol. Ther.* **2002**, *93*, 125- 133.
[28] A. M. Aronov, M. A. Murcko, *J. Med. Chem.* **2004**, *47*, 5616 – 5619.
[29] Z. Wang, B. J. Canagarajah, J. C. Boehm, S. Kassisa, M. H. Cobb, P. R. Young, S. Abdel-Meguid, J. L. Adams, E. J. Goldsmith, *Structure* **1998**, *6*, 1117 – 1128.
[30] A. Vulpetti, P. Crivori, A. Cameron, J. Bertrand, M. G. Brasca, R. D'Alessio, P. Pevarello, *J. Chem. Inf. Model.* **2005**, *45*, 1282 – 1290.

Received: August 31, 2005
Revised: November 17, 2005
Published online on January 23, 2006

