ELSEVIER

Chemistry & Biology 8 (2001) 445–457

Chemistry & Biology

www.elsevier.com/locate/chembiol

Research Paper

Approaches to solving the rigid receptor problem by identifying a minimal set of flexible residues during ligand docking[1]

Amy C. Anderson [2,*], Robert H. O'Neil [2], Toral S. Surti, Robert M. Stroud

Department of Biochemistry and Biophysics, University of California at San Francisco, Box 0448, San Francisco, CA 94143-0448, USA

Received 8 November 2000; revisions requested 16 January 2001; revisions received 19 February 2001; accepted 5 March 2001
First published online 27 March 2001

Abstract

Background: Using fixed receptor sites derived from high-resolution crystal structures in structure-based drug design does not properly account for ligand-induced enzyme conformational change and imparts a bias into the discovery and design of novel ligands. We sought to facilitate the design of improved drug leads by defining residues most likely to change conformation, and then defining a minimal manifold of possible conformations of a target site for drug design based on a small number of identified flexible residues.

Results: The crystal structure of thymidylate synthase from an important pathogenic target *Pneumocystis carinii* (PcTS) bound to its substrate and the inhibitor, BW1843U89, is reported here and reveals a new conformation with respect to the structure of PcTS bound to substrate and the more conventional antifolate inhibitor, CB3717. We developed an algorithm for determining which residues provide 'soft spots' in the protein, regions where conformational adaptation suggests possible modifications for a drug lead that may yield higher affinity. Remodeling the active site of thymidylate synthase with new conformations for only three residues that were identified with this algorithm yields scores for ligands that are compatible with experimental kinetic data.

Conclusions: Based on the examination of many protein/ligand complexes, we develop an algorithm (SOFTSPOTS) for identifying regions of a protein target that are more likely to accommodate plastically to regions of a drug molecule. Using these indicators we develop a second algorithm (PLASTIC) that provides a minimal manifold of possible conformations of a protein target for drug design, reducing the bias in structure-based drug design imparted by structures of enzymes co-crystallized with inhibitors.

Keywords: Crystal structure; Docking; Ligand design; Protein flexibility; Thymidylate synthase

1. Introduction

Computational approaches to structure-based drug design primarily rely on a high-resolution structure of the target receptor against which small molecules can be docked and scored for interactions. It is known, however, that proteins undergo conformational changes to accommodate a ligand and increase the number of protein–ligand interactions. Once observed and described, these protein conformational changes are incorporated into a few novel docking and scoring algorithms, as reported in a recent review [1], but traditionally are not included due to the lack of a comprehensive rule set that would allow a fast and accurate algorithm. The existing algorithms that allow a degree of protein flexibility include methods based on multiple structure alignment or single semi-flexible structures. Ensembles of structures for multiple-conformation docking are generated either through molecular dynamics [2,3], nuclear magnetic resonance solution structures, or by the existence of multiple crystal structures [4]. Algorithms using semi-flexible single structures are soft-docking techniques [5,6], minimization of side-chain rotamers [7,8], and side-chain minimization using Monte-Carlo simulations or molecular dynamics [9]. The majority of these algorithms incorporate flexibility for all residues

Abbreviations: TS, thymidylate synthase; PcTS, TS from *Pneumocystis carinii*; dUMP, deoxyuridine monophosphate; dTMP, deoxythymidine monophosphate; 5FU, 5-fluorouridine; rmsd, root mean square deviation; EcTS, TS from *Escherichia coli*; HIV, human immunodeficiency virus

[1] PDB coordinates have been deposited with the RSCB with accession ID: 1F28.

[2] Present address: Department of Chemistry, Dartmouth College, Burke Laboratories, Hanover, NH 03755, USA.

* Correspondence: A.C. Anderson;
E-mail: amy.c.anderson@dartmouth.edu

1074-5521/01/$ – see front matter
PII: S1074-5521(01)00023-0

in the active site. We show here that if conformational changes can be relegated to a few specific residues and their trajectories can be projected then it is possible to compute a minimal manifold of structures as input to a more exactly defined docking target resulting in better scores for known inhibitors.

Thymidylate synthase (TS) catalyzes the final step in the de novo synthesis of deoxythymidine monophosphate (dTMP) using the substrate deoxyuridine monophosphate (dUMP), and a natural cofactor, 5,10-methylene tetrahydrofolate. Because of its essential role in DNA replication, human TS is an anticancer drug target [10–14] and TS from *Pneumocystis carinii* (PcTS) is an infectious disease drug target [15]. Previous inhibitors developed against human TS have focused on substrate mimics: 5-fluorouridine (5FU) [16] and cofactor mimics: CB3717 [11,17], ZD1694 [10] and BW1843U89 [14,18]. Several potent and novel structure-based inhibitors such as AG337 [13], phenolphthalein [19] and antibacterial-specific agents have also been developed [20].

Additional disease targets of interest include TS from other pathogens including *Cryptococcus neoformans*, *Toxoplasma gondii* and *Cryptosporidium parvum*. Since TS is one of the most highly conserved enzymes it directly challenges and hones our ability to design, and/or to develop ligands that show adequate species selectivity over human TS.

