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

Alfonso T. García-Sosa[1#] and Ricardo L. Mancera*[2]

1 Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1PD, UK; 2 Western Australian Biomedical Research Institute, School of Pharmacy and School of Biomedical Sciences, Curtin University of Technology, GPO Box U1987, Perth WA 6845, Australia

Abstract

We have investigated the effects that a tightly-bound water molecules has on the *de novo* design of cyclin-dependent kinase-2 (CDK2) ligands. In particular, we have analysed the impact of a specific 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

Current address: Computer-Aided Molecular Design Lab, Guggenheim Building, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA.

* To whom correspondence should be addressed: R.Mancera@wabri.org.au

1. 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 usually 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 stabilise 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 recent 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 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. As more of the 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 analysed the crystal structures of these proteins and found that cyclin-dependant 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.

1.1 *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 catalyse the transfer of a phosphate group from ATP to a specific substrate aminoacid 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 because of its intermolecular interactions via its adenine ring to Glu 81 (Glu 75 in structure 1di8) and Leu 83 (Leu 77 in structure 1di8), as well as via its triphosphate group. 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 Figure 1a interacting with Glu 81 and Leu 83 are found in the *apo*-enzyme (1hcl). Their presence confirms the importance of these hydrogen-bonding groups in the binding site. An analysis of the ligands shown in Figure 1a reveal that these ligands 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 sitepoints) 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. Figure 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) which 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.[31]

2. Results and Discussion

2.1 *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.[32] 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 analysed using a multivariate logistic regression approach. A probabilistic model was obtained, which 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 elsewhere.[32]

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 1di8). The position of these water molecules in the binding site of CDK2 can be seen in Figure 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.

2.2 De novo *ligand design.*

Due to the pre-eminence of the interactions of known inhibitors of CDK2 with sitepoints 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 sitepoints 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 1di8, the inhibitor does not interact directly with these sitepoints but a water molecule (HOH 100) interacts with most of them, as shown in Figure 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 1di8 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 sitepoints 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 sitepoints 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 sitepoints to an incoming ligand. An analysis of the other CDK2 crystal structures reveals that HOH 100 is also found as HOH 38Z in 1h0w and HOH 582 in 1dm2. In most of the other crystal structures, this water position is occupied by a polar group in the ligands (such as a sulfonamide group).

A recent molecular dynamics study of the hydration of the empty active site of CDK2 as well as complexed with ATP and two inhibitors has been recently reported.[33] 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,[33] 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 *de novo* ligand generation. The first "standard" strategy (named A) was to generate ligands that satisfy only two or three of the three typical sitepoints (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 sitepoints *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 sitepoints *and* all the additional sitepoints 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[34] and the fact that several inhibitors interact with Asp 86.[35,36] Table 1 summarises 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.[37] Furthermore, it is important to bear in mind that *de novo* ligand design methods may suggest chemically unreasonable or synthetically unfeasible molecules. Therefore we have focused our investigation to the analysis of known molecular scaffolds.

2.3 *Evaluation of* de novo *generated ligands*

Any minimised ligand that did not make hydrogen bonds with at least two out of the three typical sitepoints (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.

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

2.4 ***Scaffold Analysis***

Table 2 lists all molecular scaffolds that were obtained with each ligand design strategy. Each scaffold is identified by a number and subdivided into two classes (a and b) depending on their binding mode. Table 3 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 schematic representation, broken lines indicate hydrogen bonds, and water molecule HOH 100 and/or the additional sitepoints would be found at the bottom of each scaffold.

We can see that the scaffolds fall naturally into different classes depending on their hydrogen-bonding patterns with the three typical protein sitepoints (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 **1a** was found in the depicted binding mode within the ligands generated with all three ligand design strategies. The same scaffold was also found with an alternative binding mode (shown in scaffold **1b)**. Since the scaffold has rotated by about 180º it now has a different hydrogen bonding pattern with the three typical sitepoints. Scaffold **1b** 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 sitepoints). These scaffolds are **1a** (observed, for example, in the ligand in 1ckp), **2a** (observed, for example, in the ligand in 1dm2), **3a** (observed, for example, in the ligands in 1aq1, 1fvt, 1h0w and 1ke5 to 1ke9), **4a** (observed, for example, in the ligand in 1jvp), **5a**, **6a**, **7a**, **8a** (observed, for example, in the ligands in 1jsv and 1h0w) and **9a** (observed, for example in the ligands in 1e1x, 1e1v, 1h0u,

1h0v and 1h0w). 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 previously determined. Several other scaffolds and their corresponding binding modes were also found with strategy A (as can be seen in Table 3). 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 has been proposed to be the highest value that *de novo* design can provide.

ATP in one of the crystal structures of CDK2 (1hck[38]) has a binding mode like that one seen for scaffold **7a** (except that ATP has a six-membered ring instead of the five-membered ring of scaffold **7a**), 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 which have the structure of a modified guanine that interacts with the binding site in the same way as that for scaffolds **6a** and **9a/b.**[37] Scaffold **9a/b** can also be seen in the ligands found in crystal structures 1e1v, 1e1x, 1h0u and 1h0v. 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 sitepoints and either water molecule HOH 100 or the additional sitepoints. 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 **7b** is contained in recently disclosed clinical candidates for drugs that inhibit CDK2,[39] and is shown in Figure 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.[39]

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 Figures 3a and 4a (generated with design strategy B) combines the hydrogen-bonding interactions of scaffolds **6b** and **7a**, as well as hydrogen bonding with water molecule HOH 100.

