

Including Tightly-Bound Water Molecules in de Novo Drug Design. Exemplification through the in Silico Generation of Poly(ADP-ribose)polymerase Ligands

Alfonso T. García-Sosa,^{†,‡} Stuart Firth-Clark,[‡] and Ricardo L. Mancera^{*,§}

Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1PD, U.K., De Novo Pharmaceuticals, Compass House, Vision Park, Histon, Cambridge CB4 9ZR, U.K., and Western Australian Biomedical Research Institute, School of Biomedical Sciences and School of Pharmacy, Curtin University of Technology, GPO Box U1987, Perth WA 6865, Australia

Received October 11, 2004

Different strategies for the in silico generation of ligand molecules in the binding site of poly(ADP-ribose)-polymerase (PARP) were studied in order to observe the effect of the targeting and displacement of tightly bound water molecules. Several molecular scaffolds were identified as having better interactions in the binding site when targeting one or two tightly bound water molecules in the NAD binding site. Energy calculations were conducted in order to assess the ligand–protein and ligand–water–protein interactions of different functional groups of the generated ligands. These calculations were used to evaluate the energetic consequences of the presence of tightly bound water molecules and to identify those that contribute favorably to the binding of ligands.

INTRODUCTION

Water plays a crucial role in determining the structure and dynamics of biomolecules, and, in particular, it can also have a major impact on the binding of ligands to their receptor proteins. Water molecules can be observed experimentally in biomolecular structures, such as those seen in the crystal structures of proteins. Although some water molecules that appear in X-ray determinations may be an artifact¹ and others may be loosely bound to the surface of the protein, a few water molecules can be said to be tightly bound to the protein surface, as revealed by their crystallographic order and the number and/or strength of their interactions with the protein.^{2,3} Most structure-based drug design and ligand–protein docking applications usually begin by stripping off all water molecules from the binding site of a target protein. When tightly bound water molecules are present, this approach may not be realistic as such water molecules provide hydrogen bonding groups that can mediate the interactions between the ligand and the protein. The ligand–water–protein hydrogen-bonding network that may be formed can help to stabilize the ligand–protein interaction and have a significant effect on the binding mode^{4–6} and even on the chemical diversity of molecules binding to a target active site.

The tightly bound water molecules seen in the binding sites of proteins have been mimicked or targeted in an increasing number of examples in the drug design literature.^{7–9} These studies reveal that displacing a tightly bound water molecule by a functional group in a ligand may improve the binding affinity, although this does not seem to be a general

rule.¹⁰ Natural substrates¹¹ and designed inhibitors¹² have been seen to make use of existing tightly bound water molecules to “bridge” their interactions with the protein. Ligand–protein docking¹³ and virtual screening of organic compounds^{14,15} have been reported to improve in the presence of tightly bound water molecules. The failure to include and target a (noncatalytic) crystallographic water molecule in the virtual screening of potential inhibitors of human carbonic anhydrase resulted in nonoptimal energy scores.¹⁶ This water molecule was later discovered to bridge the interaction in the crystal structure of an inhibitor bound to the protein.¹⁶ Water molecules have also helped to distinguish the binding of different chemical scaffolds,¹⁵ to improve the predictive ability of three-dimensional QSAR models,¹⁷ and to enhance the structural interpretation of ligand-derived pharmacophore models of the binding sites of proteins.¹⁸

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.¹⁹ The complete removal of all water molecules led to difficulties when generating any potential ligands, due to the fact that removing all tightly bound water molecules left unsatisfied hydrogen-bonding groups on the protein surface beyond physical reach for a ligand to satisfy. It became easier to generate ligands when more of the tightly bound water molecules that satisfied those hydrogen-bonding groups were placed back in the binding site. These ligands were also observed to be more chemically diverse. It was proposed that tightly bound water molecules may be in some cases more accessible for hydrogen bonding to an incoming ligand than the actual protein hydrogen-bonding groups associated with them. Water molecules may thus act as versatile hydrogen-bonding groups and reduce the conformational constraints of a particular binding site.

* Corresponding author e-mail: R.Mancera@wabri.org.au.

[†] University of Cambridge.

[‡] De Novo Pharmaceuticals.

[§] Curtin University of Technology.

[#] Current address: Computer-Aided Molecular Design Laboratory, Guggenheim 711, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905.

Drug discovery and crystallographic studies have shown that tightly bound water molecules can be displaced by an appropriate ligand atom but that this is not always necessary. For example, a series of diethylamine HIV-1 protease inhibitors have been reported in which the crystallographic water molecule that acts as an acceptor for two hydrogen bonds (from backbone donors of Ile A50 and Ile B50) and as a hydrogen-bond donor to the ligand was retained. However, a series of peptidomimetics containing a cyclic urea scaffold with a carbonyl group displace this same water molecule, interacting with both isoleucines.⁸ The targeting and displacement of water molecules is probably determined by the free energy changes arising from the favorable entropy gain resulting from displacement of the tightly bound water and subsequent transfer to the bulk and the enthalpy contributions from the hydrogen bonding interactions with the protein and/or ligand and other water molecules that the water molecule had. Therefore, the decision to remove or to keep crystallographically determined water molecules in the binding site of a ligand is not straightforward, as there are examples of both the successful targeting and displacement of these water molecules by active ligands.

Poly(ADP-ribose)polymerase (PARP) is a dimeric enzyme that helps in the repair of breaks in DNA strands by catalyzing the formation of poly(ADP-ribose). The enzyme is activated when cytotoxic agents, such as ionizing radiation and alkylating agents, cause DNA strands to break. Its catalytic domain contains a NAD binding site.²⁰ Resistance to anticancer therapy involves tumor cells becoming able to recognize and repair DNA damage through the action of PARP. Inhibitors of this enzyme may therefore enhance the effects of DNA-damaging anticancer therapy by impeding these repair processes. This makes PARP an attractive target for drug design in the context of cancer treatment. PARP inhibitors have also been suggested for the treatment of stroke.²¹

The analysis of several crystal structures of PARP and the concomitant use of a ligand-derived binding site model of this protein identified a number of tightly bound water molecules that can mediate interactions between a ligand and the protein.¹⁸ Tricyclic inhibitors of PARP had also been previously reported to make interactions with a water molecule in the binding site.²² In this paper we report our analysis of the effect of either displacing or targeting some of these tightly bound water molecules on the prediction of ligand–protein binding and the *in silico de novo* ligand generation. We have compared the docked binding modes of known ligands for PARP and analyzed their possible interactions with two tightly bound water molecules found in the ligand binding site. In addition, we have also classified molecular scaffolds generated by *in silico de novo* ligand generation based on their protein-binding features and analyzed the energetic effect of displacing the tightly bound water molecules by different chemical functional groups within ligands.

MATERIALS AND METHODS

The PDB crystal structure 1EFY²³ of PARP contains two water molecules (HOH 52 and HOH 107) in the NAD binding site that interact directly with the cocrystallized benzimidazole ligand; these water molecules have been

classified as being tightly bound to the protein surface,¹⁸ according to the WaterScore method.²⁴

A number of PARP inhibitors reported in the World Drug Index²⁵ were docked into the binding site of PARP (using PDB crystal structure 2PAX, which contains the largest crystallized ligand) with the docking program GOLD.²⁶ The tightly bound water molecules of 1EFY were then subsequently placed into the binding site of the 2PAX structure to investigate the possible interactions of the docked inhibitors with these water molecules. The final optimal binding modes were selected on the basis of their similarity to the binding mode observed in the crystal structures of PARP with bound inhibitors (1EFY, 1PAX, 2PAX, 3PAX, and 4PAX).