The quest for an accurate evaluation of novel species-specific inhibitors discovered through computational chemistry requires precise models of physically accessible conformers of the several species. Identification and conformational sampling of only the principal residues involved in ligand-induced structural changes will reduce this otherwise potentially large conformational space for each species. This reduction in enzyme conformational sampling will allow the practical search for more specific inhibitors without the bias that is imparted by the conformation of the enzyme in the ternary complex of a single inhibited crystal structure.

In order to reduce the discrepancy between docking scores of known good inhibitors to different conformations of the same active site, we sought to identify a minimal set of residues undergoing conformational changes, and to compute a manifold of structures based on different conformations for only those residues. The identification of flexible residues was first accomplished by comparing crystal structures of PcTS bound to dUMP and either CB3717 or BW1843U89 (a structure determination reported here). This knowledge was then used first to see if good scores could be correctly achieved by docking against the limited manifold of structures generated by incorporating discretely supplied rotamers of just this small subset. It was then used again in the first iteration of discovery aimed at defining an algorithm (SOFTSPOTS) that would allow conformationally flexible residues to be identified in this or in any protein target, where multiple structures either may or may not be available. Using this predictive algorithm we identify three residues in PcTS crucial to ligand adaptation. Allowing these three residues to change conformation dramatically reduced the rigid receptor problem for PcTS in computational docking and scoring. The principles developed here can be applied toward other enzymes that undergo conformational change, allowing identification of specific adaptable residues in those systems and a concomitant reduction in enzyme conformational sampling during docking.

2. Results and discussion

2.1. The crystal structure of PcTS Bound to dUMP and BW1843U89

The crystal structure of PcTS bound to dUMP and BW1843U89 was recently determined (statistics for the data collection and model refinement are found in Table 1). The enzyme (Fig. 1) forms an obligate dimer that binds two molecules of the substrate, dUMP, and two molecules of the antifolate inhibitor, BW1843U89 (Fig. 2 compares the two-dimensional structures of BW1843U89 with CB3717, another antifolate inhibitor discussed in this work). The crystal structure was solved with diffraction amplitudes extending to 1.9 Å resolution, the highest resolution data for any eukaryotic TS structure solved to date. The enzyme (MW 70 kDa) has a five-stranded β-sheet interface and an active site comprised of nine conserved helices. There are two dimers in the asymmetric unit, related by a two-fold non-crystallographic symmetry axis. The two monomers within a dimer are structurally independent and have asymmetric interactions with the

Fig. 1. Ribbon diagram of the structure of PcTS bound to dUMP and BW1843U89. The ligands are shown in black ball-and-stick representation.

Table 1
Refinement statistics for PcTS/dUMP/BW1843U89

Space group	$P2_1$
Unit cell edges	$a = 54.4$, $b = 65.9$, $c = 184.9$, $\beta = 90.2$
Resolution (Å)	50–1.9
Observed reflections	530 911
Unique reflections	104 488
Completeness (%)	90.3
$I/\sigma(I)$	14.4
R_{merge}	0.088
R_{cryst} (50–1.9 Å)	0.216
R_{free} (50–1.9 Å)	0.247
Total no. atoms	10 125
No. water molecules	499
Rmsd bonds (Å)	0.0126
Rmsd angles (°)	1.67
Average *B*-factor (Å^2)	31.07

ligands. Both monomers show the 'closed' conformation: the four C-terminal residues have moved approximately 4 Å toward the active site to interact with the antifolate and with residues in the active site. The ternary closed conformation differs from the apo or binary conformation and serves to orient the ligands for catalysis, sequester them from bulk solvent and protect the high-energy reaction intermediates [21].

2.2. Ligand binding induces change in the receptor

One dimer of PcTS binds two molecules of the substrate, dUMP, and antifolate, BW1843U89 ($K_i = 16$ nM for PcTS), although each monomer has asymmetric interactions with the ligands due to the mechanism of cooperativity described below. Electron density for both ligands was clearly visible in the initial maps (Fig. 3) and the ligands were placed unambiguously in the model of the enzyme. PcTS binds BW1843U89 by undergoing conformational changes at the active site that maximize hydrophobic interactions (Fig. 4a,c) and where, specifically, the isoindolinone of BW1843U89 induces the majority of this conformational change. There are at least five strong hydrophobic interactions between the inhibitor and the enzyme including those between Met 295, Trp 87 and Leu 205 with the benzoquinazoline. Ile 86 stacks against the isoindolinone moiety that in turn stacks against Phe 209. The isoindolinone moiety forms additional van der Waals interactions with Gly 61. The Connolly surface representation in Fig. 4c reveals the tightly closed active site and in comparison with the surface representation of PcTS/dUMP/CB3717 (Fig. 4d) indicates the extent of the conformational change in the active site. There are also four hydrogen bonds between the enzyme and the inhibitor: Asp 202 with N2, the amide nitrogen of Gly 206 and the keto oxygen O1A (both N2 and O1A are members of the benzoquinazoline moiety), Glu 65 with N12, the link between the benzoquinazoline and isoindolinone moieties, and an ordered water molecule to the keto oxygen of the isoindolinone.