An important observation is needed here before proceeding to analyse the effect of the presence of the tightly-bound water molecule. The use of more sitepoints (ligand design strategies B and C) would inevitably lead to the generation of larger ligands due to the larger distance between all sitepoints considered. Therefore, it was necessary to compare the structures and binding modes only in the region of the typical protein sitepoints (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 sitepoints. This would allow us to investigate the availability of specific molecular scaffolds and their binding modes near the common typical protein sitepoints under the influence of the tightly-bound water molecule and/or its associated additional sitepoints.

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 **17a**, 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 sitepoints and with the water

molecule. Figure 3b shows the sideview of the binding mode of the ligand shown in Figure 3a (its chemical structure is shown in Figure 4a). This ligand can interact with all three typical sitepoints and with water molecule HOH 100. We can see that these sitepoints and the water molecule lie in a common plane within the binding site. It is possible that such 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 sitepoints. Few cases were found in which a scaffold was interacting with both the Leu 83 NH and Glu 81 C=O sitepoints (scaffolds **8b** and **10a**), and there was only one scaffold that interacted with all three sitepoints (the symmetric scaffold **9a/b**). However, there were multiple instances of ligands that were generated with design strategies A and C that had scaffolds which 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 sitepoint 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 which are also able to form hydrogen bonds with the water molecule. An example of this type of molecular scaffold in a ligand structure can be seen in Figure 3c, while its corresponding chemical structure is shown in Figure 4c. The hydrogen-bonding interactions of this ligand with the typical protein sitepoints are provided by a combination of the *interactions* seen in the binding modes of scaffolds **11b** and **15b** (as shown in Table 2). Figure 3d, on the other hand, shows a ligand containing scaffold **16b** that was generated with

strategy C but that was not found with strategy B. This scaffold (in conjunction with scaffold **3a**) can be found in the ligands of crystal structures 1ke5 to 1ke9. 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 Figure 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.

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 1di8, there are other inhibitors that appear to interact with this water molecule when bound to the active site of CDK2, such as roscovitine[33,40] and isopentenyladenine.[33,41] 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.[33] 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.

3. 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 neighbouring 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 generated ligands. 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.

4. Experimental

A survey of the Protein Data Bank[42] 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 aminoacid sequence). One of these crystal structures is that of the *apo*-enzyme (1hcl), whilst the other crystal structures contain either ATP (1hck and 1fin) or an inhibitor bound to the ATP site (1aqi, 1ckp, 1di8, 1dm2, 1eiv, 1eix, 1fvt, 1h0u, 1h0v, 1h0w, 1jsv, 1jvp, 1ke5, 1ke6, 1ke7, 1ke8 and 1ke9). 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 1di8 was used. This structure was determined to a resolution of 2.2 Å.[42] 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]

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 optimisation 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 to gradually satisfy the protein binding site constraints. This process is carried out in a stochastic manner to gradually optimise the interaction properties and chemical features of the generated ligand during the annealing optimisation. The assembled ligands must satisfy user-defined geometric constraints, such as those defining hydrogen bond distances and angles for pre-selected 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.[19,20,44] 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 minimised using the Discover 3 module in InsightII 2000 (Accelrys) with the CFF forcefield.[45] 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 minimisations; however, hydrogen atoms in any aminoacid in the binding site with at least

one atom within 3.5 Å of the ligand were allowed to re-orient in order to optimise the hydrogen-bonding network between the ligand, the water molecule (if present) and the protein. The ligands were allowed full flexibility during the minimisations. Water molecules were kept in their original crystal structure positions but were allowed to re-orient their hydrogen atoms. The energy minimisations were conducted in stages as described elsewhere[18] to try to retain the original binding mode. The minimisations were stopped when the energy gradient reached a value of less than 0.01 kcal/mol/Å.

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. RLM is also a Research Fellow of Hughes Hall, Cambridge, UK. The authors would like to thank Drs. Nikolay Todorov, Stuart Firth-Clarke and Christoph Buenemann for helpful and fruitful discussions.

References

1. Davis, A.M.; Teague, S.J.; Kleywegt, G.J. *Angew. Chem. Int. Ed.*, **2003**, *42*, 2718.

2. Poornima, C.S.; Dean, P.M. *J. Comput-Aided Mol. Des.*, **1995**, *9*, 500.

3. Hendlich, M.; Bergner, A.; Günter, J.; Klebe, G. *J. Mol. Biol.*, **2003**, *326*, 607.

4. Chung, E.; Henriques, D.; Renzoni, D.; Zvelebil, M.; Bradshaw, J.M.; Waksman, G.; Robinson, C.V.; Ladbury, J.E. *Struct. Fold. Des*., **1998**, *6*, 1141.

5. Rejto, P.A.; Verkhivker, G.M *Proteins Struct. Funct. Genet*. **1997** *28*, 313.

6. Wester, M.R.; Johnson, E.F.; Marques-Soares, C.; Dijols, S.; Dansette, P.M.; Mansuy, D.; Stout, C.D. *Biochem*., **2003**, *42*, 9335.

