J Mol Model
DOI 10.1007/s00894-005-0063-1

ORIGINAL PAPER

Alfonso T. García-Sosa · Ricardo L. Mancera

The effect of a tightly bound water molecule on scaffold diversity in the computer-aided de novo ligand design of CDK2 inhibitors

Received: 8 March 2005 / Accepted: 21 July 2005

Abstract We have determined the effects that tightly bound water molecules have on the *de novo* design of cyclin-dependent kinase-2 (CDK2) ligands. In particular, we have analyzed the impact of a specific structural water molecule on the chemical diversity and binding mode of ligands generated through a de novo structure-based ligand generation method in the binding site of CDK2. The tightly bound water molecule modifies the size and shape of the binding site and we have found that it also imposed constraints on the observed binding modes of the generated ligands. This in turn had the indirect effect of reducing the chemical diversity of the underlying molecular scaffolds that were able to bind to the enzyme satisfactorily.

Keywords Hydration · Solvation · Structure-based drug design · CDK2

Introduction

The crystal structures of protein binding sites often reveal the presence of several water molecules. Some of these water molecules may be artefacts of the X-ray determination [1], while others are loosely bound to the surface of the protein. However, a few water molecules are tightly bound to the surface, as revealed by their crystallographic order and the number of their interactions with the protein [2, 3]. Most drug design and ligand docking applications start by removing all water molecules from the binding site of a target protein. This is unlikely to be realistic, particularly when tightly bound water molecules are present, as such solvent molecules provide hydrogen-bonding groups that can mediate the interactions between the ligand and the protein. The resulting formation of a water–ligand–protein hydrogen-bonding network can help stabilize the ligand–protein interaction [4–6] and may have a significant effect on the binding mode and even the chemical diversity of ligands binding to a given protein binding site.

There is an increasing number of examples in the drug-design literature where tightly bound water molecules in the binding site of proteins have been mimicked or included [7–9]. These applications reveal that displacing a tightly bound water molecule by a ligand may improve the binding affinity, although this is not always the case [10]. Other studies have shown that both natural substrates [11] and designed inhibitors [12] can make use of existing tightly bound water molecules to "bridge" their interactions with the protein. Recent literature has also been providing examples of an increasing number of molecular modeling applications that make use of water molecules. It has been reported that ligand–protein docking [13] and virtual screening of organic compounds [14, 15] can be improved by the presence of bound water molecules in the binding site of proteins. Water molecules have also been used to distinguish the binding of different chemical scaffolds to a protein [15], to improve the predictive ability of three-dimensional QSAR models [16] and to aid in the structural interpretation of ligand-derived pharmacophore models of the binding sites of proteins [17].

A study on the use of tightly bound water molecules in the de novo ligand design of molecular scaffolds for bacterial neuraminidase provided the first evidence of the influence that such water molecules can have in drug

A. T. García-Sosa (✉)
Department of Pharmacology,
University of Cambridge,
Tennis Court Road,
Cambridge CB2 1PD, UK
e-mail: atgs@cantab.net
Tel.: +1-507-5383336
Fax: +1-507-2849111

R. L. Mancera
Western Australian Biomedical Research Institute,
School of Biomedical Sciences and School of Pharmacy,
Curtin University of Technology,
G.P.O. Box U1987, Perth WA 6865, Australia
e-mail: r.mancera@wabri.org.au

A. T. García-Sosa
Computer-Aided Molecular Design Laboratory,
Guggenheim 711, Mayo Clinic College of Medicine
200 First Street SW,
Rochester, MN 55905, USA

design [18]. It was observed that the complete removal of all water molecules led to difficulties when generating any potential ligands. This was due to the fact that removing all tightly bound water molecules left their now unsatisfied hydrogen-bonding groups beyond physical reach for a ligand to satisfy. The more water molecules that were identified as tightly bound were allowed in the binding site, the easier it became to generate ligands, which were also observed to be more chemically diverse. It was proposed that, in some cases, tightly bound water molecules may in fact be more accessible for hydrogen bonding to an incoming ligand than the actual protein hydrogen-bonding groups associated with them. Water molecules may thus behave as versatile hydrogen-bonding groups and reduce the conformational constraints of a particular binding site.

A recent validation study on the use of computer-aided de novo drug design showed that the Skelgen algorithm [19, 20] was able to generate representative molecular scaffolds of most inhibitor classes for a number of proteins of pharmaceutical interest [20]. In this work we have analyzed the crystal structures of these proteins and found that cyclin-dependent kinase 2 (CDK2) contained a particularly relevant tightly bound water molecule. We then proceeded to investigate in detail the effect of the presence of this water molecule during the in silico generation of representative molecular scaffolds. We report our analysis of the variation in chemical diversity and binding mode of these molecular scaffolds.

CDK2 binding site analysis

CDK2 is an enzyme implicated in cell division whose deregulated activity is thought to contribute to the initiation and progression of several diseases such as cancer, neurodegenerative and inflammatory disorders [21]. Cyclin-dependent kinases catalyze the transfer of a phosphate group from ATP to a specific substrate amino-acid residue (serine or threonine), and the majority of drug discovery research in this area has been aimed at trying to produce small molecules that mimic ATP and bind competitively to its binding site [22–26]. Despite concerns about the selectivity of inhibitors among kinases based on designs using the ATP site, the discovery and ultimate development of potent and selective inhibitors, such as the anti cancer drugs Gleevec and Iressa, have helped validate kinase inhibition as a therapeutic strategy [21, 27], with many known ligands reported in the literature [28–30]. The binding mode of ATP has thus served as the basis for the search for new inhibitors of CDK2.

Materials and methods

A survey of the Protein Data Bank (pdb) [31] was used to obtain a selection of 20 X-ray crystal structures of CDK2 (no mutations, a resolution below 2.5 Å and the same amino acid sequence). One of these crystal structures is that of the *apo* enzyme (pdb code 1 hcl), whilst the others contain either ATP (pdb codes 1 hck and 1 fin) or an inhibitor bound to the ATP site (pdb codes 1 aqi, 1 ckp, 1 di8, 1 dm2, 1 eiv, 1 eix, 1 fvt, 1 h0u, 1 h0v, 1 h0w, 1 jsv, 1 jvp, 1 ke5, 1 ke6, 1 ke7, 1 ke8 and 1 ke9). For the present study, and in accordance with the validation study of the computer aided de novo drug design algorithm that we employ [20], the crystal structure with pdb code 1 di8 was used. This structure was determined to a resolution of 2.2 Å [32]. It presents an intermediate orientation of the hydrogen bonding groups in the hinge strand connecting the N and C terminal domains in the ATP binding site [20]. The ATP molecule has intermolecular interactions via its adenine ring to Glu 81 (Glu 75 in structure 1 di8) a nd Leu 83 (Leu 77 in structure 1 di8), as well as via its triphosphate group. The residues in structure 1 di8 were renumbered according to the numbering observed in all the other structures.