Computational *de novo* ligand generation was carried out using the program Skelgen.^{27–29} This program can construct or modify a ligand incrementally within a ligand binding site using a Monte Carlo simulated annealing optimization algorithm. The program makes use of common ring and acyclic fragments (derived from the World Drug Index) to assemble a molecular structure through a set of chemical rules. The changes carried out during the simulated annealing procedure to incrementally modify a molecular structure are fragment additions, deletions, and mutations as well as molecular translations, rotations, and torsional conformational changes. These changes are carried out stochastically in order to gradually optimize the interaction features and chemical properties of the generated ligands. In addition, hydrogen bond, aromatic, and/or steric geometric constraints can be imposed in order to ensure that ligands satisfy interactions with specific groups in the protein. A full description of this program can be found elsewhere.^{27–29} For each one of the ligand-generation strategies undertaken, a total of 200 runs were carried out to generate an ideal number of 200 ligands. Any failure of Skelgen to successfully generate a ligand can be attributed to difficulties in assembling a suitable molecular scaffold that can satisfy all steric and hydrogen-bonding constraints imposed on the system.

All energy minimizations were carried out using the Discover 3 module in InsightII³⁰ and the CFF force field.³¹ The planarity of aromatic or conjugated systems in some ligands was enforced with additional torsional or out-of-plane restraints. The protein structure was kept rigid in its original crystal structure geometry; however, the hydrogen atoms of all protein amino acids with at least one atom within 3.5 Å of any ligand atom were allowed to reorient during the minimizations in order to optimize the hydrogen-bonding network between the ligand, the water molecule(s) (if present), and the protein. The ligands were allowed full flexibility, while water molecules could only reorient their hydrogen atoms while retaining their original crystallographic position. Care was taken to ensure that the original binding modes of the ligands were retained during the energy minimization by using initial hydrogen-bonding constraints and reduced van der Waals interactions, as described elsewhere.¹⁹ Energy minimizations were stopped when the energy gradient reaches a value of less than 0.01 kcal mol⁻¹ Å⁻¹.

The ligands generated by Skelgen were filtered to remove all those molecules that did not contain a functional amide group interacting with Ser 904 O γ and an aromatic ring. Most active PARP ligands contain these chemical functionalities,

which are considered to be essential for a strong interaction with the protein.^{32–34} This measure was taken to ensure that any generated ligand belonged to one of the known classes of PARP inhibitors, so that the effect of introducing tightly bound water molecules can be assessed with respect to molecules of known activity. The remaining ligands were subsequently clustered based on the type of chemical groups that the amide group is embedded in and the different aromatic ring types found within the ligands. The interaction energies of all the ligands were also calculated and ranked using the Goldscore scoring function implemented in the program GOLD,²⁶ including the receptor water molecules (whose orientation was optimized by energy minimization as described above) if they were present in the structure generation strategy. GOLD scores were obtained after docking the ligand in question and retrieving the Goldscore of the ligand pose with the lowest binding energy. In the case of the known inhibitors, no water molecules were included during the docking optimizations, and the best binding modes were readily identified according to their similarity to the binding motif of crystallized ligands.

For each ligand, the functional group that displaced a tightly bound water molecule in order to interact with hydrogen-bonding groups on the protein was “mutated” into a hydrogen atom. If two functional groups had displaced two water molecules, then the two functional groups were mutated both together and separately. This mutation was done in order to calculate the contribution to the energy of ligand–protein binding that each different functional group had. Functional groups within rings were converted into a $-\text{CH}_2-$ leaving the cyclic structure intact. Single-point potential energy calculations were thus carried out to compare the “normal” and “mutant” ligand–protein complexes. The transformations were carried out using the program Insight II,³⁰ and the energy calculations were done using the CFF force field,³¹ as before. In addition, single-point energy calculations were also carried out after removing the water molecule(s), if any was present during the *in silico* ligand generation procedure. This was done in order to calculate the contribution to the energy of ligand–protein binding of each of the tightly bound water molecules. These calculations do not include any corrections for the entropy change, but it is not clear whether the displacement of a tightly bound water molecule always results in a gain in entropy⁷ since there are reports that this entropy change may vary greatly.³⁵

RESULTS AND DISCUSSION

The catalytic site of PARP is a narrow flat cleft where the hydrogen bonds that dictate specificity are formed on a narrow region on one side of the cleft. Tyr 907, located on one long side of the cavity, guarantees that the site remains closed with planar ligands best suited for forming favorable aromatic interactions inside the binding site. Ligands usually exhibit the binding mode shown in Figure 1, with a *cis* amide group interacting via its NH_2 or NH with Gly 863 $\text{C}=\text{O}$ and via its $\text{C}=\text{O}$ with both Gly 863 NH and the hydroxyl Ser 904 $\text{O}\gamma$.

Mutation of any of the residues, Tyr 907 (Y907N), Lys 903 (K903Q, K903E), and Glu 988 with Lys 903 (K903E/E988K), results in the nearly complete loss of activity.³⁶ The catalytic activity of Glu 988 consists of activating an

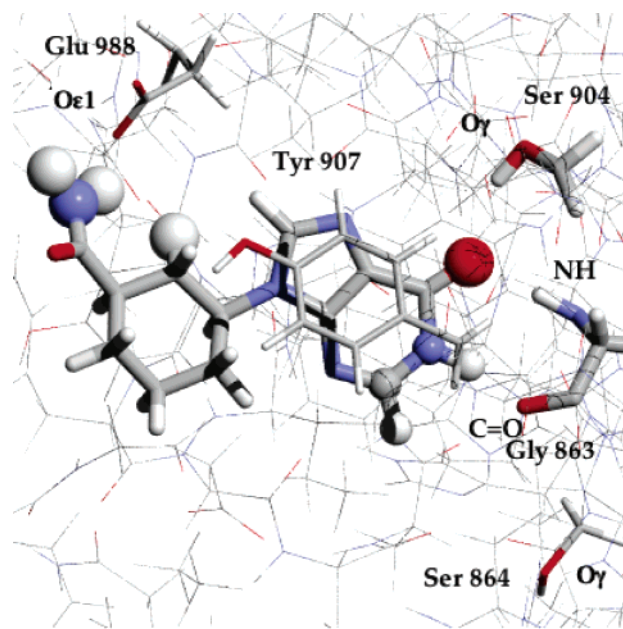


Figure 1. Main hydrogen-bonding groups in the nicotinamide binding site of PARP. Small spheres indicate atoms within hydrogen-bonding distance (2.5–3.5 Å) to Gly 863, and medium spheres are atoms within hydrogen-bonding distance to Ser 904 or Glu 988.

attacking hydroxyl group on the nucleophile poly(ADP-ribose) attacking NAD and simultaneously stabilizing the transition state.

The crystal structures of PARP with bound inhibitors reveal the presence of several water molecules. There are two tightly bound water molecules in the 1EFY crystal structure: HOH 52 and HOH 107.¹⁸ These water molecules are seen to mediate the interactions of various ligands with the protein.^{18,22} Furthermore, the positions of these water molecules in the binding site of PARP match projection points from a ligand-derived pharmacophore model of the binding site.¹⁸ Consequently, these water molecules can be considered as part of the ligand–protein interactions of known inhibitors and should therefore be considered in a ligand design strategy. The position of these tightly bound water molecules in the binding site of PARP can be seen in Figure 2.

The chemical structures of the inhibitors that were docked into the binding site of PARP can be found in Table 1. A superposition of their optimal binding modes with respect to the binding site of PARP can be seen in Figure 3, where the tightly bound water molecules have been included for reference. As expected, all ligands make hydrogen bonds with Gly 863 $\text{C}=\text{O}$, Gly 863 NH , and Ser 904 $\text{O}\gamma$.