7. Marrone, T.J.; Briggs, J.M.; McCammon, J.A. *Ann. Rev. Pharmacol. Toxicol.*, **1997**, *37*, 71.

8. Lam, P.Y.S.; Jadhav, P.K.; Eyermann, C.J.; Hodge, C.N.; Ru, Y.; Bacheler, L.T.; Meek, J.L.; Otto, M.J.; Rayner, M.M.; Wong, Y.N.; Chang, C.H.; Weber, P.C.; Jackson, D.A.; Sharpe, T.R.; Ericksonviitanen, S. *Science*, **1994**, *263*, 380.

9. Chen, J.M.; Xu, S.L.; Wawrzak, Z.; Basarab, G.S.; Jordan, D.B. *Biochem.*, **1998**, *37*, 17735.

10. Mikol, V.; Papageorgiou, C.; Borer, X. *J. Med. Chem.*, **1995**, *38*, 3361.

11. Cherbavaz, D.B.; Lee, M.E.; Stroud, R.M.; Koschl, D.E. *J. Mol. Biol*., **2000**, *295*, 377.

12. Finley, J.B.; Atigadda, V.R.; Duarte, F.; Zhao, J.J.; Brouillette, W.J.; Air, G.M.; Luo, M. *J. Mol. Biol.*, **1999**, *293,* 1107.

13. Rarey, M.; Kramer, B.; Lengauer, T. *Proteins Struct. Funct. Genet.*, **1998**, *34*, 17.

14. Schnecke, V.; Kuhn, L.A. *Perspect. Drug Discov. Des.*, **2000**, *20,* 171.

15. Pospisil, P.; Kuoni, T.; Scapozza, L.; Folkers, G. *J. Recept. Signal Transduct. Res*., **2002**, *22*, 141.

16. Pastor, M.; Cruciani, G.; Watson, K.A. *J. Med. Chem.*, **1997**, *40*, 4089.

17. Lloyd, D.G.; García-Sosa, A.T.; Alberts, I.L; Todorov, N.P.; Mancera, R.L. *J. Comput.-Aided Mol. Des.*, **2004**, *18*, 19.

18. Mancera, R.L. *J. Comput.-Aided Mol. Des*., **2002**, *16*, 479.

19. Todorov.; N.P.; Dean, P.M. *J. Comput.-Aided Mol. Des*., **1998**, *12*, 335.

20. Stahl, M.; Todorov, N.P.; James, T.; Mauser, H.; Boehm, H.-J.; Dean, P.M. *J. Comput.-Aided Mol. Des*., **2002**, *16*, 459.

21. Gray, N.S.; Wodicka, L.; Thunissen, A.-M.W.H.; Norman, T.C.; Kwon, S.; Espinoza, F.H.; Morgan, D.O.; Barnes, G.; LeClerc, S.; Meijer, L.; Kim, S.-H.; Lockhart, D.J.; Schultz, P.G. *Science*, **1998**, *281*, 533.

22. Knockaert, M; Greengard, P; Meijer, L *Trends Phamacol. Sci.*, **2002**, *23*, 417.

23. Metz, W.A. *Bioorg. Med. Chem. Lett*., **2003**, *13*, 2953.

24. Beattie, J.F.; Breault, G.A.; Ellston, R.P.A.; Green, S.; Jewsbury, P.J.; Midgley, C.J.; Naven, R.T.; Minshull, C.A.; Pauptit, R.A.; Tucker, J.A.; Pease, J.E. *Bioorg. Med. Chem. Lett*., **2003**, *13*, 2955.

25. Breault, G.A.; Ellston, R.P.A.; Green, S.; James, S.R.; Jewsbury, P.J.; Midgley, C.J.; Pauptit, R.A.; Minshull, C.A.; Tucker, J.A.; Pease, J.A. *Bioorg. Med. Chem. Lett*., **2003**, *13*, 2961.

26. Anderson, M.; Beattie, J.F.; Breault, G.A.; Breed, J.; Blyth, K.F.; Culshaw, J.D.; Ellston, R.P.A.; Green, S.; Minshull, C.A.; Norman, R.A.; Pauptit, R.A.; Stanway, J.; Thomas, A.P.; Jewsbury, P.J. *Bioorg. Med. Chem. Lett*., **2003**, *13*, 3021.

27. Westwell, A.D. *Drug Discov. Today*, **2003**, *8*, 1094.

28. McGovern, S.L.; Shoichet, B.K. *J. Med. Chem*., **2003**, *46*, 1478.

29. Sayle, K.L.; Bentley, J.; Boyle, F.T.; Calvert, A.H.; Cheng, Y.Z.; Curtin, N.J.; Endicott, J.A.; Golding, B.T.; Hardcastle, I.R.; Jewsbury, P.J.; Mesguiche, V.; Newell, D.R.; Noble, M.E.M.; Parsons, R.J.; Pratt, D.J.; Wang, L.Z.; Griffin, R.J. *Bioorg. Med. Chem. Lett*., **2003**, *13*, 3079 .

30. Moravec, J.; Krystof, V.; Hanus, J.; Havlicek, L.; Moravcova, D.; Fuksova, K.; Kuzma, M.; Lenobel, R.; Otyepka, M.; Strnad, M. *Bioorg. Med. Chem. Lett*., **2003**, *13*, 2993.