Figure 1a shows the superposition (using residues Glu 81 and Leu 83 as reference) of the above-listed crystal structures, including their water molecules. All the ligands share a common flat orientation in the binding site. The two water molecules seen in Fig. 1a interacting with Glu 81 and Leu 83 are found in the *apo* enzyme (1 hcl). Their presence confirms the importance of these hydrogen bonding groups in the binding site. An analysis of the ligands in Fig. 1a reveal that they are surrounded by water molecules that make various interactions with both the ligand and the protein. The phosphate group in particular exhibits this feature, as this highly charged group occupies regions of the binding site where extensive clusters and networks of water molecules can be seen.

Figure 1b shows the (ATP) binding site of CDK2 with the inhibitor 4 [3 hydroxyanilino] 6,7 dimethoxyquinazoline (as found in 1di8), with all hydrogen bonding groups (which we refer to hereafter as site points) that are available for ligand generation, as well as all identified tightly bound water molecules (see below). Most known inhibitors of CDK2 interact with the backbone groups Glu 81 C=O, Leu 83 NH and Leu 83 C=O. Fig. 1b shows these groups as hydrogen-bonding groups, as well as other groups in the enzyme binding site. The fused ring of the ligand interacts with the β-strand (residues 81–84, hinge region) that links the two domains of the protein. It is interesting to note that this ligand seems to form two –CH…O– hydrogen bonds with the protein. This kind of non-standard hydrogen bond has been previously observed in heterocyclic kinase ligands [33].

Computer-aided de novo ligand design was carried out using the program Skelgen [19, 20]. This is a program that can incrementally construct and/or modify a ligand in the binding site of a target protein using a Monte Carlo simulated annealing optimization algorithm. The program uses a set of common ring and acyclic fragments that are assembled together into a ligand structure following chemical rules. Ligand structures are modified through fragment additions, fragment removals and fragment mutations, as well as by molecular translations and rotations and conformational changes in torsional space. These modifications allow for previously incorporated fragments to be removed or replaced with different fragments, allowing the ligand gradually to satisfy the protein

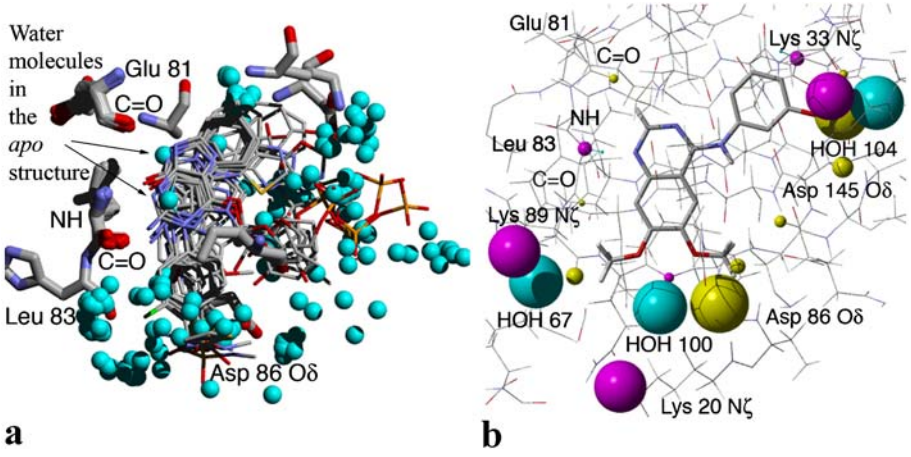


Fig. 1 **a** Superposition of the ligands found in the binding site of the crystal structures of CDK2. *Cyan spheres* represent crystallographically observed water molecules. *Green*=chlorine, *yellow*=sulphur, *red*=oxygen, *blue*=nitrogen, *grey*=carbon, *orange*=phosphorous. **b** The binding site of CDK2 (pdb code 1di8) with its co-crystallized inhibitor. *Yellow* spheres indicate hydrogen bond acceptors and magenta spheres represent hydrogen bond donors. *Cyan spheres* represent tightly bound water molecules. The size of the spheres is directly proportional to the degree of solvent accessibility of the hydrogen-bonding group

binding-site constraints. This process is carried out in a stochastic manner to optimize the interaction properties and chemical features of the generated ligand gradually during the annealing optimization. The assembled ligands must satisfy user-defined geometric constraints, such as those defining hydrogen-bond distances and angles for preselected donor and acceptor groups and the steric constraints imposed by the structure of the binding site. Full details of this algorithm can be found elsewhere [19, 20, 34]. The program was used to generate 200 molecular structures for each ligand design strategy (see below), producing a total of 600 scaffolds.

The ligand structures generated with Skelgen were minimized using the Discover 3 module in InsightII 2000 (Accelrys) with the CFF force field [35]. Additional torsional or out-of-plane restraints were used to ensure the planarity of aromatic or conjugated systems in some ligands. The protein was kept rigid in its original crystal structure conformation throughout the minimizations. However, hydrogen atoms in any amino acid in the binding site with at least one atom within 3.5 Å of the ligand were allowed to reorient in order to optimize the hydrogen-bonding network between the ligand, the water molecule (if present) and the protein. The ligands were allowed full flexibility during the minimizations. Water molecules were kept in their original crystal-structure positions but were allowed to reorient their hydrogen atoms. The energy minimizations were conducted in stages as described elsewhere [18] to try to retain the original binding mode. The minimizations were stopped when the energy gradient reached a value of less than 0.01 kcal mol^{-1} $Å^{-1}$.