The ligand found in the 2PAX structure, 4-amino-1,8-naphthalimide, has an amine group that displaces HOH 52 and replaces its hydrogen-bonding interactions with the protein. Recently reported inhibitors were designed to displace HOH 52 in order to interact with the carboxylate group of Glu 988.³⁷ Inhibitor 4 displaces HOH 107 but does so with a benzyl ring, i.e., without replacing the hydrogen-bonding interactions of HOH 107. Interestingly, none of the inhibitors studied was able to displace both HOH 52 and HOH 107.

Other inhibitors do not displace these water molecules. For example, HOH 52 is retained and targeted by chemical

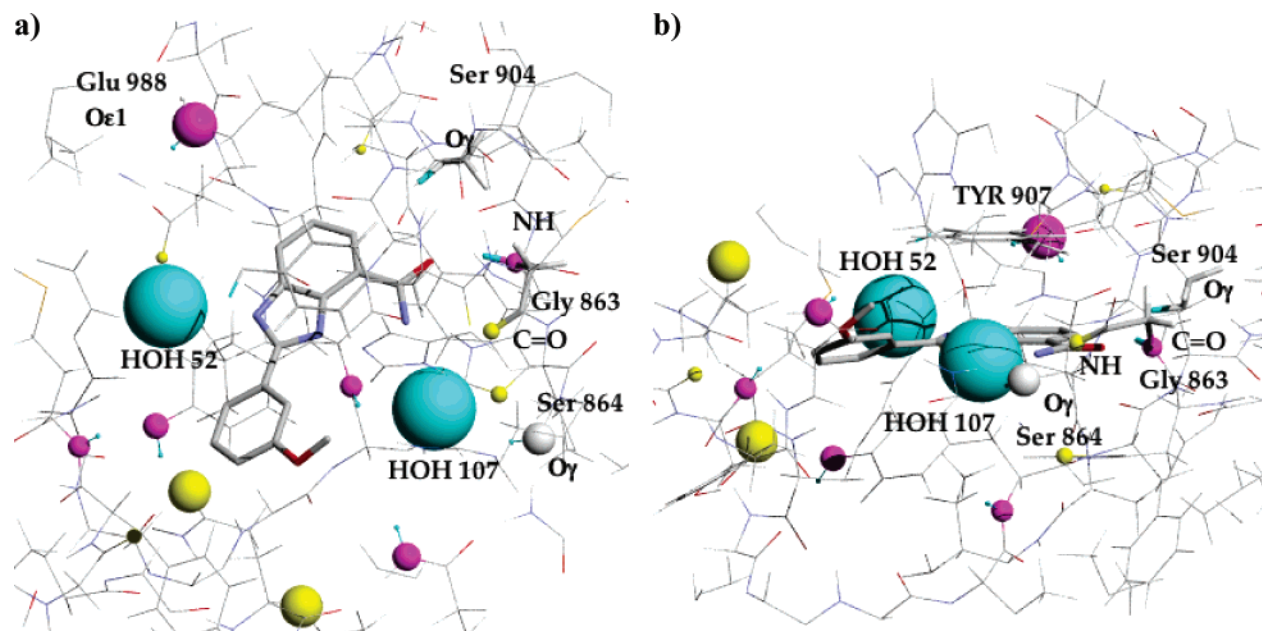
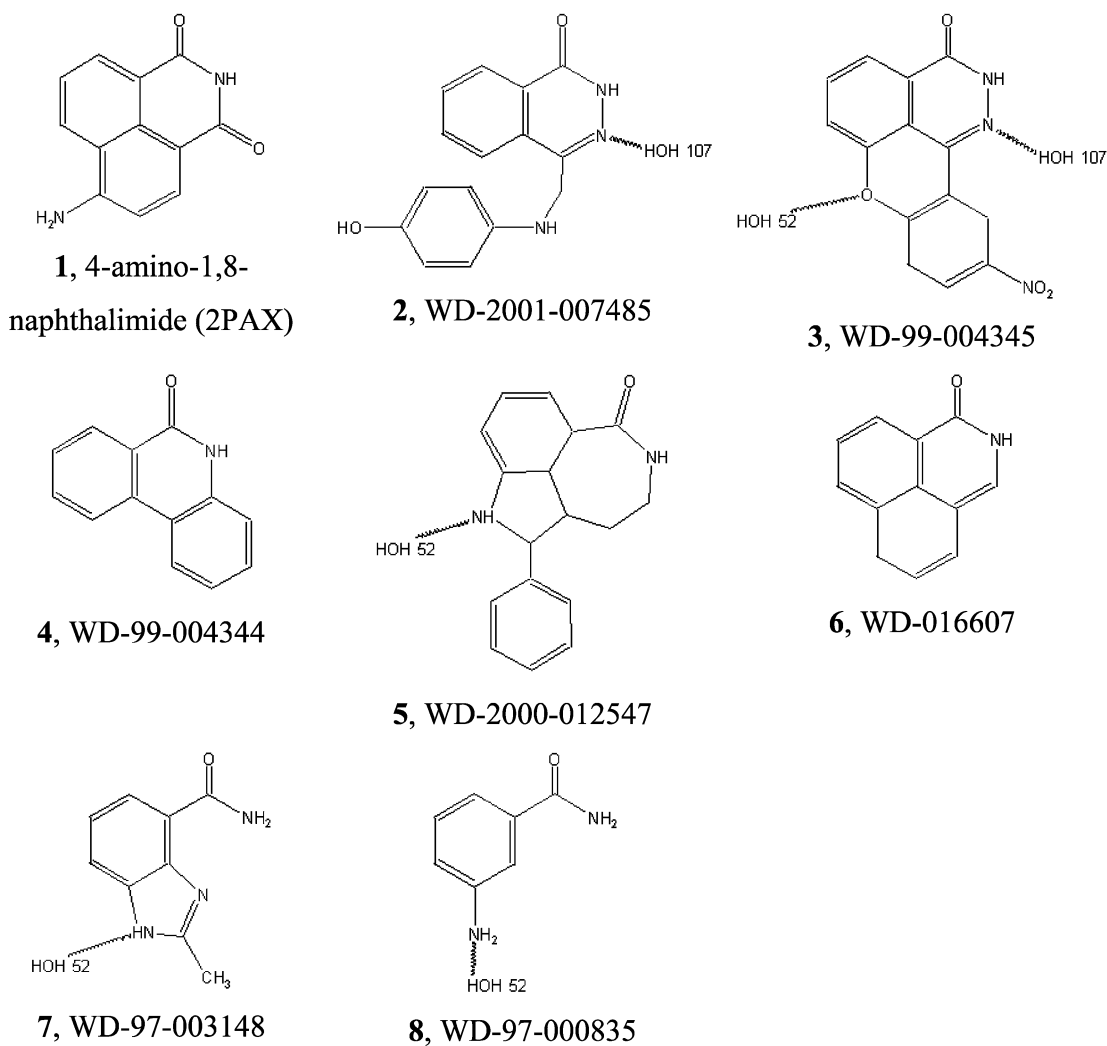


Figure 2. (a) The (nicotinamide) binding site of PARP with inhibitor **4** (sticks) and tightly bound water molecules (cyan spheres). (b) Side-view. Hydrogen-bonding groups (site-points) are highlighted as yellow spheres for hydrogen-bond acceptors, as purple spheres for hydrogen-bond donors, and as white spheres for amphiprotic groups.