31. Pierce, A.C; Sandretto, K.L.; Bemis, G.W. *Proteins Struct. Funct. Genet.*, **2002**, *49*, 567.

32. García-Sosa, A.T.; Mancera, R.L.; Dean, P.M. *J. Mol. Model.*, **2003**, *9*, 172.

33. Kříž, Z.; Otyepka, M.; Bártová, I.; Koča, J. *Proteins Struct. Funct. Bioinf.*, **2004**, *55*, 258.

34. Kryštof, V.; Strnad, M. *Chem. Listy*, **2001**, *95*, 295.

35. Davies, T.G.; Pratt, D.J.; Endicott, J.A.; Johnson, L.N.; Noble, M.E.M. *Pharmacol. Therap.*, **2002**, *93*, 125.

36. Hardcastle, I.R.; Golding, B.T.; Griffin, R.J. *Annu. Rev. Pharmacol. Toxicol.*, **2002**, *42*, 325.

37. Gibson, A.E.; Arris, C.E.; Bentley, J.; Boyle, T.; Curtin, N.J.; Davies, T.G.; Endicott, J.A.; Golding, B.T.; Grant, S.; Griffin, R.J.; Jewsbury P.; Johnson, L.N.; Mesguiche, V.; Newell, D.R.; Noble, M.E.M.; Tucker, J.A.; Whitfield, H.J. *J. Med. Chem*., **2002**, *45*, 3381.

38. Schulze-Gahmen, U.; De Bondt, H.L.; Kim, S.-H. *J. Med. Chem*., **1996**, *39*, 4540.

39. Misra, R.N.; Xiao, H.; Kim, K.S.; Han, W.-C.; Barbosa, S.A.; Hunt, J.T.; Rawlins, D.B.; Shan, W.; Ahmed, S.Z.; Qian, L.; Chen, B.-C.; Zhao, R.; Bednarz, M.S.; Kellar, K.A.; Mulheron, J.G.; Batorsky, R.; Roongta, U.; Kamath, A.; Marathe, P.; Ranadive, S.A.; Sack, J.S.; Tokarski, J.S.; Pavletich, N.P.; Lee, F.Y.F.; Webster, K.R.; Kimball, S.D. *J. Med. Chem*., **2004**, *47*, 1719.

40. De Azevedo, W.F.; Mueller-Dieckman, H.-J.; Schulze-Gahmen, U.; Worland, P.J.; Sausville, E.A.; Kim, S.H. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*, 2735.

41. Schulze-Gahmen, U.; Brandsen, J.; Jones, H.D.; Morgan, D.; Meijer, L.; Veselý, J.; Kim, S.-H. *Proteins Struct. Funct. Genet.*, **1995**, *22*, 378.

42. Berman. H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. *Nucl. Acids. Res*., **2000**, *28*, 235.

43. Shewchuk, L.; Hassell, A.; Wisely, B.; Rocque, W.; Holmes, W.; Veal, J.; Kuyper, L.F. *J. Med. Chem*., **2000**, *43*, 133.

44. Todorov, N.P.; Dean, P.M. *J. Comput.-Aided Mol. Des*., **1997**, *11*, 175.

45. Dinur, U.; Hagler, A.T. In Reviews in Computational Chemistry; Lipkowitz, K.B.; Boyd, D.B., Eds., VCH Publishers Inc.: USA, 1991. Vol. 2.

Figure captions

Figure 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 (1di8) with its co-crystallised 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.

Figure 2. Chemical structure of the BMS-387032 inhibitor.

Figure 3. a) Ligand generated in strategy B that combines the fragments shown in scaffolds **6b** and **7a** in Table 2. **b)** Side view of the binding mode of the previous ligand, showing all typical sitepoints 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 **11b** and **15b**; it also interacts with water molecule HOH 100. **d)** Ligand generated in strategy C that contains scaffold **16b** superimposed into the binding site, shown clashing with water HOH 100.

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

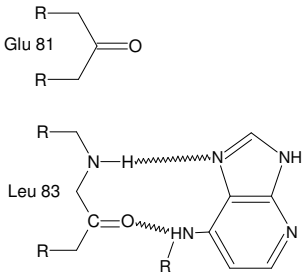
Table 1. Summary of strategies for *de novo* ligand generation

Strategy	**Sitepoints used**
A	Glu 81 C=O
= Standard	Leu 83 NH
	Leu 83 C=O
B	Glu 81 C=O
= Including water	Leu 83 NH
	Leu 83 C=O
	HOH 100
C	Glu 81 C=O
= Additional sitepoints	Leu 83 NH
	Leu 83 C=O
	Asp 86 Oδ
	Asp 86 N
	Lys 89 Nζ

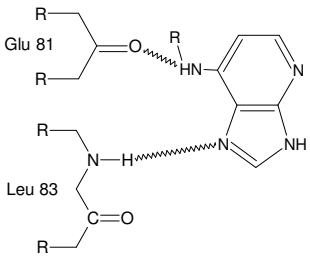
Table 2. Classification of molecular scaffolds according to the ligand design strategy