Results and discussion

Identification of tightly-bound water molecules

We have recently introduced a multivariate logistic method called WaterScore to discriminate between tightly bound and displaceable water molecules in the binding sites of proteins [36]. Structural properties of water molecules in crystal structures such as the temperature B-factor, the solvent-accessible contact surface area, the number of protein atom contacts and the hydrogen-bond energy were analyzed using a multivariate logistic regression approach. A probabilistic model was obtained that can predict the likelihood of a water molecule being tightly bound to a binding site through the following equation:

$$P(Y=1) = \frac{\exp[A]}{(1+\exp[A])} \quad (1)$$

with

$$A = a - b_1 * Bf - b_2 * SCSA + b_3 * NPAC \quad (2)$$

where *Bf* is the B factor of a water molecule, *SCSA* is its solvent-accessible contact surface area, and *NPAC* is the number of protein atomic contacts. *P(Y=1)* is the probability of a water molecule being classified as tightly bound, and the values of the different coefficients are a=76.442, b_1=5.278, b_2=2.166 and b_3=84.458. We can see that this model reflects the fact that tightly bound water molecules will tend to have low B-factors, small solvent accessible contact surface areas and a large number of protein atomic contacts. Full details of this model can be found in the article published earlier [36].

By applying the above method to the crystal structure under study we found that three water molecules that are close to the ligand are predicted to be tightly bound: HOH 67, HOH 100 and HOH 104 (numbering as assigned in crystal structure 1 di8), as they had scores of 1.0. The position of these water molecules in the binding site of CDK2 can be seen in Fig. 1b. These water molecules are seen to participate in hydrogen bonding to important hydrogen-bonding groups in the binding site, as we discuss further below.

De novo ligand design

Due to the importance of the interactions of known inhibitors of CDK2 with site-points Glu 81 C=O, Leu 83 NH and Leu 83 C=O, a typical strategy for de novo ligand design involves generating molecular scaffolds that satisfy these three groups [20]. There are several additional site-points in the vicinity of the above groups that are available for hydrogen-bonding and that may or may not be used by a bound ligand: Asp 86 Oδ, Asp 86 N and Lys 89 Nζ. In the crystal structure of 1 di8, the inhibitor does not interact directly with these site-points but a water molecule (HOH 100) interacts with most of them, as shown in Fig. 1b.

Figure 1b also reveals that the tightly bound water molecules plays different roles in the binding of the inhibitor to CDK2. Water molecule HOH 67 interacts with Lys 89 Nζ, but it does not interact with the inhibitor in 1 di8 and is in fact too far away to have a significant direct role in the binding of a ligand. Water molecule HOH 104 engages in hydrogen bonding with the inhibitor, but it does not obstruct the site-points that it interacts with Lys 33 Nζ and Asp 145 Oδ) and is, consequently, unlikely to have a significant direct role in the binding of a ligand. Water molecule HOH 100 interacts directly with the inhibitor and with site-points Asp 86 Oδ and Asp 86 N, while being less than 4.5 Å away from Lys 20 Nζ. This water molecule partially blocks access to these site-points to an incoming ligand. An analysis of the other CDK2 crystal structures reveals that HOH 100 is also found as HOH 38Z in the structure pdb code 1h0w and HOH 582 in the structure pdb code 1 dm2. In most of the other crystal structures this water position is occupied by a polar group in the ligand (such as a sulfonamide group).

A molecular dynamics study of the hydration of the empty active site of CDK2 as well as complexed with ATP and two inhibitors has been reported recently [37]. A number of identified tightly bound water molecules are replaced by the purine ring of ATP and the inhibitors. In particular, a water molecule (which corresponds to HOH 100 in 1di8) was seen to interact strongly with Asp 86 and was identified as a key tightly bound water molecule mediating the interaction between the protein and the inhibitors [37], supporting our own finding that HOH 100 is a tightly bound water molecule.

On the basis of the above observations, we defined three strategies for ligand generation. The first "standard" strategy (named A) was to generate ligands that satisfy only two or three of the three typical site-points (Leu 83 NH, Leu 83 C=O and Glu 81 C=O). This is the same ligand design strategy adopted in an earlier validation study of Skelgen [20]. The second strategy (named B) was to generate ligands that also satisfy these same site-points *and* water molecule HOH 100 (which can act as a hydrogen-bond donor or acceptor). The third strategy (named C) was to generate ligands that satisfy the above three typical site-points *and* all the additional site-points that water molecule HOH 100 would otherwise block (Asp 86 Oδ, Asp 86 N and Lys 89 Nζ). This last strategy was adopted in order to generate ligands that would mimick the interactions of water molecule HOH 100 with the protein. This approach has been demonstrated by the higher activities of –OH substituted purine-like inhibitors [38] and the fact that several inhibitors interact with Asp 86 [39, 40]. Table 1 summarizes these three strategies that we adopted for *de novo* ligand generation. Each strategy allows the generation of ligands under different constraints as the shape and interaction properties of the binding site are modified in the presence or absence of the water molecule.

It should be borne in mind that the above ligand generation strategies did not aim to fill the entirety of the binding site, but rather attempted to find molecular scaffolds that would satisfy the specific hydrogen-bond interactions mentioned above. Satisfying these interactions alone does not lead to high-affinity inhibitors, because binding affinity is also achieved through lipophilic interactions between the planar, mostly heterocyclic ring systems carrying the donor and acceptor groups that bind to the hinge region (Leu 83 and Glu 81) and surrounding aliphatic side chains [41]. Furthermore, it is important to bear in mind that *de novo* ligand design methods may suggest synthetically unfeasible molecules. Therefore, we have focused our investigation to the analysis of molecular scaffolds that are known to be synthetically feasible.

Evaluation of de novo generated ligands

Any minimized ligand that did not make hydrogen bonds with at least two of the three typical site-points (Leu 83 NH, Leu 83 C=O and Glu 81 C=O) was discarded, as this has been observed to be an important requirement for biological activity [20]. A further condition was that ligands were not allowed to use a hydroxyl group (–OH) to satisfy Leu 83 NH and Glu 81 C=O simultaneously, since compounds of this type are known but have not led to any CDK2 inhibitors of pre-clinical interest [20]. When water molecule HOH 100 was present, ligands were further required to form a hydrogen bond to it.

Table 1 Summary of strategies for de novo ligand generation

Strategy	Site-points used
A= standard	Glu 81 C=O
	Leu 83 NH
	Leu 83 C=O
B= including water	Glu 81 C=O
	Leu 83 NH
	Leu 83 C=O
	HOH 100
C=additional site-points	Glu 81 C=O
	Leu 83 NH
	Leu 83 C=O
	Asp 86 Oδ
	Asp 86 N
	Lys 89 Nζ

Once all ligands had been minimized and filtered, the molecular scaffolds involved in the interactions with the protein site-points (and the water molecule HOH 100, if present) were extracted, and any duplicates were removed. The scaffolds were then classified manually according to their hydrogen-bonding patterns and binding modes.