Substrate interactions in the PcTS/dUMP/BW1843U89 complex (Fig. 4a) include hydrogen bonds between the phosphate moiety and both arginine residues contributed from the opposite monomer (Arg 153′ and Arg 154′), Arg 199, Ser 200 and an ordered water molecule. The ribose of dUMP forms hydrogen bonds to Tyr 242 and His 240 and the pyrimidine base forms hydrogen bonds from O4 to an ordered water molecule and from O4 and N3 to Asn 210. Asn 202 forms a hydrogen bond to O2. This structure of PcTS bound to dUMP and BW1843U89 reveals the canonical set of hydrogen bonds between TS and dUMP and indicates the tightly closed nature of the BW1843U89 active site (Fig. 4a,c). The structure of PcTS bound to dUMP and CB3717 [15] (Fig. 4b) has many fewer hydrogen bonds between the enzyme and substrate, indicating a more open active site complex. Hydrogen bonds present in the BW1843U89-inhibited structure but not in the CB3717-inhibited structure include: NH2 of Arg 199 and OP2, Ser 200 and OP1, Nε of both Arg 154′ and 153′ to OP2 and OP3, respectively, His 240 and the ribose O3′ and from ordered water molecules to O4 and OP3.

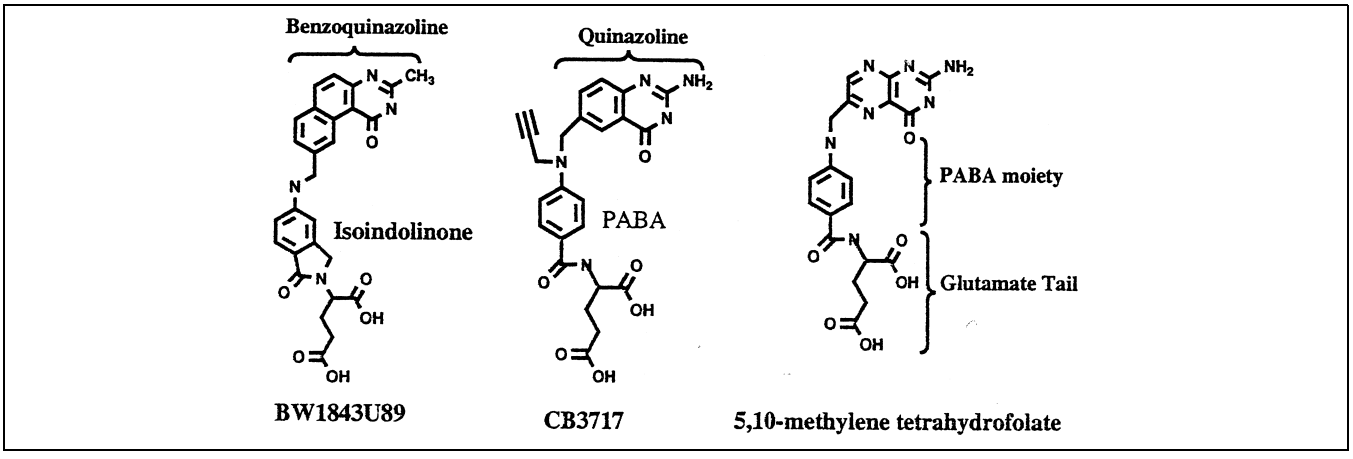

Fig. 2. Two-dimensional structures of the cofactor analogs.

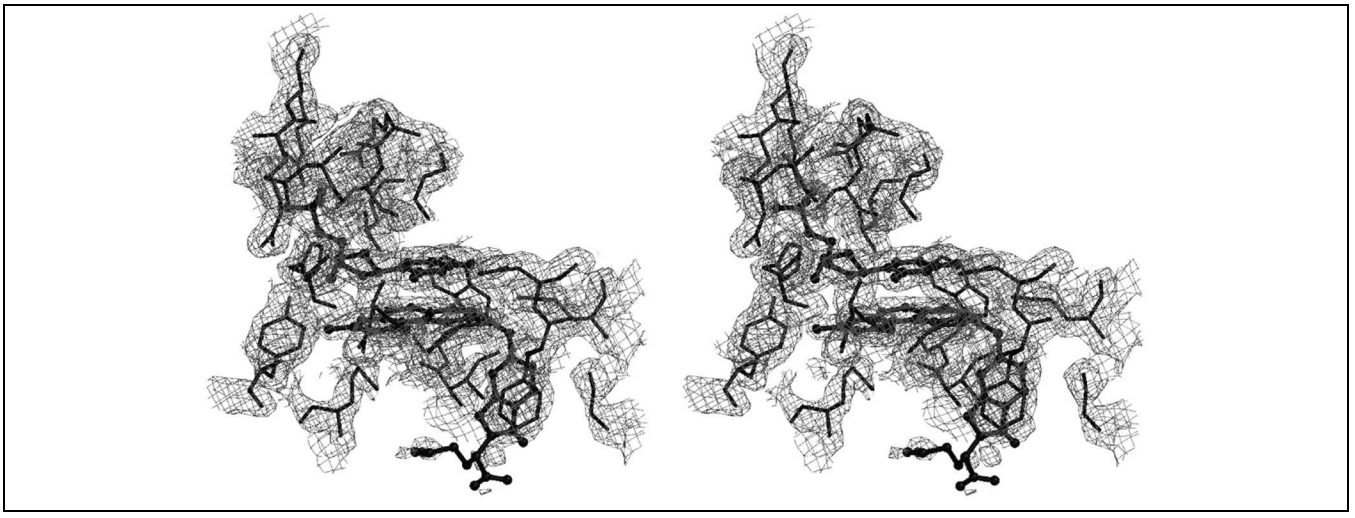

Fig. 3. Electron density for the ligands and residues at the active site. The ligands are shown in black ball-and-stick representation.