Table 1. Representative Inhibitors of PARP and Their Schematic Interactions with the Two Tightly Bound Water Molecules Found in the Binding Site



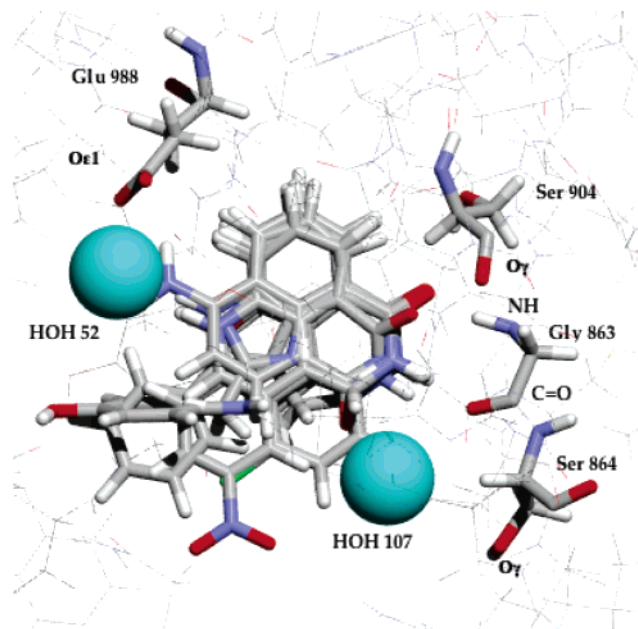


Figure 3. Superposition of docked inhibitors in the binding site of PARP. Tightly bound water molecules are shown as cyan spheres.

groups in inhibitors **3**, **5**, **7**, and **8**. HOH 107 also bridges the interactions of inhibitors **2** and **3** with the protein. Inhibitor **6** is different from the other ligands: it neither displaces nor interacts with any of the water molecules. We can generally conclude that known inhibitors of PARP can either displace or target tightly bound water molecules, and these two possibilities should be considered in any ligand design strategy.

de Novo Ligand Generation. The above description of the PARP active site clearly indicates that any ligand should have compulsory hydrogen-bonding interactions with three protein groups: Gly 863 C=O, Gly 863 NH, and Ser 904 O γ . We designed nine separate ligand generation strategies by using appropriate hydrogen-bonding constraints to impose interactions with these hydrogen-bonding groups, in addition to a combination of other possible hydrogen-bonding groups in the protein and water molecules HOH 52 and HOH 107 (which could act as either hydrogen-bond donors or acceptors). Consequently, these strategies involve the displacement and/or targeting of these water molecules. The former was achieved by additionally including those hydrogen-bonding groups that these water molecules either interact with or block access to. HOH 52 blocks access to Glu 988 O ϵ 1 and HOH 107 interacts with Ser 864 O γ . The ligand design strategies that we used are outlined in Table 2.

The core molecular scaffolds identified for each cluster of designed ligands can be seen in Table 3. These scaffolds are drawn schematically, and they all consist of heterocyclic aromatic five- or six-membered rings with multiple substitutions.

Table 2 also indicates which molecular scaffolds were observed in the output from each ligand design strategy. We now proceed to describe in detail the output of the *in silico* ligand generation with the various above-described strategies.

Strategy 1. Only the three compulsory hydrogen-bonding site-points were considered here. Due to the optimal location

of these site-points in the binding site of PARP, Skelgen was able to generate ligands quickly and effectively. Most ligands tended to be small, as was expected by the close proximity of the three site-points, which can be satisfied by small chemical fragments. The ligands contained mainly molecular scaffolds A, C, and D (see Table 3 for their chemical structures).

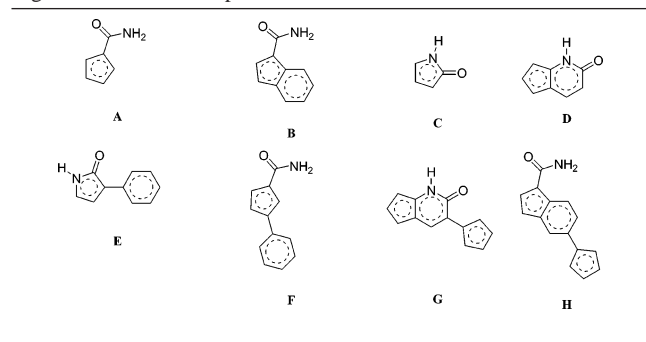
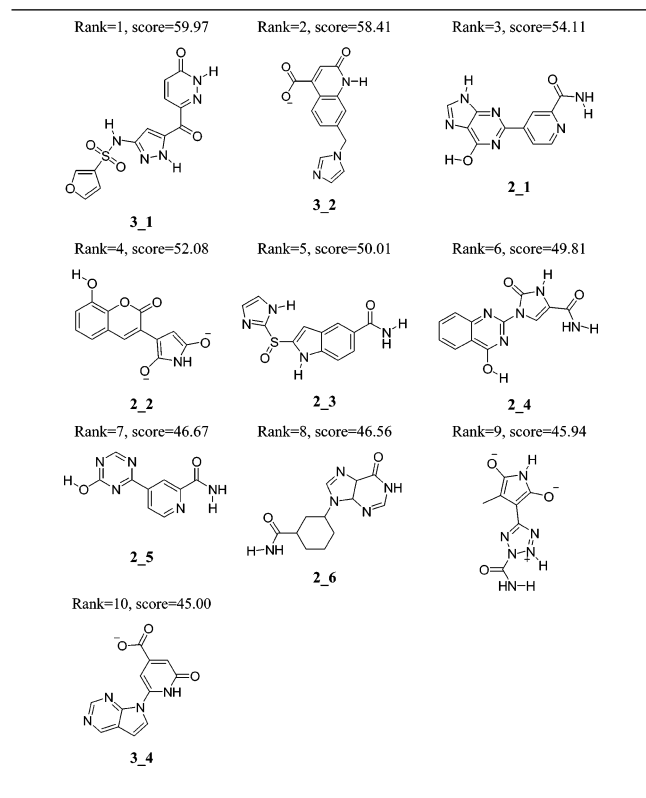
Strategy 2. The three compulsory hydrogen-bonding site-points plus Glu 988 O ϵ 1 were used here. Glu 988 O ϵ 1 lies opposite the three compulsory site-points (on the left in Figure 2(a) and toward the back of the binding site in Figure 2(b)). The presence of this additional site-point was expected to result in significantly larger ligands (and their associated molecular scaffolds) with a wider chemical diversity. The ligands generated have indeed a rich chemical diversity, as revealed by the fact that ligands containing all molecular scaffolds (except C) were present. Ligands with molecular scaffold C may not have been found due simply to insufficient sampling. Molecular scaffold E is the most prevalent one, which suggests that it can be used effectively to target the three compulsory site-points on one side of the binding site and the additional site-point on the other side. The results with this strategy also confirm that adequate sampling of the fragment set was being carried out by Skelgen, as most molecular scaffolds seen in PARP inhibitors were found by this *in silico* approach.

Strategy 3. The three compulsory hydrogen-bonding site-points plus HOH 52 were used here. HOH 52 has its hydration site in front of Glu 988 O ϵ 1, thus blocking access to this site-point. The ligands generated tended to be relatively small, predominantly with molecular scaffolds A and C. Some medium-sized ligands were also generated, having molecular scaffolds D and E. The interactions of the ligands with HOH 52 were achieved by appropriate substituents on the rings of the scaffolds. In comparison with strategy 2 above, the ligands generated with strategy 3 had a higher proportion of large and/or extended aromatic ring systems. Ligands with molecular scaffolds B, D, and E were generated both in the absence (strategy 2) or presence (strategy 3) of HOH 52, suggesting that these scaffolds are rather versatile and appropriate substitutions can be made to design ligands whose binding could either displace or target HOH 52. Top-ranking ligands generated with strategies 2 and 3 can be found in Table 4, where they are ranked according to their GOLD score. The molecules in Table 4 have been depicted in approximately the same orientation as that observed for the ligand in Figure 1. The ranking tells us that both strategies 2 and 3 (that is, either displacing or targeting HOH 52, respectively) potentially result in ligands that score equally well.