Scaffold analysis

Table 2 shows the chemical structures of these different scaffolds and illustrates schematically the possible two binding modes that were obtained with any (or all) of the three different ligand-design strategies. Within this

Table 2 Molecular scaffold classes and their binding modes (with *X*=O, N, S)

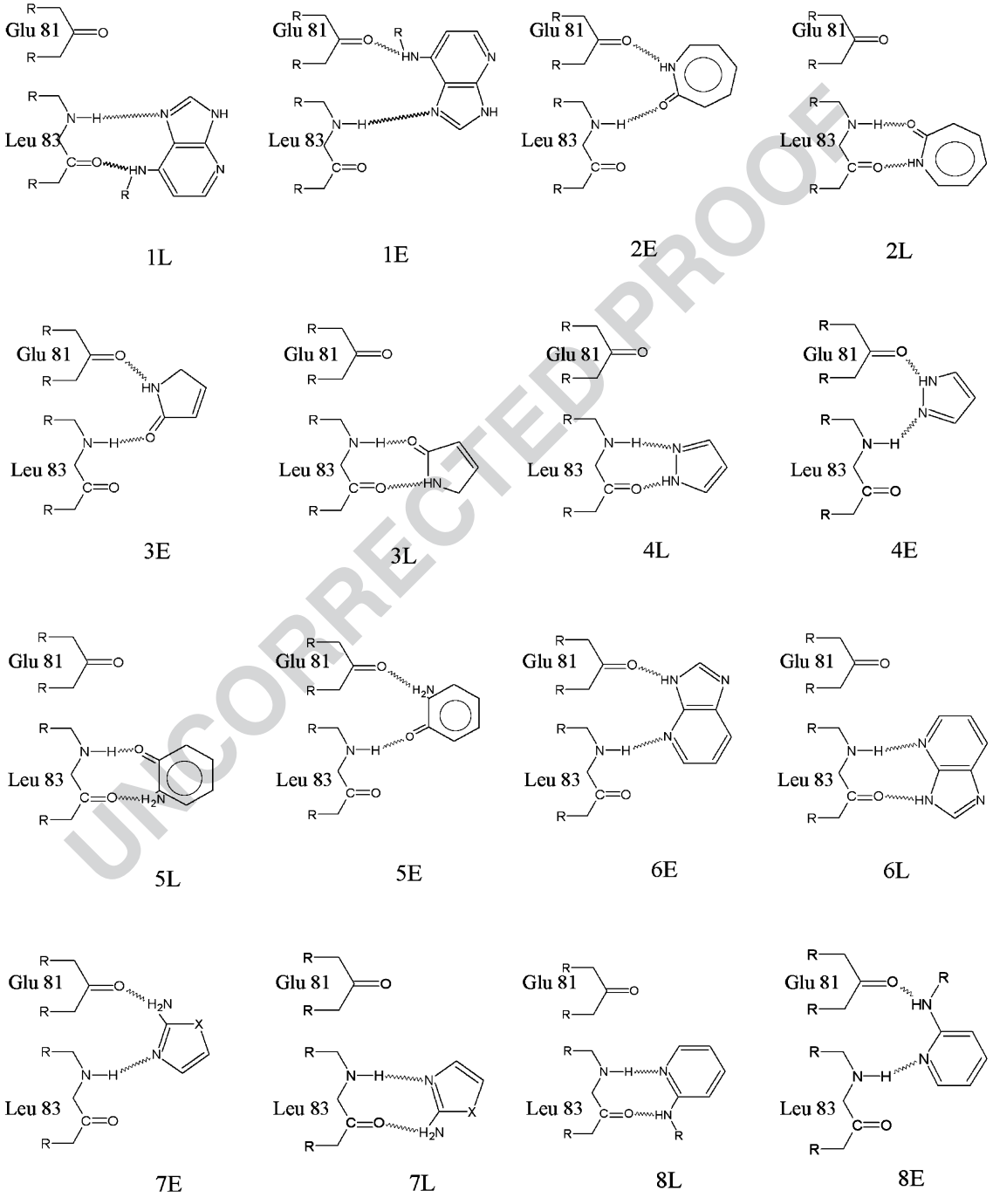
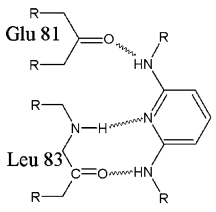
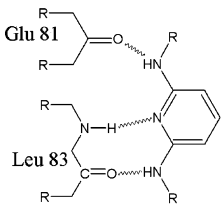


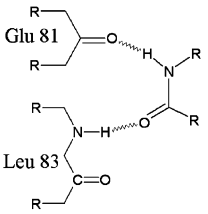
Table 2 (Continued)