A = Standard	B = Including water	C = Additional sitepoints
1a, 1b, 2a, 2b, 3a, 3b, 4a, 4b, 5a, 5b, 6a, 6b, 7a, 7b, 8a, 8b, 9a, 9b, 10a, 10b, 11a, 11b, 12a, 13a, 13b, 14a, 14b, 15a, 15b, 16a, 16b, 17a, 17b	**1a, 2b, 3b, 4a, 5a, 6b, 8a, 8b, 9a, 9b, 10a, 10b, 11b, 13b, 14b, 15b, 17a**	**1a, 1b, 2b, 3a, 3b, 4a, 4b, 5a, 5b, 6a, 6b, 7a, 7b, 8a, 8b, 9a, 9b, 10a, 10b, 11a, 11b, 13b, 14b, 15a, 15b, 16b, 17b**

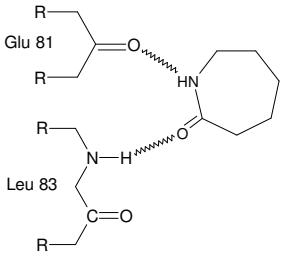
Table 3. Molecular scaffold classes and their binding modes (with X = O, N, S)



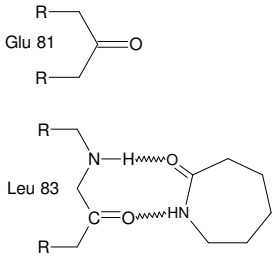
1a



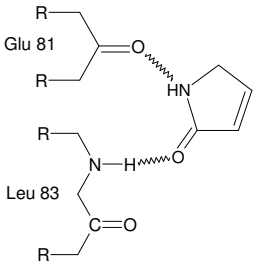
1b



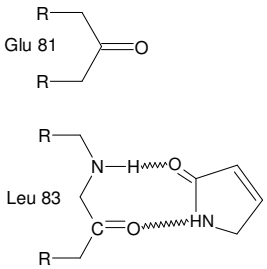
2a



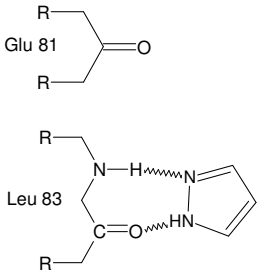
2b



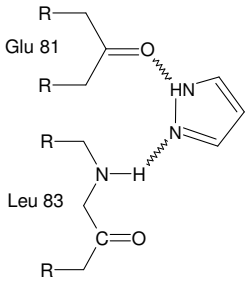
3a



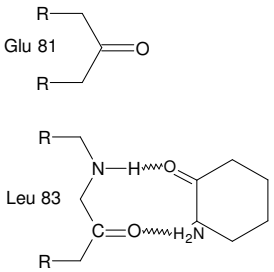