Thymidylate synthase converts the substrate, dUMP, to a product, dTMP, in a reaction which involves a covalent Michael addition of the active site cysteine (Cys 173 for PcTS) to C6 of the pyrimidine ring of dUMP. In this structure of PcTS bound to dUMP and BW1843U89, a covalent bond forms in monomer A of each dimer in the asymmetric unit, but not in monomer B. A similar situation occurs in the structure of PcTS bound to dUMP and CB3717, for which we described a structure-based half-the-sites reactivity mechanism [22]. The half-the-sites reactive mechanism uses a relay of changes between subunits and across the dimer interface to allow a covalent bond to form for only one monomer of the dimeric enzyme. When one covalent bond forms, a second is prevented from forming simultaneously in the opposite monomer.

2.3. Failure of the rigid receptor hypothesis

Two computational experiments using DOCK 4.0, developed at UCSF, that dock and score ligands in active site pockets reveal the inadequacies of the rigid receptor model for TS. In one experiment, CB3717, in a rigid conformation, is docked into the PcTS site from the crystal structure of PcTS/dUMP/CB3717 and also into the structure of PcTS from the dUMP/BW1843U89 complex. The same swapping experiment is performed with the BW1843U89 compound and the energy scores for all four experiments are shown in Table 2 (analogous to a ΔG calculation, a negative score indicates a stronger interaction). Force-field scoring algorithms for these experiments, for which the relative values are most relevant, are based on both electrostatic and van der Waals interactions [23]. The results show that both compounds score well when docked into the site from which they were derived and score poorly when docked into the 'opposite' PcTS structure. Simple minimization of the protein/inhibitor complexes at this step does not relieve this problem since there is significant steric hindrance that is relieved only by the choice of different rotamers for particular residues. The radius of convergence of minimization is small and cannot surmount the energy barriers between rotameric states of residues. Similarly, including hydration corrections for the effect of solvent in the active site does not relieve this problem since hydration corrections improve the scores of docked ligands when the ligand can actually fit into the active site and is not hindered by steric

Table 2
Force-field scores from computational chemistry experiments

Structure from which site is derived	Docked compound		Top 500 compounds from ACD (low/high)
	CB3717	BW1843U89	
Rigid receptor/rigid ligands			
PcTS/dUMP/CB3717	*−52.9*[a]	−44.7	−45.3/−60.8
PcTS/dUMP/BW1843U89	−32.8	*−56.5*	−37.8/−48.7
Rigid receptor/flexible ligands			
PcTS/dUMP/CB3717	*−57.5*	−48.6	−52.0/−69.8
PcTS/dUMP/BW1843U89	−51.7	*−56.6*	−48.3/−59.2
Flexible receptor/rigid ligands			
Remodeled PcTS/dUMP/CB3717		−51.5	
Remodeled PcTS/dUMP/BW1843U89	−49.7		

[a]Scores in italics represent those from a compound docked into its native receptor.