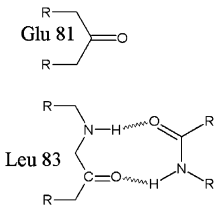
9EL



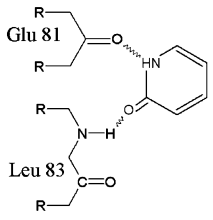
9EL



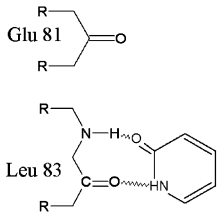
10E



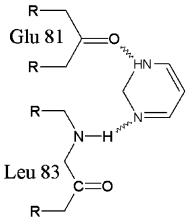
10L



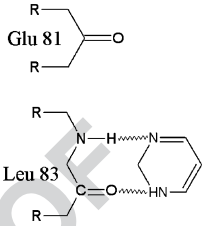
11E



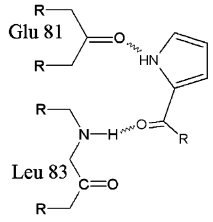
11L



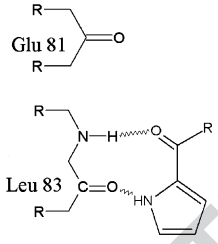
12E



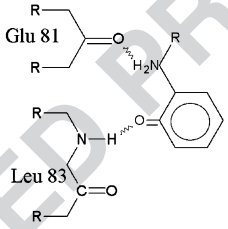
12L



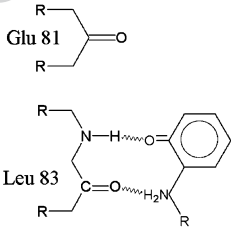
13E



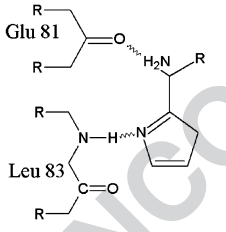
13L



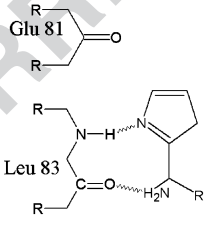
14E



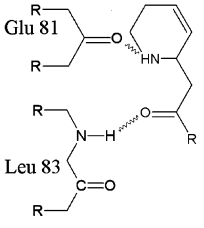
14L



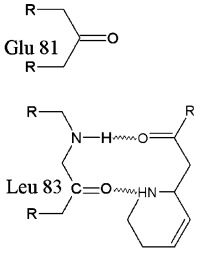
15E



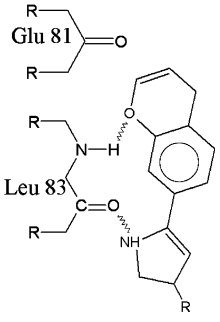
15L



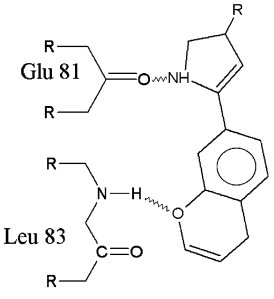
16E



16L



17L



17E

schematic representation, broken lines indicate hydrogen bonds, and water molecule HOH 100 and/or the additional site-points would be found at the bottom of each scaffold. Table 3 lists all molecular scaffolds that were obtained with each ligand-design strategy. Each scaffold is identified by a number and subdivided into three classes (**E**, **L**, and **EL**) depending on their binding mode, respectively, interacting with Glu 81, only with Leu 83, and both Glu 81 and Leu 83.

We can see that the scaffolds fall naturally into different classes depending on their hydrogen-bonding patterns with the three typical protein site-points (Glu 81 C=O, Leu 83 C=O and Leu 83 NH), the nature of the donor and acceptor atoms on the scaffold, the number of bonds separating them, and the chemical type of ring they have.

For example, scaffold **1L** was found in the binding mode depicted within the ligands generated with all three ligand design strategies. The same scaffold was also found with an alternative binding mode (shown in scaffold **1E**). Since the scaffold has rotated by about 180° it now has a different hydrogen-bonding pattern with the three typical site-points. Scaffold **1E** was found with ligand-design strategies A and C, but not with B (which included water molecule HOH 100).

All of the nine molecular scaffolds that were found in a previous validation study of Skelgen [20] were identified with ligand-design strategy A (using the typical site-points). These scaffolds are **1L** (observed, for example, in the ligand in 1ckp), **2E** (observed, for example, in the ligand in 1dm2), **3E** (observed, for example, in the ligands in 1 aq1, 1 fvt, 1 h0w and 1 ke5 to 1 ke9), **4L** (observed, for example, in the ligand in 1 jvp), **5L**, **6E**, **7E**, **8L** (observed, for example, in the ligands in 1 jsv and 1 h0w) and **9EL** (observed, for example in the ligands in 1e1x, 1e1v, 1h0u, 1h0v and 1 h0w). It is worth noting that the earlier Skelgen validation study [20] also correctly identified five chemical classes or binding motifs for CDK2 that had been determined previously. Several other scaffolds and their corresponding binding modes were also found with strategy A (as can be seen in Table 2). Interestingly, some of these scaffolds and binding modes were also found in the other two strategies (B and C). The ability to identify new chemical entities with new binding motifs is perhaps the highest value that de novo design can provide.

ATP in one of the crystal structures of CDK2 (1 hck [42]) has a binding mode like that seen for scaffold **7E** (except that ATP has a six-membered ring instead of the five-membered ring of scaffold **7E**), with the two nitrogens of its six-membered ring participating in hydrogen-bonding: N1 accepts a hydrogen bond from Leu 83 NH and N6 donates a hydrogen bond to Glu 81 C=O. A series of inhibitors have been reported that has the structure of a modified guanine that interacts with the binding site in the same way as that for scaffolds **6E** and **9EL** [41]. Scaffold **9EL** can also be seen in the ligands found in crystal structures 1 e1v, 1 e1x, 1 h0u and 1 h0v. This scaffold is symmetrical, and was generated with all three ligand-design strategies, which indicates that it is a versatile scaffold that allows ligands that contain it to interact with all of the typical protein site-points and either water molecule HOH 100 or the additional site-points. The above examples illustrate the agreement that exists between the experimentally observed binding modes of ATP and inhibitors of CDK2 and those of ligands generated in silico in this study.

The molecular structure of scaffold **7L** is contained in recently disclosed clinical candidates for drugs that inhibit CDK2 [43], and is shown in Fig. 2. The crystal structure has not yet been disclosed for the structure of CDK2 complexed with this inhibitor. Our modeled binding mode of this molecule in the binding site of CDK2 found that the core interactions with Glu 81 and Leu 83 are preserved and that the piperazine ring of the ligand occupies the position of water molecule HOH 100, indicating that this inhibitor displaces this water molecule upon binding. The reported pictorial representation of the crystal structure of this inhibitor bound to CDK2 appears to confirm this prediction [43].

Ligands that combine several of the molecular scaffolds and binding modes shown in Table 2 are of interest because they are likely to have appropriate interactions with the protein that would enhance ligand binding. Such ligands might be useful in the search for new lead compounds. For example, the ligand shown in Figs. 3a and 4a (generated with design strategy B) combines the hydrogen-bonding interactions of scaffolds **6L** and **7E**, as well as hydrogen bonding with water molecule HOH 100.