3b



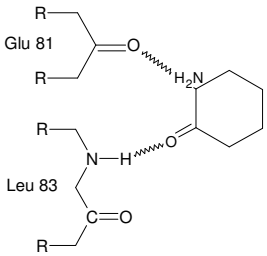
4a



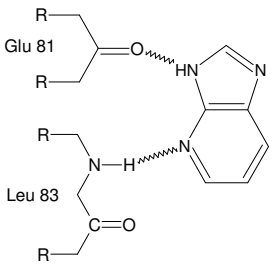
4b



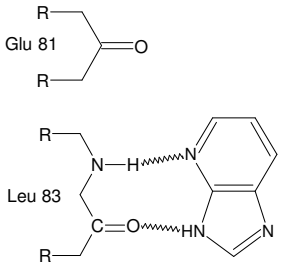
5a



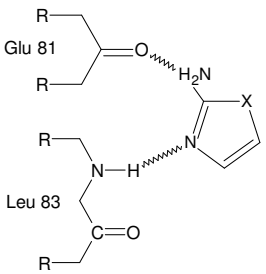
5b



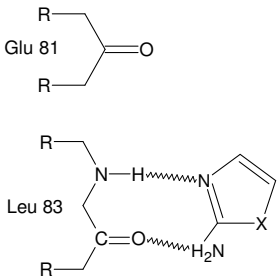
6a



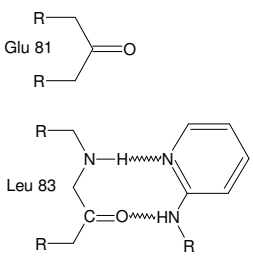
6b



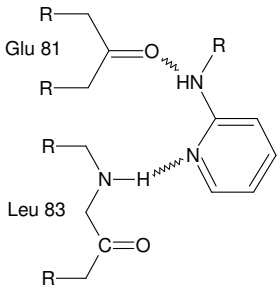
7a



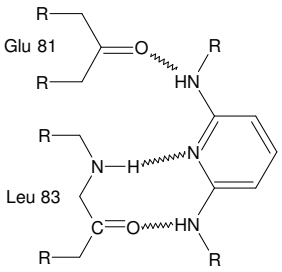
7b



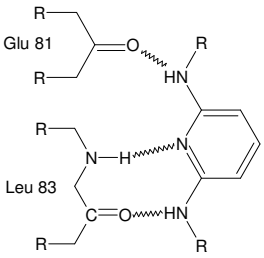
8a



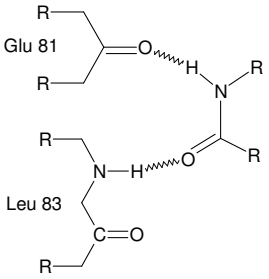
8b



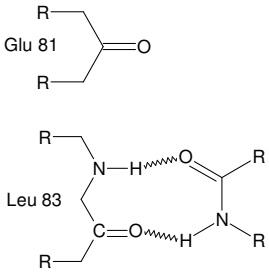
9a



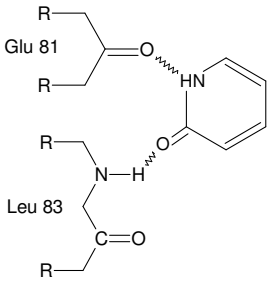
9b



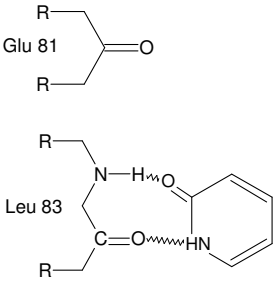
10a



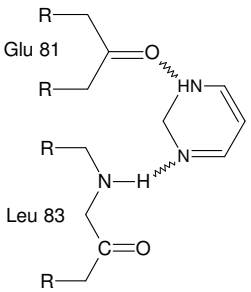
10b



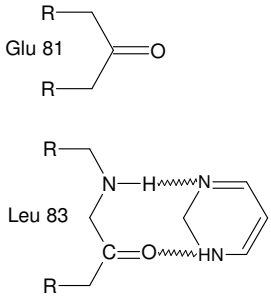
11a



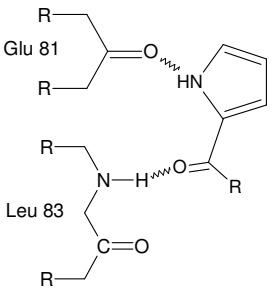
11b



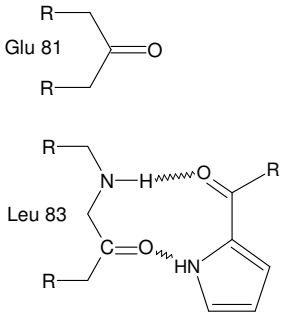