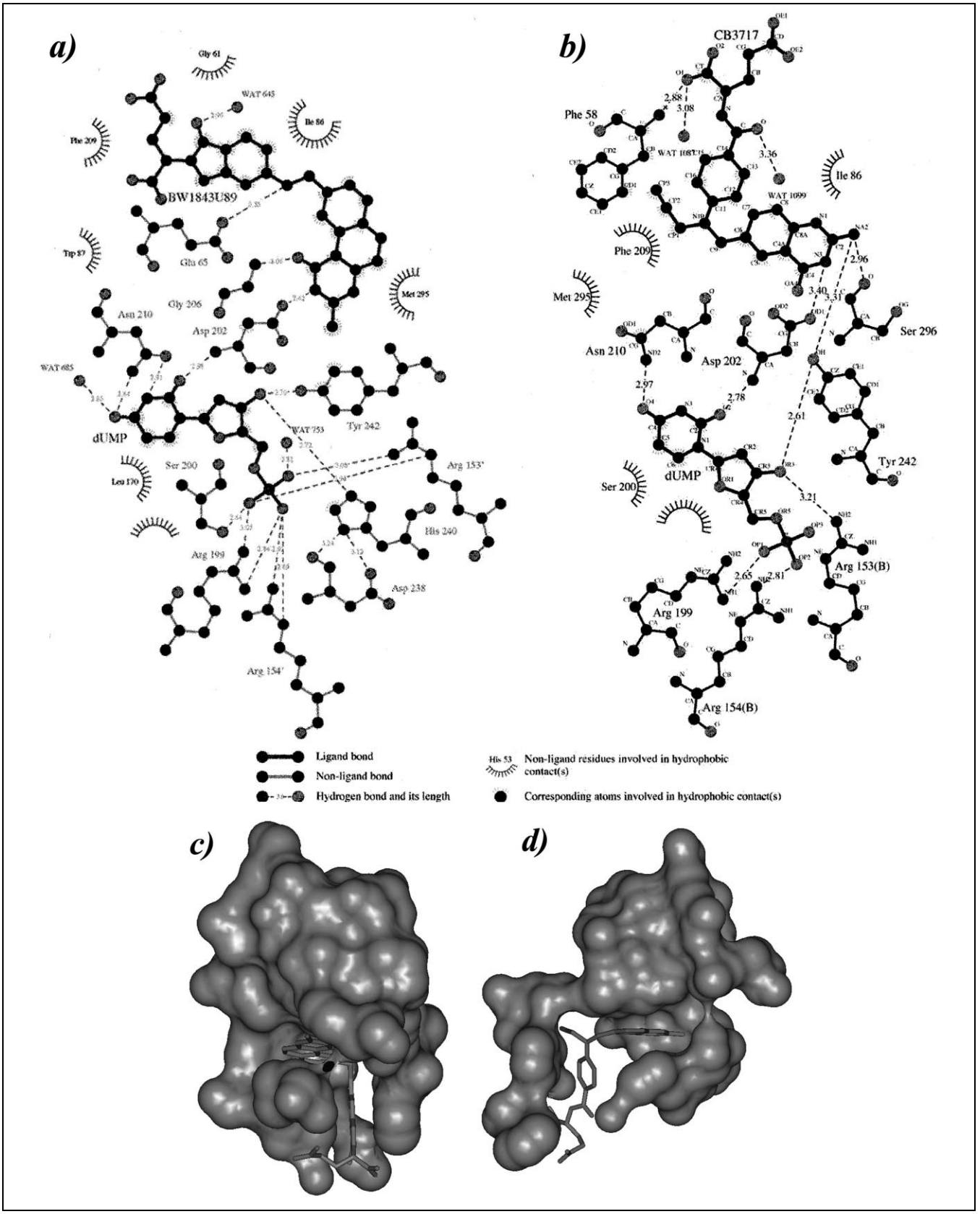

Fig. 4. LIGPLOT schematic of the interactions between PcTS and (a) dUMP and BW1843U89 and (b) dUMP and CB3717. Connolly surface representations for the active sites surrounding (c) BW1843U89 and (d) CB3717.