An important observation is needed here before proceeding to analyze the effect of the presence of the tightly bound water molecule. The use of more site-points (ligand-design strategies B and C) would inevitably lead to the generation of larger ligands due to the larger distance between all site-points considered. Therefore, it was necessary to compare the structures and binding modes only in the region of the typical protein site-points (Glu 81 C=O, Leu 83 C=O and Leu 83 NH) in order to distinguish the effects produced by the presence of water molecule HOH 100 or the additional site-points. This would allow us to investigate the availability of specific molecular scaffolds and their binding modes near the common typical protein site-points under the influence of the tightly bound water molecule and/or its associated additional site-points.

t3.1 **Table 3** Classification of molecular scaffolds according to the ligand design strategy

t3.2 A = standard	B = including water	C = additional site-points
t3.3 1L, 1E, 2E, 2L, 3E, 3L, 4L, 4E, 5L, 5E, 6E, 6L, 7E, 7L, 8L, 8E, 9EL, 10E, 10L, 11E, 11L, 12E, 13E, 13L, 14E, 14L, 15E, 15L, 16E, 16L, 17L, 17E	1L, 2L, 3L, 4L, 5L, 6L, 8L, 8E, 9EL, 10E, 10L, 11L, 13L, 14L, 15L, 17L	1L, 1E, 2L, 3E, 3L, 4L, 4E, 5L, 5E, 6E, 6L, 7E, 7L, 8L, 8E, 9EL, 10E, 10L, 11E, 11L, 13L, 14L, 15E, 15L, 16L, 17E

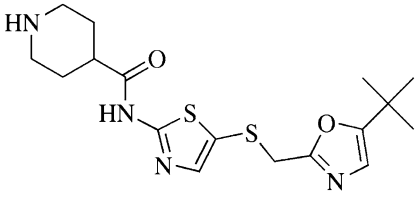


Fig. 2 Chemical structure of the BMS-387032 inhibitor

All of the scaffolds generated with ligand-design strategy B (with water molecule HOH 100) were also generated with design strategies A and C (except for scaffold **17L**, which was not generated with design strategy C). On the other hand, several scaffolds and/or their alternative binding modes were only generated with design strategies A and C, but not with design strategy B. This suggests that it is more difficult (i.e. the chemical diversity is more limited) to find a ligand that can interact with the typical protein site-points and with the water molecule. Fig. 3b shows the side view of the binding mode of the ligand shown in Fig. 3a (its chemical structure is shown in Fig. 4a). This ligand can interact with all three typical site-points and with water molecule HOH 100. We can see that these site-points and the water molecule lie in a common plane within the binding site. It is possible that such an arrangement introduces what we have named as *geometric constraints* on the placement of molecular scaffolds, where only certain scaffolds can be used in ligands that can satisfy all hydrogen-bonding interactions and have substituent groups at an appropriate hydrogen-bonding distance from the water molecule (close enough to form a hydrogen bond but not too close to give rise to a steric clash).

Nearly all of the scaffolds generated with ligand-design strategy B interact with the protein by forming hydrogen bonds with both the Leu 83 NH and Leu 83 C=O site-points. Few cases were found in which a scaffold was interacting with both the Leu 83 NH and Glu 81 C=O site-points (scaffolds **8E** and **10E**), and there was only one scaffold that interacted with all three site-points (the symmetric scaffold **9EL**). However, there were multiple instances of ligands that were generated with design strategies A and C that had scaffolds that made hydrogen bonds to both Leu 83 NH and Glu 81 C=O. It then becomes apparent that including the tightly bound water molecule HOH 100 as an interaction site-point restricts the binding modes available to molecular scaffolds and, in doing so, restricts the chemical diversity of the scaffolds that can be generated. Consequently, in the presence of water molecule HOH 100 it is easier to generate ligands that form hydrogen bonds with both Leu 83 NH and Leu 83 C=O and that are also able to form hydrogen bonds with the water molecule.

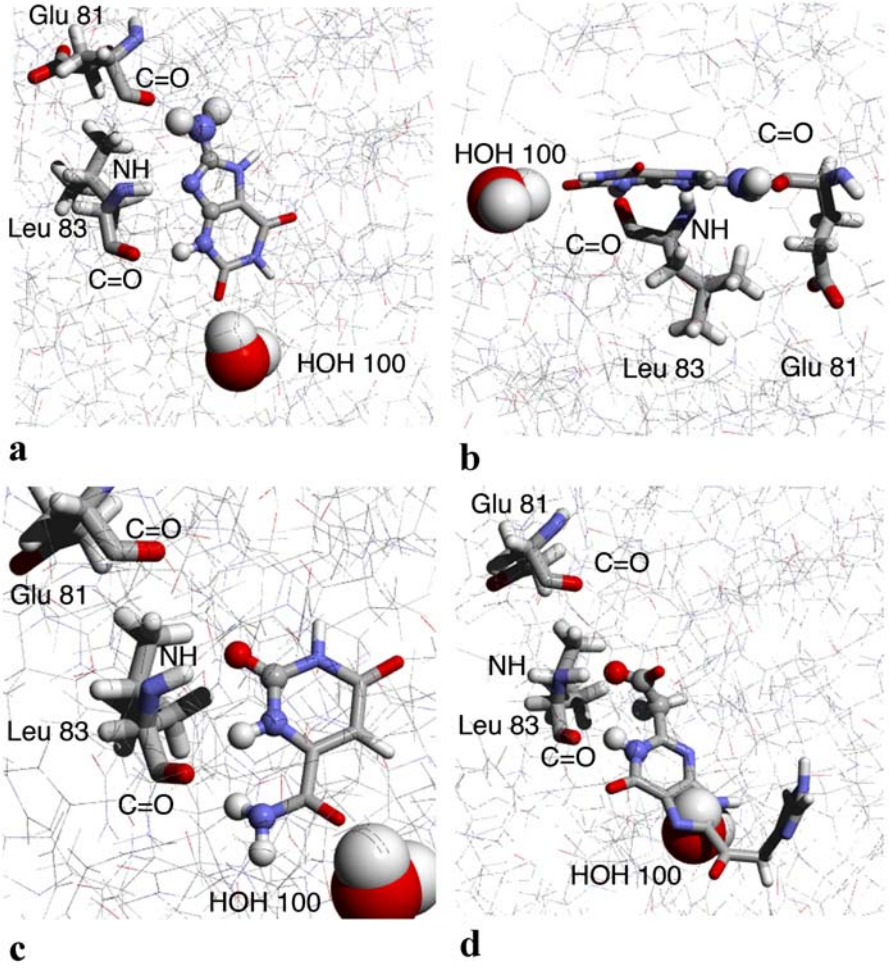


Fig. 3 **a** Ligand generated in strategy B that combines the fragments shown in scaffolds **6L** and **7E** in Table 2. **b** Side view of the binding mode of the previous ligand, showing all typical site-points and water molecule HOH 100 in the same plane within the binding site. **c** Ligand generated in strategy B that combines the interactions seen for scaffolds **11L** and **15L**, it also interacts with water molecule HOH 100. **d** Ligand generated in strategy C that contains scaffold **16L** superimposed into the binding site, shown clashing with water HOH 100