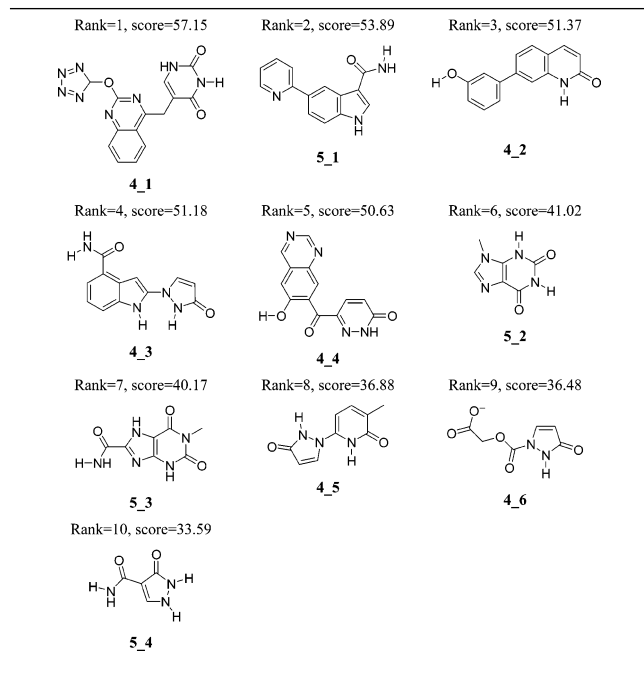
Strategy 4. The three compulsory hydrogen-bonding site-points plus Ser 864 O γ were used here. The hydroxyl of Ser 864 O γ is located near the backbone C=O group of the compulsory site-point Gly 863 (see Figure 2). Most of the ligands that were generated contained molecular scaffold E, with only a couple of ligands containing scaffold C and one ligand with scaffold D. If we compare this strategy with strategy 1, we can see that the addition of Ser 864 O γ resulted in somewhat larger ligands but with essentially only one associated molecular scaffold type.

Table 2. Site-Points Used and Molecular Scaffolds Found in the Different Ligand Generation Strategies

run	site-point groups	predominant scaffold(s)	strong presence of scaffold(s)	comments
1	3 compulsory (Gly863 O + Gly 863 N + Ser 904 O γ)	A	C, D	only small ligands found
2	3 compulsory + Glu 988 O ϵ 1	E	A, B, D, F, G, H	very diverse, most scaffolds found
3	3 compulsory + HOH 52	A, C	B, D, E	moderately diverse
4	3 compulsory + Ser 864 O γ	E		only a few large ligands in one cluster found
5	3 compulsory + HOH 107	C	A, D	mostly small ligands
6	3 compulsory + Glu 988 O ϵ 1 + Ser 864 O γ	G		few large ligands
7	3 compulsory + HOH 52 + HOH 107	C, E	A, D	many small ligands
8	3 compulsory + Glu 988 O ϵ 1 + HOH 107	C, E		few ligands
9	3 compulsory + HOH 52 + Ser 864 O γ	D, E	C, G	moderately diverse

Table 3. Core Molecular Scaffolds in Each Cluster of Ligand-Generation Output**Table 4.** GOLD Ranking of Ligands Generated with Strategies 2 and 3

Strategy 5. The three compulsory hydrogen-bonding site-points plus HOH 107 were used here. HOH 107 has its hydration site in front of Ser 864 O γ , thus blocking access to this site-point. It was easier to generate ligands with this strategy compared with strategy 4, probably due to less restrictive geometric constraints since HOH 107 is located in the immediate vicinity of Gly 863. Molecular scaffold C was found predominantly in the generated ligands, followed

Table 5. GOLD Ranking of Ligands Generated with Strategies 4 and 5

by scaffolds A and D. Most ligands tended to be small molecules. The only common scaffold found between strategy 5 and strategy 1 was scaffold C. Top-ranking ligands generated with strategies 4 and 5 can be found in Table 5, where they are ranked according to their GOLD score. In analogy to strategies 2 and 3, this score was calculated in the presence of HOH 107.

As was the case in the comparison in Table 4, both strategies produce ligands that score well, either displacing HOH 107 in strategy 5 or targeting HOH 107 in strategy 4.

Strategy 6. The three compulsory hydrogen-bonding site-points plus Glu 988 O ϵ 1 and Ser 864 O γ were used here. Skelgen was able to generate only relatively few ligands, due to difficulties in generating ligands that could satisfy all site-points. Most ligands contained molecular scaffold G, thus tending to be large molecules with extended and fused aromatic rings. It is worth noting that the first fused ring in molecular scaffold G, containing the core amide functional group that interacts with the three compulsory site-points, is the same one found in scaffold D (which contains two fused rings). Ligands with molecular scaffolds G and D differ only in the number of rings and the way in which they are linked. Therefore, it would seem that the most likely way to target both additional site-points is through appropriate substitutions on an additional phenyl (or other rings) on

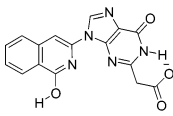
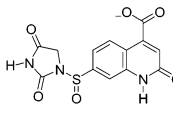
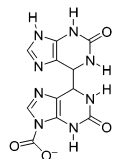
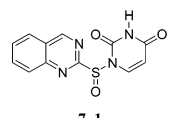
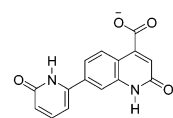
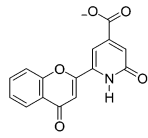
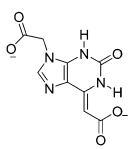
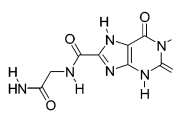
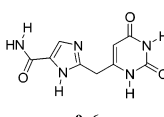
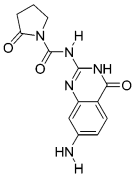
scaffold D (turning it into scaffold G). A further observation here is that the removal of all water molecules in the binding site of PARP does not seem to be a good strategy as only very large ligands seem to be able to interact with all site-points.

Strategy 7. The three compulsory hydrogen-bonding site-points plus HOH 52 and HOH 107 were used here. The inclusion of both water molecules reduced the distances between the compulsory site-points and the water site-points, and, as a consequence, the ligands generated tended to be of intermediate size. Molecular scaffolds C and E were predominant in the generated ligands, although there were also several ligands with scaffolds A and D. A comparison of this strategy, which includes both water molecules, with strategy 3 (which only includes HOH 52) and strategy 5 (which only includes HOH 107) reveals that the ligands generated by all three strategies share molecular scaffolds A, C, and D. On the other hand, a comparison with strategies 8 (which includes HOH 107 and Glu 988 O ϵ 1) and 9 (which includes HOH 52 and Ser 864 O γ), which are explained below, reveals that the ligands generated by all three strategies share molecular scaffolds C and E. These observations suggest that ligands that target the additional site-points (Glu 988 O ϵ 1 and Ser 864 O γ) will tend to have a reduced chemical diversity in their underlying molecular scaffolds. The chemical diversity increases when either of the water molecules replaces these site-points.