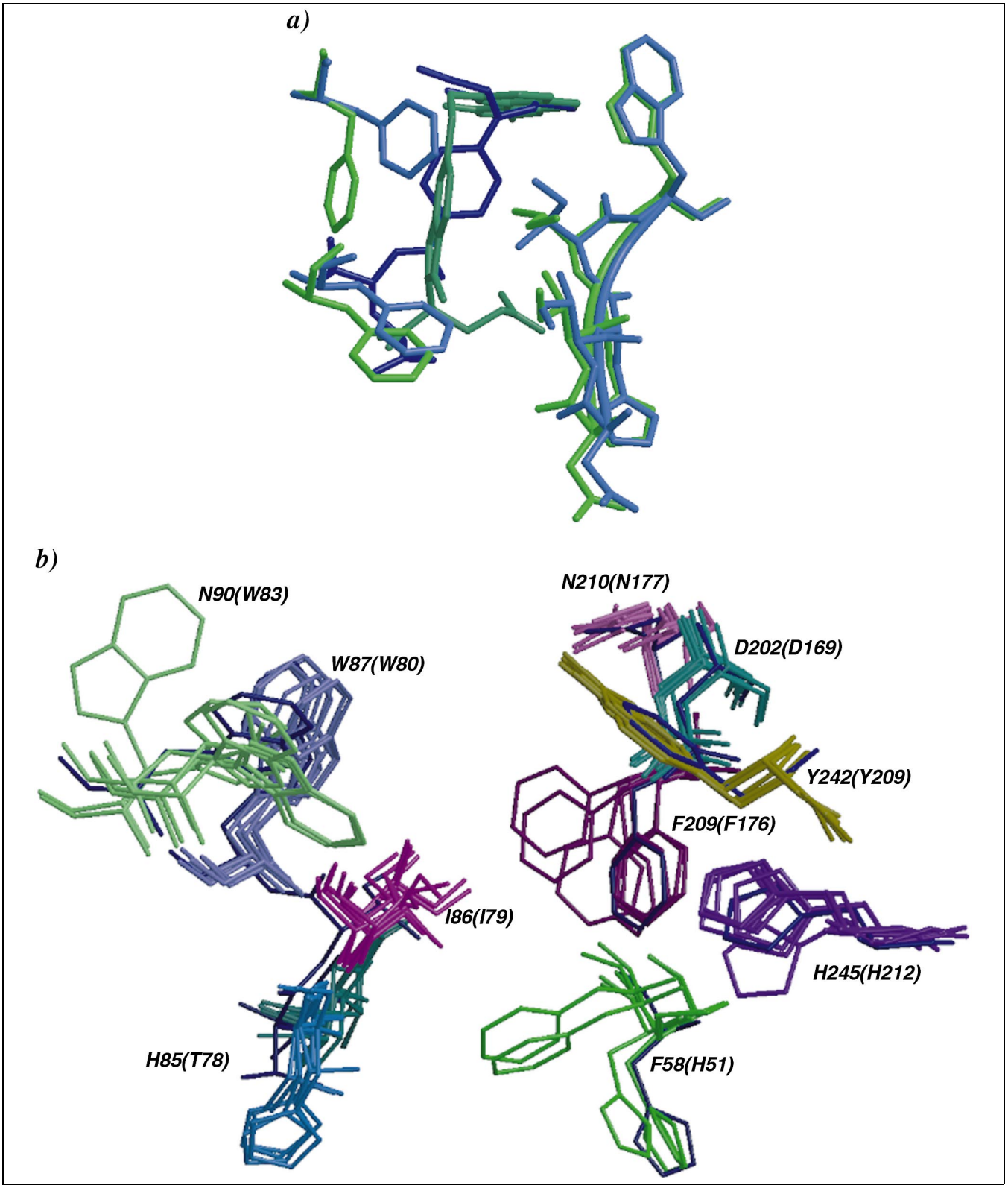

Fig. 5. (a) Ligand-induced conformational changes between PcTS/BW1843U89 (green) and PcTS/CB3717 (blue). (b) Superposition of seven structures of TS: TS from *Escherichia coli* bound to dUMP and either CB3717, BW1843U89 or ZD1694, PcTS bound to dUMP and either CB3717 or BW1843U89 and TS from *Lactobacillus casei* bound to A156 or MR29. Individual residues related by sequence alignment are shown in a single color and all residues of EcTS in the unbound, apo, form are shown in dark blue.

Table 3
Derivation of an algorithm for prediction of flexible side chains in TS

Residues in active site forming interactions with inhibitor	Hydrophobic residues and loop residues	Hydrophobic residues with available space in site and loop residues	Altered conformers revealed by comparison of structures
F58	F58	F58	F58
E65			
I86	I86	I86	I86
W87	W87		
D202			
L205	L205	L205	
G206			
F209	F209	F209	F209
M295	M295	M295	

repulsion between its atoms and the protein atoms. More drastic enzyme rearrangement is needed in order to correct the discrepancies of the cross-docked ligands (hence the rigid receptor problem). Knegtel et al. [4] also observe lower energy scores when a ligand is docked into its respective site and higher energy scores when docked into the same site with a new conformation. Yet we know from solution experiments that both compounds are potent PcTS inhibitors: CB3717 has an IC_{50} value of 90 nM for PcTS [12] and BW1843U89 has a K_i value of 16 nM. Therefore choosing one structure and applying the rigid receptor assumption imposes a bias for the outcome of the experiment and excludes potentially excellent inhibitors from consideration.

A second computational experiment investigated how many molecules from a database search would receive good energy scores in both conformations of PcTS. A subset of molecules (110 070 compounds) from the Available Chemicals Database (ACD) 97.0 was docked, maintaining a rigid conformation, into the active sites of both crystal structures and scored using the same force-field scoring algorithm. Only 42 compounds are common to both lists of the top 500 compounds and furthermore the ranking order of these 42 compounds is not preserved. Again, employing only one of two enzyme conformations as a receptor imparts a bias into the selection of low-energy scoring candidate compounds.

Another conclusion from the database docking experiment is that BW1843U89, which has an energy score of −44.7, would not be selected in the list of 500 top-scoring compounds if the PcTS/dUMP/CB3717 model were the only one available, despite a 16 nM K_i value in solution. Likewise, CB3717, with an energy score of −32.8, would not be selected in the list of 500 top-scoring compounds if the PcTS/dUMP/BW1843U89 model were the only one available, despite its IC_{50} of 90 nM. Conversely, BW1843U89 receives a better score than any other small molecule in the database when docked into its own receptor site and CB3717 receives a score within the top 40 ligands when docked into its own site.

A third experiment examined whether treating the ligands as flexible molecules but maintaining a rigid site ameliorated the rigid receptor problem. The same swapping experiment outlined above for the rigid ligands was repeated with flexible ligands. In this experiment, torsion angles of the ligand were allowed to vary (results tabulated in Table 2). In this scenario the energy scores for the two ligands are closer (−51.7 for CB3717 docked into the BW1843U89 model and −56.6 for BW1843U89 docked into its own site) when the site is derived from the BW1843U89 model, a tightly closed active site. However, when the scores of the single flexible ligands are compared to the scores from the subset of the ACD when docked flexibly into the sites, it is evident that the score for CB3717 falls at the bottom of the list of top 500 scoring compounds and therefore would be unlikely to be selected for further investigation. The energy scores for the two ligands are still not comparable when the more open active site from the CB3717 model is used (−57.5 for CB3717 docked into its own site, −48.6 for BW1843U89 docked into the site from the CB3717 model). The score for BW1843U89 docked flexibly into the site from PcTS/dUMP/CB3717 does not fall within the range of the top 500 scoring compounds and therefore BW1843U89 probably would not be selected for further investigation. In addition, on closer examination of the docked conformation of the flexible ligands, the final conformation of the ligand is not the crystallographically observed conformation (rmsd 3.18 Å for BW1843U89 and rmsd 3.78 Å for CB3717). In conclusion, treating the ligand flexibly, either with discrete torsion angles or as a fully flexible ligand, may allow a conformation that will fit the altered active site. That ligand conformation, however, may exhibit large deviations from crystallographically-observed conformations.