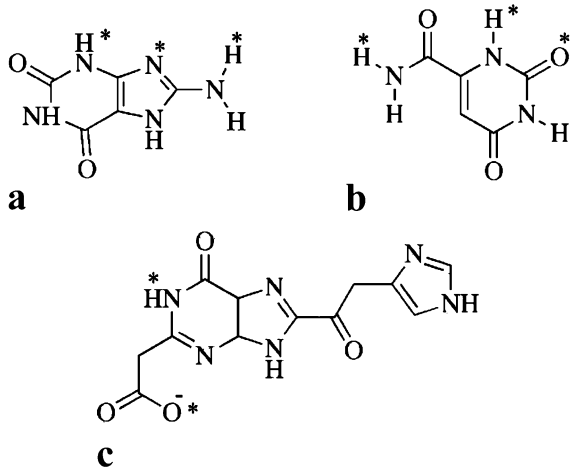


Fig. 4 **a** Chemical structure of the ligand shown in Fig. 3a. Atoms marked with a *star* (*) represent atoms within hydrogen-bonding distance from the protein. **b** Chemical structure of the ligand shown in Fig. 3c. **c** Chemical structure of the ligand seen in Fig. 3d

An example of this type of molecular scaffold in a ligand structure can be seen in Fig. 3c, while its corresponding chemical structure is shown in Fig. 4c. The hydrogen-bonding interactions of this ligand with the typical protein site-points are provided by a combination of the interactions seen in the binding modes of scaffolds **11L** and **15L** (as shown in Table 2). Fig. 3d, on the other hand, shows a ligand containing scaffold **16L** that was generated with strategy C but that was not found with strategy B. This scaffold (in conjunction with scaffold **3E**) can be found in the ligands of crystal structures 1 ke5 to 1 ke9. We can see that the scaffold itself has a steric clash with water molecule HOH 100, preventing it from being incorporated into any ligand in a design strategy that incorporates this water molecule. The chemical structure of the ligand is shown in Fig. 4d. All of the above observations portray a picture where a structure-based drug-design strategy that includes tightly bound water molecules may have a significant effect on the types of molecular scaffolds and ligands (and their binding modes) that can be generated. In this context, the fact that a number of scaffolds were seen to adopt alternative binding modes was due to the fact that only small ligands were generated within the specified region of the binding site. One would not expect all of these alternative binding modes to be observed in larger ligands occupying a greater extent of the binding site. Thus, reducing the available binding modes for a scaffold would reduce the chemical diversity in larger ligands generated by constructing molecules based on those scaffolds.

It is difficult to assess whether ligands that possess appropriate molecular scaffolds that would allow them to interact with water molecule HOH 100 have better binding affinities to CDK2. In addition to the anilinoquinazoline ligand found in 1 di8, there are other inhibitors that appear to interact with this water molecule when bound to the active site of CDK2, such as roscovitine [37, 44] and isopentenyladenine [37, 45]. Nonetheless, it has been suggested that the appropriate replacement of tightly bound water molecules in the active site of CDK2 may result in an increase in binding affinity [37]. As more crystal structures become available, it may be possible to determine unambiguously the binding mode of some of the molecular scaffolds that we have investigated in this study. Furthermore, an experimental determination of the binding constants of ligands containing such scaffolds would help establish the relative importance of water molecule HOH 100 as it bridges the interaction between the ligands and the protein.

Conclusions

We have studied the effects that an experimentally observed tightly bound water molecule has on the computer-aided *de novo* design of CDK2 ligands. Ligand generation was carried out to satisfy a set of typical and widely used protein hydrogen-bonding groups and either a neighboring tightly bound water molecule or its associated hydrogen-bonding groups (which are not accessible when the water molecule is present). This *in silico* approach has yielded a significant number of known binding motifs, some of which can be observed in known active compounds. A number of new binding motifs have also been generated, corroborating the utility of *de novo* ligand design for suggesting novel chemical entities.

We have observed that the tightly bound water molecule modifies the size and shape of the binding site and, more importantly, we have also found that it imposes constraints on the observed binding modes of the ligands generated. This is due to the specified requirement that generated ligands have to interact (through hydrogen bonding) with this water molecule. Ligands generated under these conditions exhibit more restricted hydrogen-bonding patterns within the binding site, which in turn is translated into a reduced chemical diversity of the underlying molecular scaffolds.

Complementary to the finding that, in some cases, tightly bound water molecules satisfy hydrogen-bonding groups that would be otherwise inaccessible to an incoming ligand [18], we have concluded that tightly bound water molecules may have an influential role in determining the binding modes and chemical diversity of molecular scaffolds. These findings have implications for drug-design strategies that make use of tightly bound water molecules as potential hydrogen-bonding groups.

Acknowledgements ATGS is grateful to Consejo Nacional de Ciencia y Tecnología (CoNaCyT, México) for the award of a post-graduate scholarship and to the Universities UK for an Overseas Research Scheme award. The authors would like to thank Dr. Nikolay P. Todorov, Dr. Stuart Firth-Clarke and Dr. Christoph Buenemann for helpful and fruitful discussions.