12a



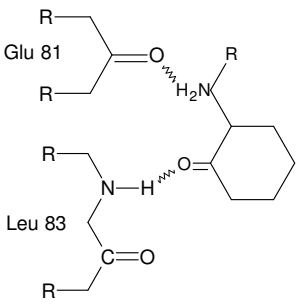
12b



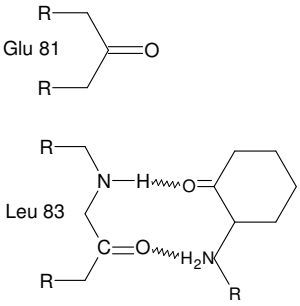
13a



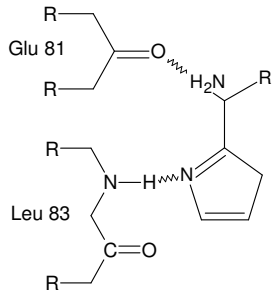
13b



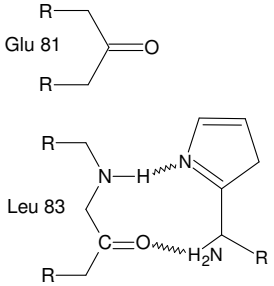
14a



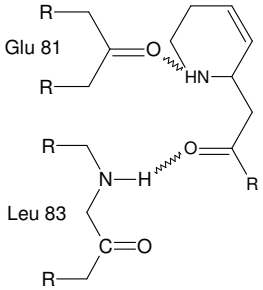
14b



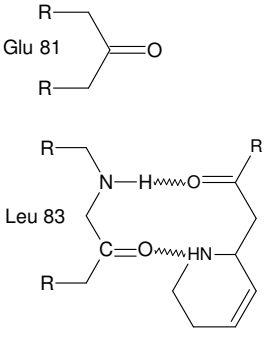
15a



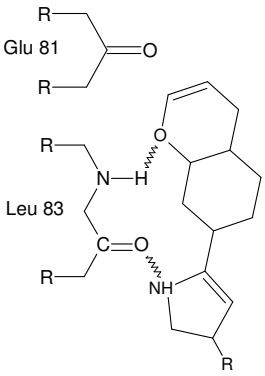
15b



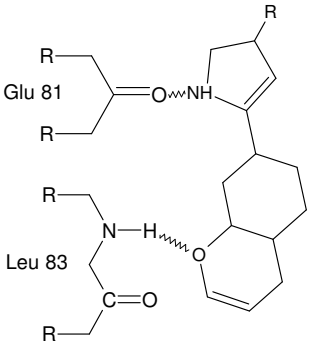
16a



16b



17a



17b

Figure 1

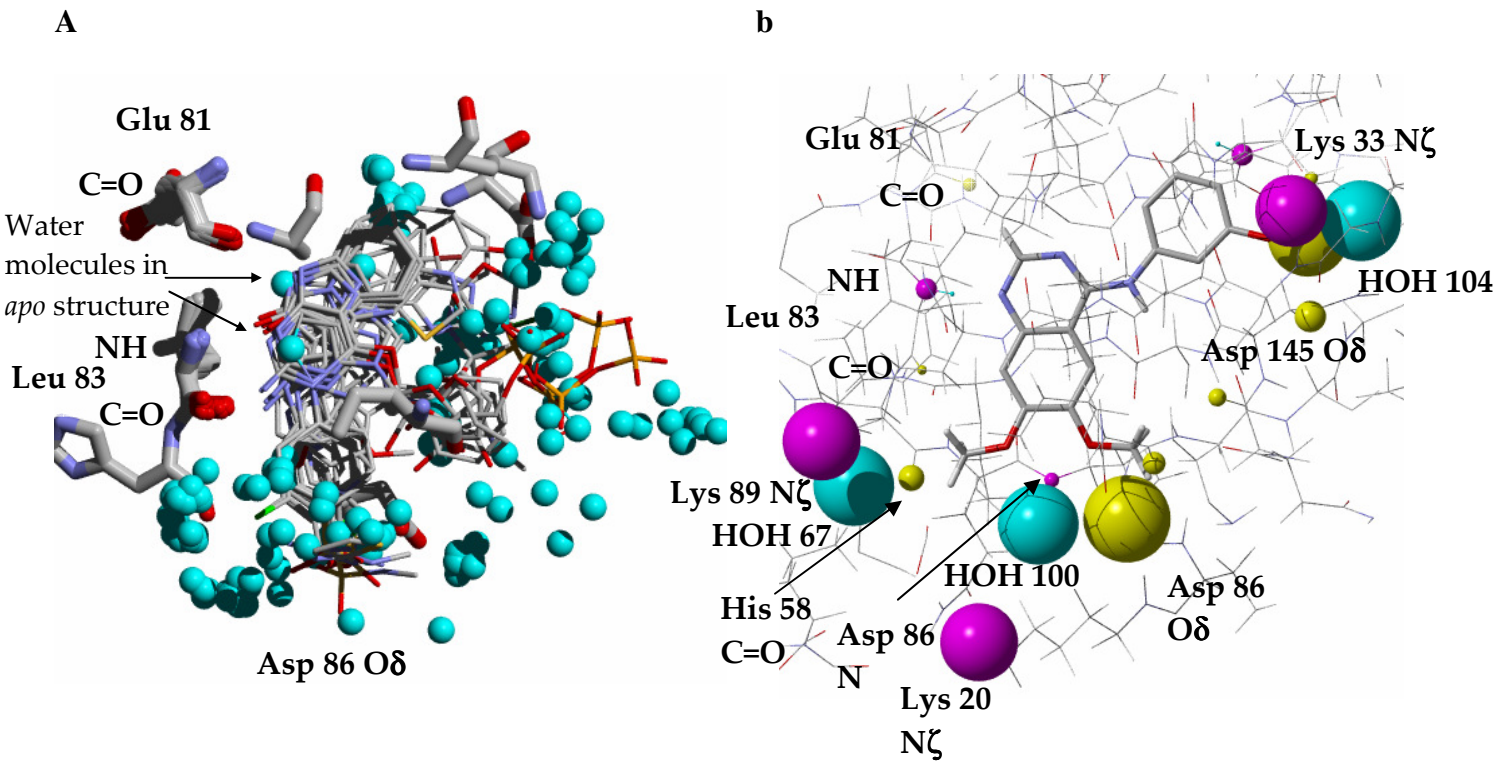

A
Glu 81
C=O
Water
molecules in
apo structure
NH
Leu 83
C=O
Asp 86 Oδ
b
Glu 81
C=O
Lys 33 Nζ
NH
HOH 104
Leu 83
Asp 145 Oδ
C=O
Lys 89 Nζ
HOH 67
HOH 100
Asp 86
Oδ
His 58
C=O
Asp 86
N
Lys 20
Nζ

Figure 2

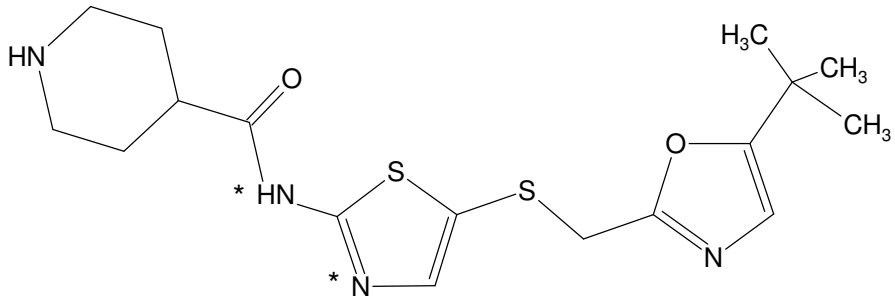
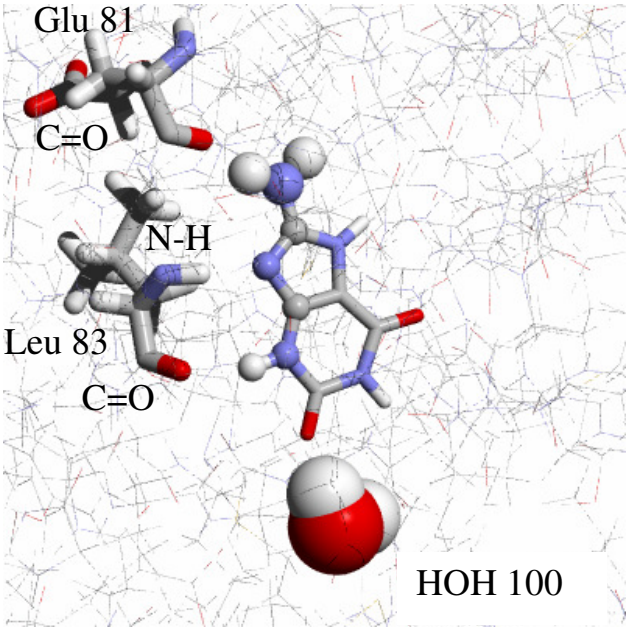
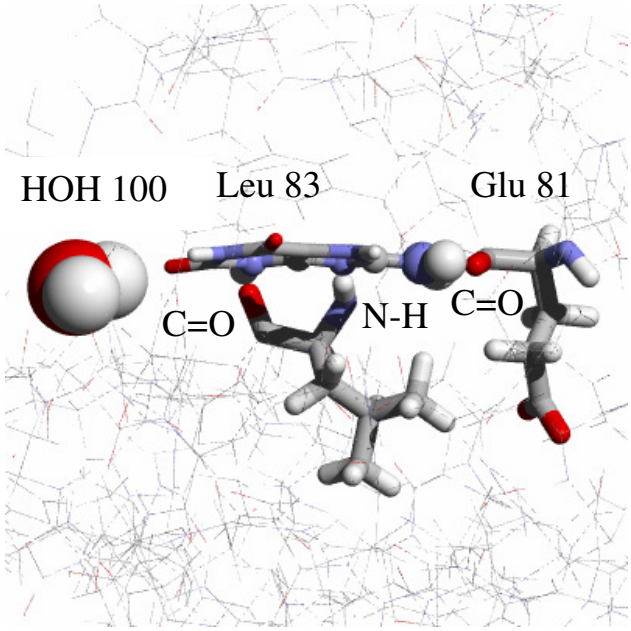