2.4. Identification of ligand-induced conformational changes using structure comparison

With the goal of developing a minimal manifold of accessible conformations of an active site for more accurate docking and scoring experiments, principal residues responsible for conformational changes were identified with a structure-based comparison. Conformational

changes are noticeable in a comparison of structures of PcTS bound to either BW1843U89 or CB3717, an antifolate closely related to the cofactor. The two PcTS structures share the same crystal packing arrangement and have an overall rms deviation of 0.66 Å for all Cα atoms. A few key hydrophobic residues in PcTS undergo conformational changes in order to form ideal van der Waals interactions with the isoindolinone moiety of BW1843U89 (see Fig. 5a). The benzoquinazoline moiety of BW1843U89 lies in the same space and orientation, immediately below the pyrimidine and ribose rings of dUMP, as the quinazoline ring of CB3717. The benzoquinazoline moiety, although larger than the quinazoline, or the pterin group of the cofactor (see Fig. 2), does not induce any enzyme accommodation since the third ring occupies the same space as the propargyl group. The isoindolinone moiety of BW1843U89, however, lies in a plane 90° from the *p*-aminobenzoic acid moiety of CB3717. Importantly, Phe 209 rotates 90° away, changing its rotamer, from its conformation in the CB3717 structure to form a strong stacking arrangement with the isoindolinone while Ile 86 translates 1.2 Å toward the active site, continuing the stacking arrangement. The two neighboring residues on each side of Ile 86 also translate 0.8–1.3 Å toward the active site. The backbone of Phe 58, near the glutamate tail, shifts 0.7 Å away from the inhibitor in order to avoid a steric conflict. This shift removes the potential hydrogen bond between the backbone of Phe 58 and the glutamate tail that is formed in the PcTS/CB3717 complex. Finally, residues 243–251 move 1.2–1.4 Å toward the active site in the BW1843U89 complex, completing the closure of the hydrophobic pocket comprised of the isoindolinone, Phe 209, Leu 205 and Met 293.

PcTS and TS from *Escherichia coli* (EcTS) undergo similar conformational changes on binding BW1843U89, emphasizing the importance of the increased hydrophobic interactions of the enzyme with the inhibitor and especially with the isoindolinone moiety. Two structures of EcTS were compared: EcTS/dUMP/CB3717 [24] and EcTS/dUMP/BW1843U89 [20,21], both in the same space group and with the same unit cell dimensions. Similar to PcTS, Phe 176 from EcTS (PcTS:F209) rotates 90° to stack against the isoindolinone of BW1843U89 and His 51 (PcTS:F58) shifts 1.1 Å away from the active site to remove steric conflicts with the glutamate tail. Preserving the similarity in conformational change, Ile 79 (PcTS:I86) stacks against the opposite face of the isoindolinone in both structures.

Structures of TS bound in ternary complexes with varying ligands were compared (Fig. 5b) in order to determine the relative active site plasticity over several species and ligand combinations. The ligands involved include CB3717, BW1843U89, ZD1694 (all antifolate derivatives) as well as A156 and MR29, two phenolphthalein derivatives with very different binding motifs. It is evident from this comparison that Phe 209, Ile 86, and Asn 90 (Trp in prokaryotic species) change conformation depending on the species and ligand bound. Asn 90 (for which the prokaryotic substitution, Trp, undergoes conformational change) changes its rotamer in the presence of inhibitors that are unrelated to the antifolate scaffold. The polar residues Asn 210, Asp 202, and Tyr 242 remain relatively fixed, independent of the ligand bound or the species studied. Phe 58 (PcTS) is a histidine in the prokaryotic species and this side chain adopts a different conformation. His 245 adopts a different conformation in the structure of PcTS bound to dUMP and CB3717 but otherwise remains fixed. The apo structure of EcTS is shown in Fig. 5b in dark blue and represents the conformational change from the unbound form of the enzyme to the ternary bound form, also revealing the flexibility of the hydrophobic residues. From this comparison it is evident that the hydrophobic residues are subject to ligand-induced accommodation while the polar residues do not undergo conformational changes in order to form hydrogen bonds.

In subsequent experiments (Fritz et al., in preparation) we show, using entirely novel inhibitors of TS unrelated to any of those mentioned above, that the same residues provide flexibility and accommodate binding, independently suggesting that common elements of the protein target site do provide for plasticity in a general way.

2.5. Can we predict flexible residues within a single crystal structure?

In order to predict flexible active site residues using a single crystal structure, we derived a simple selection scheme to identify the residues that are most crucial to the conformational change. This scheme is based on a structural evaluation of ligand-induced conformational changes from several enzymes including thymidylate synthase, glycogen phosphorylase, aldose reductase and human immunodeficiency virus (HIV) reverse transcriptase. The principles have been incorporated in an algorithm for flexible side-chain selection, outlined in Table 3 and described in detail below.

1. Include all residues in the active site capable of interacting with the inhibitor.
2. Select hydrophobic residues and residues belonging to loop regions from this list.
3. Retain loop residues and further select for hydrophobic residues that allow room for conformational changes, determined by assessing the steric hindrance of rotamers of the side chain.
4. If two structures of the enzyme are available, compare the two structures to validate the choice of residues from steps one through three and add any additional residues exhibiting a conformational change. In three cases of our application, including TS, this latter razor did not lead to new additions.

After all residues in the active site have been tabulated (step 1), a series of eliminations leaves a small number of residues predicted to be capable of conformational change. In the first step of elimination, step 2 of the algorithm, hydrophobic residues and loop residues are maintained from the initial list of active site residues.

Conformational changes of hydrophobic residues that improve van der Waals interactions with inhibitor atoms are emphasized in an analysis of enzymes that undergo the largest ligand-induced changes. These include seven structures of TS, and therefore are included in an algorithm for predicting flexible residues. In PcTS, Phe 209 rotates 90° to form a stacking interaction with the isoindolinone and Ile 86 shifts 1.3 Å to continue the stacking interaction on the other side of the isoindolinone.