References

1. Davis AM, Teague SJ, Kleywegt GJ (2003) Angew Chem Int Ed 42:2718-2736
2. Poornima CS, Dean PM (1995) J Comput-Aided Mol Des 9:500–512
3. Hendlich M, Bergner A, Günter J, Klebe G (2003) J Mol Biol 326:607–620
4. Chung E, Henriques D, Renzoni D, Zvelebil M, Bradshaw JM, Waksman G, Robinson CV, Ladbury JE (1998) Struct Fold Des 6:1141–1151
5. Rejto PA, Verkhivker GM (1997) Proteins Struct Funct Genet 28:313–324
6. Wester MR, Johnson EF, Marques-Soares C, Dijols S, Dansette PM, Mansuy D, Stout CD (2003) Biochem 42:9335–9345
7. Marrone TJ, Briggs JM, McCammon JA (1997) Ann Rev Pharmacol Toxicol 37:71–90
8. Lam PYS, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, Bacheler LT, Meek JL, Otto MJ, Rayner MM, Wong YN, Chang CH, Weber PC, Jackson DA, Sharpe TR, Ericksonviitanen S (1994) Science 263:380–384
9. Chen JM, Xu SL, Wawrzak Z, Basarab GS, Jordan DB (1998) Biochem 37:17735–17744
10. Mikol V, Papageorgiou C, Borer X (1995) J Med Chem 38: 3361–3367
11. Cherbavaz DB, Lee ME, Stroud RM, Koschl DE (2000) J Mol Biol 295:377–385
12. Finley JB, Atigadda VR, Duarte F, Zhao JJ, Brouillette WJ, Air GM, Luo M (1999) J Mol Biol 293:1107–1119
13. Rarey M, Kramer B, Lengauer T (1998) Proteins Struct Funct Genet 34:17–28
14. Schnecke V, Kuhn LA (2000) Perspect Drug Discov Des 20: 171–190
15. Pospisil P, Kuoni T, Scapozza L, Folkers G (2002) J Recept Signal Transduct Res 22:141–154
16. Pastor M, Cruciani G, Watson KA (1997) J Med Chem 40:4089–4102
17. Lloyd DG, García-Sosa AT, Alberts IL, Todorov NP, Mancera RL (2004) J Comput-Aided Mol Des 18:89–100
18. Mancera RL (2002) J Comput-Aided Mol Des 16:479–499
19. Todorov NP, Dean PM (1998) J Comput-Aided Mol Des 12: 335–349
20. Stahl M, Todorov NP, James T, Mauser H, Boehm H-J, Dean PM (2002) J Comput-Aided Mol Des 16:459–478
21. Gray NS, Wodicka L, Thunissen A-MWH, Norman TC, Kwon S, Espinoza FH, Morgan DO, Barnes G, LeClerc S, Meijer L, Kim S-H, Lockhart DJ, Schultz PG (1998) Science 281:533–538
22. Knockaert M, Greengard P, Meijer L (2002) Trends Phamacol Sci 23:417–425
23. Metz WA (2003) Bioorg Med Chem Lett 13:2953–2953
24. Beattie JF, Breault GA, Ellston RPA, Green S, Jewsbury PJ, Midgley CJ, Naven RT, Minshull CA, Pauptit RA, Tucker JA, Pease JE (2003) Bioorg Med Chem Lett 13:2955–2960
25. Breault GA, Ellston RPA, Green S, James SR, Jewsbury PJ, Midgley CJ, Pauptit RA, Minshull CA, Tucker JA, Pease JA (2003) Bioorg Med Chem Lett 13:2961–2966
26. Anderson M, Beattie JF, Breault GA, Breed J, Blyth KF, Culshaw JD, Ellston RPA, Green S, Minshull CA, Norman RA, Pauptit RA, Stanway J, Thomas AP, Jewsbury PJ (2003) Bioorg Med Chem Lett 13:3021–3026
27. Westwell AD (2003) Drug Discov Today 8:1094–1095
28. McGovern SL, Shoichet BK (2003) J Med Chem 46:1478–1483
29. Sayle KL, Bentley J, Boyle FT, Calvert AH, Cheng YZ, Curtin NJ, Endicott JA, Golding BT, Hardcastle IR, Jewsbury PJ, Mesguiche V, Newell DR, Noble MEM, Parsons RJ, Pratt DJ, Wang LZ, Griffin RJ (2003) Bioorg Med Chem Lett 13:3079–3082
30. Moravec J, Krystof V, Hanus J, Havlicek L, Moravcova D, Fuksova K, Kuzma M, Lenobel R, Otyepka M, Strnad M (2003) Bioorg Med Chem Lett 13:2993–2996
31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucl Acids Res 28:235–242
32. Shewchuk L, Hassell A, Wisely B, Rocque W, Holmes W, Veal J, Kuyper LF (2000) J Med Chem 43:133–138
33. Pierce AC, Sandretto KL, Bemis GW (2002) Proteins Struct Funct Genet 49:567–576
34. Todorov NP, Dean PM (1997) J Comput-Aided Mol Des 11: 175–192
35. Dinur U, Hagler AT (1991) In: Lipkowitz KB, Boyd DB (eds) Reviews in Computational Chemistry, vol 2. VCH Publishers Inc, USA
36. García-Sosa AT, Mancera RL, Dean PM (2003) J Mol Model 9:172–182
37. Kríz Z, Otyepka M, Bártová I, Koca J (2004) Proteins Struct Funct Bioinf 55:258–274
38. Krystof V, Strnad M (2001) Chem Listy 95:295–300
39. Davies TG, Pratt DJ, Endicott JA, Johnson LN, Noble MEM (2002) Pharmacol Therap 93:125–133
40. Hardcastle IR, Golding BT, Griffin RJ (2002) Annu Rev Pharmacol Toxicol 42:325–348
41. Gibson AE, Arris CE, Bentley J, Boyle T, Curtin NJ, Davies TG, Endicott JA, Golding BT, Grant S, Griffin RJ, Jewsbury P, Johnson LN, Mesguiche V, Newell DR, Noble MEM, Tucker JA, Whitfield HJ (2002) J Med Chem 45:3381–3393
42. Schulze-Gahmen U, De Bondt HL, Kim S-H (1996) J Med Chem 39:4540–4546
43. Misra RN, Xiao H, Kim KS, Han W-C, Barbosa SA, Hunt JT, Rawlins DB, Shan W, Ahmed SZ, Qian L, Chen B-C, Zhao R, Bednarz MS, Kellar KA, Mulheron JG, Batorsky R, Roongta U, Kamath A, Marathe P, Ranadive SA, Sack JS, Tokarski JS, Pavletich NP, Lee FYF, Webster KR, Kimball SD (2004) J Med Chem 47:1719–1728
44. de Azevedo WF, Mueller-Dieckman H-J, Schulze-Gahmen U, Worland PJ, Sausville EA, Kim SH (1996) Proc Natl Acad Sci USA 93:2735–2740
45. Schulze-Gahmen U, Brandsen J, Jones HD, Morgan D, Meijer L, Vesely J, Kim S-H Proteins Struct Funct Genet 22:378–391