Figure 3

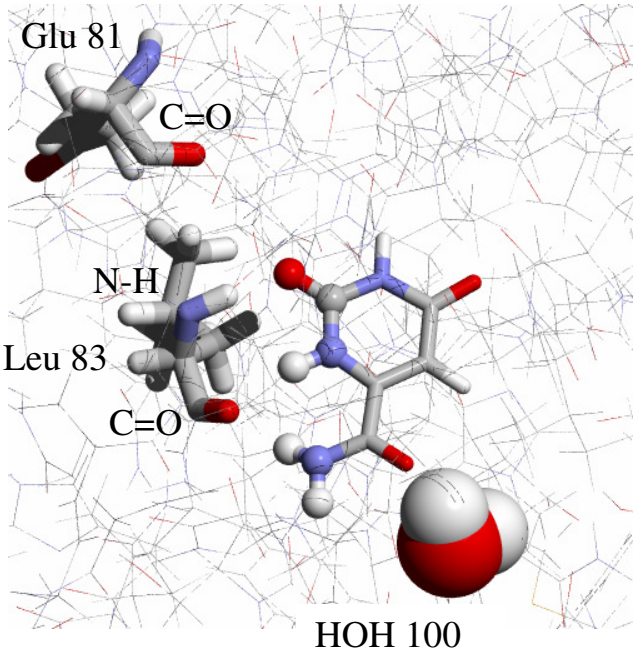
a



b



C



d

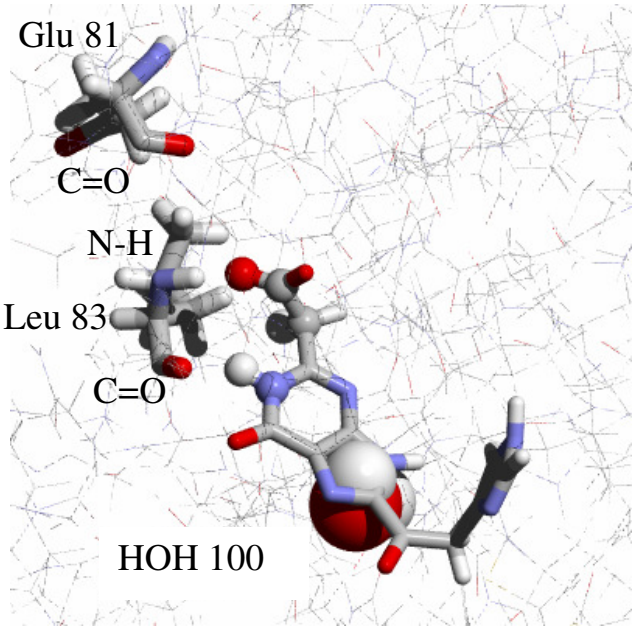
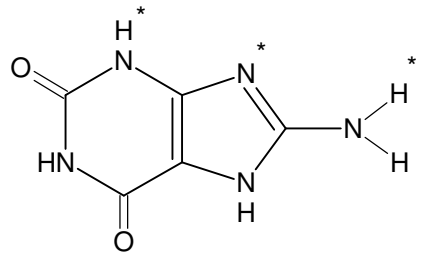
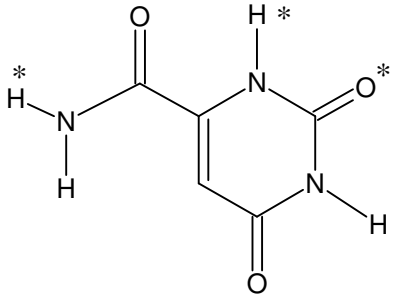


Figure 4

A



b



C