Strategy 8. The three compulsory hydrogen-bonding site-points plus Glu 988 O ϵ 1 and HOH 107 were used here. Most of the ligands generated have molecular scaffolds C or E, with only one ligand having scaffold D. If a comparison is made between these ligands and those obtained with strategy 6 (which uses all five hydrogen-bonding site-points and no water molecules), one can see that there are no molecular scaffolds in common (only scaffold G could be found in the ligands generated with strategy 6). As mentioned above when comparing strategies 4 and 5, the inclusion of HOH 107 (which is more accessible than Ser 864 O γ) reduces the geometric hydrogen-bonding constraints. In addition, its presence not only facilitated ligand generation but also increased the chemical diversity of the underlying molecular scaffolds. However, the chemical diversity seen in ligands generated with this strategy is significantly reduced in comparison with that of ligands generated with strategy 2 (where the same four site-points were used but no water molecule was present). One of the inevitable conclusions here is that ligands that are able to displace HOH 52 (strategy 2) will exhibit the widest possible chemical diversity in their underlying molecular scaffolds; however, the concomitant targeting of HOH 107 has a deleterious effect on this chemical diversity (strategy 8). The alternative option, which is the concomitant displacement of HOH 107 (strategy 6), reduces the chemical diversity of the underlying molecular scaffolds further. An obvious question arises here: is it then energetically advantageous to include HOH 107 and, indeed, HOH 52? We try to address this question with the calculations of water molecule and functional group contributions to the binding energy presented below.

Strategy 9. The three compulsory hydrogen-bonding site-points plus Ser 864 O γ and HOH 52 were used here. The ligands generated were of medium size and made use predominantly of molecular scaffolds D and E, although a

Table 6. GOLD Ranking of Ligands Generated with Strategies 6–9

Rank=1, score=61.09  6_1	Rank=2, score=59.20  9_1	Rank=3, score=56.96  9_2
Rank=4, score=55.51  7_1	Rank=5, score=53.44  9_3	Rank=6, score=53.32  9_4
Rank=7, score=52.67  9_5	Rank=8, score=48.53  7_2	Rank=9, score=44.54  9_6
Rank=10, score=42.48  9_7		

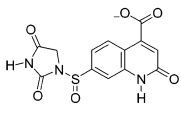
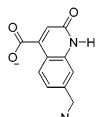
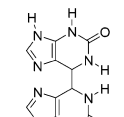
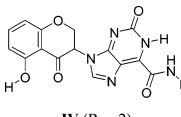
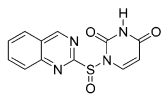
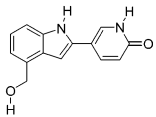
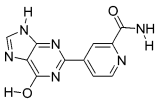
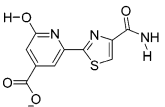
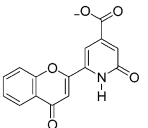
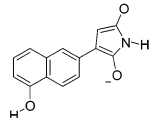
few small ligands have scaffold C and some large-sized ligands have scaffold G.

Table 6 compares the top GOLD scores for ligands generated in the last four strategies. Ligands from strategy 9 (with HOH 52) dominate, plus a couple of ligands from strategy 7 (with both HOH 52 and HOH 107). It should be noticed that these strategies produced the best scoring ligands, compared to ligands produced with strategies that displaced both water molecules.

Ligands containing molecular scaffolds F and H and, to some extent, scaffold B, were difficult to generate, and only strategy 2 was successful in generating ligands with all these scaffolds (strategy 3 also generated some ligands with scaffold B). At least in the case of PARP, this suggests that having a fairly large binding site with a few site-points allows for various scaffolds to be used successfully.

Ligands containing the related molecular scaffolds D or G could be generated in nearly all of the runs. This suggests that these fused-ring scaffolds (which incorporate the core amide group within one of the rings) are versatile enough to be able to target nearly all binding environments (with or without water molecules) with appropriate substitutions. The same can also be said about ligands containing the related molecular scaffolds C and E (which contain the core amide group within their single ring).

Table 7. Selection of Ligands Ranked within the Top 10% of All Strategies Combined using Goldscore

Rank=1, score=59.2  I (Run 9)	Rank=2, score=58.41  II (Run 3)	Rank=3, score=56.96  III (Run 9)
Rank=4, score=56.55  IV (Run 2)	Rank=5, score=55.51  V (Run 7)	Rank=6, score=54.37  VI (Run 2)
Rank=7, score=54.11  VII (Run 2)	Rank=8, score=53.92  VIII (Run 2)	Rank=9, score=53.32  IX (Run 9)
Rank=10, score=52.28  X (Run 8)		

Ligand-Protein Binding Energies. All ligand-protein complexes were minimized and subsequently scored using GOLD. A selection of the ligands ranked within the top 10% on the basis of their Goldscore binding energies are shown in Table 7.

Most of the ligands were generated using strategies 2 and 9. The bottom 10% ranked ligands were generated predominantly using strategies 1 and 5 (results not shown), which tended to generate small ligands. This probably reflects a common deficiency of scoring functions, whereby their additive nature results in low scores for small ligands and usually high scores for large ligands such as those generated in strategy 9. However, there are several ligands that are not large in size but that were ranked in the top 10%. Ligands I, II, and V to X are medium-sized molecules that were ranked in the top 10%. It is interesting to note that some strategies resulted in the generation of ligands in both the top 10% and bottom 10% of the rank list. This is the case of a few of the ligands generated using strategies 3, 5, and 7. Ligands generated with these strategies were usually ranked in the bottom 10%, possibly due to the prevalence of small-sized ligands, such as those of clusters A and C. Interestingly, half of the ligands ranked in both the top and bottom 10% were generated with strategies that included at least one water molecule. This indicates that the targeting or displacement of water molecules does not intrinsically improve the scoring of the ligand-protein interactions.

Table 8. Representative Energy Changes (in kcal mol⁻¹) upon Replacement (with a Hydrogen Atom) of Functional Groups that Had Displaced Tightly Bound Water Molecules

functional group	OH	NH in aromatic ring	NH in amide	C=O in carboxylate	C=O in amide
$\Delta\Delta E_{fg}^b$	10.44	27.02	6.83	26.12	12.12
	11.16	16.85	11.44	28.06	12.84
	11.30	3.57		29.04	
				33.97	
				9.26	
				1.12	
average	10.51	15.81	9.14	29.30	12.43

There is further evidence that strategy 1 (which only made use of the three compulsory site-points) is not a suitable one: none of the ligands generated are ranked in the top 10%. On the other hand, strategy 7 (which made use of both HOH 52 and HOH 107) resulted in predominantly small- and medium-sized ligands that do not have many interactions with the protein and usually appear in the bottom 10%.

It is also important to note that ligands generated with strategies 2 and 9 were a prominent part of the molecules ranked in the top 10%, despite the fact that they do not have extremely large sites.

Water Displacement and Water Targeting. The molecular mechanics energies of binding of ligands generated in the absence and presence of the water molecules were dissected in order to compute the energy contributions that arise, respectively, from displacing or targeting these water molecules. For the former case, Table 8 shows the representative energy changes associated with removing the interacting functional groups by mutating them into hydrogen atoms ($\Delta\Delta E_{fg}^b$), thereby providing the energy changes associated with their hydrogen-bonding interactions with the protein. This was readily calculated as

$$\Delta\Delta E_{fg}^b = \Delta E^b_{\text{norm}} - \Delta E^b_{\text{mut}} \quad (1)$$

where ΔE^b_{norm} is the binding energy of the ligand-protein complex (calculated by subtracting the energies of the ligand and the protein from the energy of the complex) and ΔE^b_{mut} is the binding energy for the modified ligand (calculated by subtracting the energies of the ligand and the protein from the energy of the modified complex). We should mention that the C-H reference state (upon mutation of a hydrogen-bonding group to hydrogen) is not strictly speaking the same for donors and acceptors due to the small electrostatic energy difference that arises from the small partial charge on the added H atom.

An analysis of the energy changes that accompany the mutations of various functional groups into noninteracting atoms, as shown in Table 8, reveals that different functional groups tend to have specific interaction energies with the protein upon displacement of a tightly bound water molecule. For example, OH groups have an average interaction energy of -11 kcal mol⁻¹ (approximately, the energy of two hydrogen bonds), C=O groups in amides have an average interaction energy of -12 kcal mol⁻¹ (approximately, the energy of two hydrogen bonds), and C=O in carboxylate groups have an average interaction energy of -29 kcal mol⁻¹ (approximately, the energy of four hydrogen bonds). In order for the displacement of a tightly bound water molecule to

Table 9. Representative Energy Changes (in kcal mol⁻¹) upon Removal of Tightly Bound Water Molecules that Bridged the Ligand–Protein Interaction

functional group	NH		N sp2 in ring	NH in ring	C=O in		
	OH	amide			aromatic ring	carboxylate	
$\Delta\Delta E_w^b$	16.24	12.42	15.60	9.78	11.03	9.10	23.42
	18.97	17.50		11.00		14.64	21.02
		19.81		14.24		14.68	
				19.68		13.56	
				18.53			
			13.85				
average	17.60	16.58	15.60	14.51	11.03	12.99	22.22

be energetically favorable overall, the energy of interaction of a functional group of a ligand bound to the protein would have to be more negative than the interaction energy of a tightly bound water molecule with the protein. It has been calculated that a hydrophilic cavity in the interior of a protein has an energy threshold of -12 kcal mol⁻¹³⁸ for finding a water molecule occupying the site. In other words, if a water molecule has interactions in an interior cavity with the protein of at least -12 kcal mol⁻¹ (roughly, two hydrogen bonds), it will occupy that site. Therefore, if a functional group on a ligand can provide the same amount of energy, it is possible that a particular ligand may replace the water molecule in the cavity of the protein. From the data shown in Table 8 we can see that hydroxyl groups may not necessarily provide this amount of energy and may therefore find it difficult to replace a water molecule in a hydrophilic cavity. Certain amines in aromatic rings and predominantly carbonyl (especially carboxyl) groups may be more likely to provide this energy of interaction and therefore displace the water molecule.

In the case of water targeting, Table 9 shows representative energy changes associated with removing the water molecule(s) for each kind of functional group that was found to interact with the water molecules ($\Delta\Delta E_w^b$). This was also readily calculated as

$$\Delta\Delta E_w^b = \Delta E^b_{\text{wat}} - \Delta E^b_{\text{no_wat}} \quad (2)$$

where ΔE^b_{wat} is the binding energy of the ligand–protein complex including water molecules (calculated by subtracting from the energy of the complex the energies of the protein, the ligand and the water molecule(s)), and $\Delta E^b_{\text{no_wat}}$ is the binding energy of the ligand–protein complex without the water molecules (calculated by subtracting from the energy of the complex the energies of the protein and the ligand).

From the data in Table 9 we can see that each water molecule has a different interaction energy in the ligand–water–protein complex depending on the nature of the functional group on the ligand that interacts with the water molecule. If we consider that the ideal energy for a water hydrogen bond is around -6 kcal mol⁻¹,³⁹ we can then conclude that water molecules are able to make one or two hydrogen bonds with the protein, while the energy of the hydrogen bonds made between ligand functional groups and the protein varies considerably between functional groups as well as within each functional group. Consequently, if there is a suitable hydration site available to a water molecule in a ligand–protein complex, the neglect of including such water molecule would be detrimental to the energy of

interaction of such complex. The energy values in Table 9 give a measure of the energy of interaction which is lost by not considering a bridging water molecule, which ranges between 11 and 22 kcal mol⁻¹ depending on the functional group in a ligand targeting the water molecule.

CONCLUSIONS

The binding site of PARP has two tightly bound water molecules in PDB structure 1EFY: HOH 52 and HOH 107. An analysis of the binding mode of various ligands in several crystal structures and the docking of a representative set of other known inhibitors revealed that PARP inhibitors can either displace or target these tightly bound water molecules.

The *in silico* generation of putative PARP ligands revealed that the targeting of HOH 52 was able to enhance the chemical diversity compared to those ligands generated when all water molecules were ignored or displaced. However, the displacement of this water molecule allowed for the generation of ligands with the largest possible chemical diversity. On the other hand, the concomitant targeting of HOH 107 had a deleterious effect by reducing even further the chemical diversity of the ligands generated.

The docking GOLD scores of the top-ranking ligands generated when targeting the water molecules were roughly the same as those scores of ligands generated when displacing the water molecules. These results indicate that although the displacement or targeting of tightly bound water molecules can be beneficial for the enhancement of chemical diversity in the ligand structures generated for PARP, there is no significant difference when comparing the binding energies of ligands computed with docking scoring functions.

Mutation of ligand functional groups provided values that allowed for determining the molecular mechanics energy change associated with the displacement of a water molecule from a hydration site, with carboxyl, carbonyl, and some aromatic amines providing more than -12 kcal mol⁻¹ of energy (equivalent to two hydrogen bonds), the amount of energy that is needed to overcome the interaction of a water molecule in a hydrophilic site. On the other hand, hydroxyl groups may not in general be appropriate functional groups for the displacement of a tightly bound water molecule.

In addition, the energy contribution of water molecules to different ligand–water–protein complexes was calculated, which allowed for estimating the amount of energy that is always neglected when a tightly bound water molecule is not included in the molecular mechanics calculation of the binding energy of a ligand–protein complex.

This work illustrates some of the effects of displacing, neglecting, or targeting of water molecules in *de novo* *in silico* ligand generation methods, showing the influence on the chemical diversity of the generated ligands, the feasibility of a ligand generation strategy, the binding energies calculated with docking scoring functions and molecular mechanics, and the energy changes and patterns associated with the displacement of such water molecules by certain functional groups in ligand–protein complexes. We have observed that the targeting and/or displacement of tightly bound water molecules in an *in silico* *de novo* ligand generation process can have a profound effect on the chemical diversity of the ligands generated. The energetic balance between displacing and targeting a tightly bound water molecule depends on

the functional groups that displace such water molecules and the water–protein and water–ligand energies of interaction when such water molecules are included.

ACKNOWLEDGMENT

A.T.G.S. would like to thank Drs. Per Källblad and Nikolay Todorov for useful discussions and to acknowledge CONACyT (Mexico) for the award of a post-graduate scholarship and Universities UK for the award of an Overseas Research Scheme Award.

REFERENCES AND NOTES

- Davis, A. M.; Teague, S. J.; Kleywegt, G. J. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem., Int. Ed.* **2003**, *42*, 2718–2736.
- Poornima, C. S.; Dean, P. M. Hydration in drug design I. Multiple hydrogen-bonding features of water molecules in mediating protein–ligand interactions. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 500–512.
- Hendlich, M.; Bergner, A.; Günter, J.; Klebe, G. Relibase: Design and development of a database for comprehensive analysis of protein–ligand interactions. *J. Mol. Biol.* **2003**, *326*, 607–620.
- Chung, E.; Henriques, D.; Renzoni, D.; Zvelebil, M.; Bradshaw, J. M.; Waksman, G.; Robinson, C. V.; Ladbury, J. E. Mass spectrometric and thermodynamic studies reveal the role of water molecules in complexes formed between SH2 domains and tyrosyl phosphopeptides. *Struct. Fold. Des.* **1998**, *6*, 1141–1151.
- Rejto, P. A.; Verkhivker, G. M. Mean field analysis of FKBP12 complexes with FK506 and rapamycin: Implications for a role of crystallographic water molecules in molecular recognition and specificity. *Proteins: Struct., Funct., Genet.* **1997**, *28*, 313–324.
- Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dijols, S.; Dansette, P. M.; Mansuy, D.; Stout, C. D. Structure of mammalian cytochrome P450C5 complexed with diclofenac at 2.1 angstrom resolution: Evidence for an induced fit model of substrate binding. *Biochemistry* **2003**, *42*, 9335–9345.
- Marrone, T. J.; Briggs, J. M.; McCammon, J. A. Structure-based drug design: Computational advances. *Annu. Rev. Pharmacol.* **1997**, *37*, 71–90.
- 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. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, 380–384.
- Chen, J. M.; Xu, S. L.; Wawrzak, Z.; Basarab, G. S.; Jordan, D. B. Structure-based design of potent inhibitors of scytalone dehydratase: displacement of a water molecule from the active site. *Biochemistry* **1998**, *37*, 17735–17744.
- Mikol, V.; Papageorgiou, C.; Borer, X. The role of water molecules in the structure-based design of (5-hydroxynorvaline)-2-cyclosporine – synthesis, biological activity and crystallographic analysis with cyclophilin-A. *J. Med. Chem.* **1995**, *38*, 3361–3367.
- Cherbavaz, D. B.; Lee, M. E.; Stroud, R. M.; Koschl, D. E. Active site water molecules revealed in the 2.1 angstrom resolution structure of a site-directed mutant of isocitrate dehydrogenase. *J. Mol. Biol.* **2000**, *295*, 377–385.
- Finley, J. B.; Atigadda, V. R.; Duarte, F.; Zhao, J. J.; Brouillette, W. J.; Air, G. M.; Luo, M. Novel aromatic inhibitors of influenza virus neuraminidase make selective interactions with conserved residues and water molecules in the active site. *J. Mol. Biol.* **1999**, *293*, 1107–1119.
- Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: Placing discrete water molecules during protein–ligand docking predictions. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 17–28.
- Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspect. Drug Discov. Des.* **2000**, *20*, 171–190.
- Pospisil, P.; Kuoni, T.; Scapozza, L.; Folkers, G. Methodology and problems of protein–ligand docking: Case study of dirydoorotate dehydrogenase, thymidine kinase, and phosphodiesterase 4. *J. Recept. Signal Transduct. Res.* **2002**, *22*, 141–154.
- Grüneberg, S.; Stubbs, M. T.; Klebe, G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: Strategy and experimental confirmation. *J. Med. Chem.* **2002**, *45*, 3588–3602.
- Pastor, M.; Cruciani, G.; Watson, K. A. A strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative structure–activity relationship analysis. *J. Med. Chem.* **1997**, *40*, 4089–4102.
- Lloyd, D. G.; García-Sosa, A. T.; Alberts, I. L.; Todorov, N. P.; Mancera, R. L. The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 89–100.
- Mancera, R. L. De novo ligand design with explicit water molecules: an application to bacterial neuraminidase. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 479–499.
- The Protein Kinase Factsbook. Protein-Tyrosine Kinases*; Hardie, G., Hanks, S., Eds.; Academic Press: GB, 1995.
- Eliasson, M. J. L.; Sampei, K.; Mandir, A. S.; Hurn, P. D.; Traystman, R. J.; Bao, J.; Pieper, A.; Wang, Z. Q.; Dawson, T. M.; Snyder, S. H.; Dawson, V. L. Poly(ADP-ribose)polymerase gene disruption renders mice resistant to cerebral ischemia. *Nat. Med. (N. Y.)* **1997**, *3*, 1089–1095.
- Canan Koch, S. S.; Thoresen, L. H.; Tikhe, J. G.; Maegley, K. A.; Almassy, R. J.; Li, J.; Yu, X.-H.; Zook, S. E.; Kumpf, R. A.; Zhang, C.; Boritzki, T. J.; Mansour, R. N.; Zhang, K. E.; Calabrese, C. R.; Curtin, N. J.; Kyle, S.; Thomas, H. D.; Wang, L.-Z.; Calvert, A. H.; Golding, B. T.; Griffin, R. J.; Newell, D. R.; Webber, S. E.; Hostomsky, Z. Novel tricyclic poly(ADP-ribose) polymerase-1 inhibitors with potent anticancer chemopotentiating activity: Design, synthesis, and X-ray cocrystal structure. *J. Med. Chem.* **2002**, *45*, 4961–4974.
- White, A. W.; Almassy, R.; Calvert, A. H.; Curtin, N. J.; Griffin, R. J.; Hostomsky, Z.; Maegley, K.; Newell, D. R.; Srinivasan, S.; Golding, B. T. Resistance-modifying agents. 9. Synthesis and biological properties of benzimidazole inhibitors of the DNA repair enzyme poly(ADP-ribose) polymerase. *J. Med. Chem.* **2000**, *43*, 4084–4097.
- García-Sosa, A. T.; Mancera, R. L.; Dean, P. M. WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein–ligand complexes. *J. Mol. Model.* **2003**, *9*, 172–182.
- Daylight 2004 World Drug Index, Daylight Chemical Information Systems, 2004.
- Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
- Todorov, N. P.; Dean, P. M. A branch-and-bound method for optimal atom-type assignment in de novo ligand design. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 335–349.
- Stahl, M.; Todorov, N. P.; James, T.; Mauser, H.; Boehm, H.-J.; Dean, P. M. A validation study on the practical use of automated de novo design. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 459–478.
- Todorov, N. P.; Dean, P. M. Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 175–192.
- InsightII, Accelrys, Inc., San Diego, CA
- Dinur, U.; Hagler, A. T. In *Reviews in Computational Chemistry*, 2nd ed.; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers Inc.: U.S.A., 1991.
- Banasik, M.; Komura, H.; Shimoyama, M.; Ueda, K. Specific inhibitors of poly(ADP-ribose)synthetase and mono(ADP-ribosyl)transferase. *J. Biol. Chem.* **1992**, *267*, 1569–1575.
- Huang, M. J.; Maynard, A.; Turpin, J. A.; Graham, L.; Janini, G. M.; Covell, D. G.; Rice, W. C. Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins. *J. Med. Chem.* **1998**, *41*, 1371–1381.
- Constantino, G.; Macchiarulo, A.; Camaioni, E.; Pellicciari, R. Modeling of poly(ADP-ribose)polymerase (PARP) inhibitors. Docking of ligands and quantitative structure–activity relationship analysis. *J. Med. Chem.* **2001**, *44*, 3786–3794.
- Denisov, V. P.; Venu, K.; Peters, J.; Horlein, H. D.; Halle, B. Orientational disorder and entropy of protein in protein cavities. *J. Phys. Chem. B* **1997**, *101*, 9380–9389.
- Ruf, A.; Rolli, V.; de Murcia, G.; Schulz, G. E. The mechanism of the elongation and branching reaction of poly(ADP-ribose)polymerase as derived from crystal structures and mutagenesis. *J. Mol. Biol.* **1998**, *278*, 57–65.
- Tikhe, J. G.; Webber, S. E.; Hostomsky, Z.; Maegley, K. A.; Ekkers, A.; Li, J.; Yu, X.-H.; Almassy, R. J.; Kumpf, R. A.; Boritzki, T. J.; Zhang, C.; Calabrese, C. R.; Curtin, N. J.; Kyle, S.; Thomas, H. D.; Wang, L.-Z.; Calvert, A. H.; Golding, B. T.; Griffin, R. J.; Newell, D. R. Design, synthesis, and evaluation of 3,4-dihydro-2H-[1, 4]-diazepino[6,7,1-h]indol-1-ones as inhibitors of poly(ADP-ribose) polymerase. *J. Med. Chem.* **2004**, *47*, 5467–5481.
- Zhang, L.; Hermans, J. Hydrophilicity of cavities in proteins. *Proteins: Struct., Funct., Genet.* **1996**, *24*, 433–438.
- Hoof, R. W. W.; Sander, C.; Vriend, G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins: Struct., Funct., Genet.* **1996**, *26*, 363–376.