There are several additional examples that reveal an emphasis on hydrophobic residues in ligand-induced conformational change. Bay W1807, a potent inhibitor for glycogen phosphorylase, is almost completely buried within the protein due to the rearrangement of several hydrophobic residues for improved interaction with the inhibitor [25]. In further examples, hydrophobic residues in the active site of aldose reductase undergo significant shifts in response to the inhibitors zopolrestat [26] and tolrestat [27] and HIV reverse transcriptase binds dissimilar ligands through accommodation of hydrophobic residues [28]. Davis and Teague [29] recently reviewed the important role of hydrophobic interactions in ligand-induced conformational changes.

Included in step 2, polar residues are removed from the list since hydrogen bond formation does not appear to drive conformational changes as much as hydrophobic residues, as evidenced from the comparison of these TS structures and from structures of many other enzymes published in the literature [29]. There are three enzyme-mediated hydrogen bonds from PcTS to BW1843U89. The residues that form hydrogen bonds to the inhibitor do so if the distance and geometry are correct but do not distort the enzyme conformation if not. Also, it is evident that the enzyme forms two more hydrogen bonds to CB3717 than it does to BW1843U89 and yet BW1843U89 ($K_i = 16$ nM) has a smaller inhibition constant for PcTS than CB3717 ($K_i = 90$ nM). Therefore, sacrificing two hydrogen bonds while gaining hydrophobic interactions can result in a smaller inhibition constant.

Hydrogen bond formation in the enzyme glycogen phosphorylase also does not appear to determine either the conformation of the enzyme or the efficacy of its inhibitors. Glucose 6-phosphate, the natural substrate, has two additional hydrogen bonds as compared to the potent inhibitor, Bay W1807, yet Bay W1807 has optimized hydrophobic interactions with Phe 196 and Val 45′ in order to effectively inhibit the enzyme [25]. Electrostatic interactions and hydrogen bond formation play a very minor role for inhibitors of aldose reductase [27] and HIV reverse transcriptase [28].

Residues in loop regions that are capable of interacting with the inhibitor are also retained as potential candidates for plastic adaptation to a ligand since these residues are less constrained than residues in the core of the enzyme. For TS, hydrophobic residues Ile 86, Phe 209, Trp 87, Met 295, Leu 205, as well as Phe 58, a loop residue, are retained at this step.

In step 3, a second series of elimination, loop residues and hydrophobic residues with space for rotamer variation are retained as potential candidates. For each hydrophobic residue, each side-chain rotamer is tested and a bump check is used to determine whether sufficient space exists for that conformation. For this study, the rotamer library, totaling 153 rotameric states for all residues, was determined from 240 high resolution (1.7 Å or higher) protein structures and was filtered to remove misfit conformations [30]. If no other rotamer position is valid for a particular side chain, that residue is eliminated from the list of candidates. This elimination scheme disallows residue Trp 87 from TS, reducing the list of flexible candidates to Phe 209, Ile 86, Met 295 and Leu 205 as well as the loop residue, Phe 58. Without multiple crystal structures, these residues (present in column three of Table 3) would represent the list of potential flexible candidates.

In the fourth and final step, column four of Table 3, any existing crystal structures of the receptor are compared. This comparison can be viewed as an independent criterion to validate the original choice of flexible residues, or can add or eliminate residues from that list. For TS, hydrophobic residues Phe 209 and Ile 86 are retained as potential flexible residues, as well as Phe 58, because of drastic conformational changes that take place in the experimental structures at those sites. In the comparison of structures of PcTS/dUMP/BW1843U89 or CB3717, there are no major conformational changes near the benzoquinazoline binding pocket and therefore, for the ensuing comparison of docking BW1843U89 versus CB3717 only, Met 295 and Leu 205, two residues that interact with the benzoquinazoline group, are not included as flexible residues. When other ligands, including other pterin or quinazoline-based inhibitors, are docked, the contributions from Met 295 and Leu 205 should be considered.

This scheme for predicting flexible residues at the active site was also validated when applied to other enzymes known to exhibit ligand-induced accommodation. Applying this predictive scheme to aldose reductase yields hydrophobic residues Phe 122, Tyr 48, Trp 20, Leu 300, Trp 79 and Trp 111 after step 2 where all hydrophobic and loop residues interacting with the inhibitor are retained as potential candidates. Step 3 retains residues with space for rotamer variation, eliminating all other residues and leaving Phe 122 and Leu 300 as potential flexible residues. These two remaining candidate flexible residues, Phe 122 and Leu 300, have been noted by the authors to undergo major conformational changes in a comparison of struc-

tures of the enzyme bound to either tolrestat or sorbinol [27].

While these rules are more complex than simply using the crystallographic temperature factors associated with residues, they are also more complete. It should be noted that crystallographic temperature factors do not necessarily define all flexible regions in a protein. A surface loop on a protein may have high temperature factors, depicting a range of motion, but a buried residue undergoing a rotamer shift may have low temperature factors in each alternative position. In PcTS, Phe 209 has a low temperature factor (average 12.8 for the residue, 22.2 for the molecule) in rotamer one in the structure with dUMP and CB3717 and also in rotamer two (average 24.9 for the residue, 31.07 for the molecule) in the structure with dUMP and BW1843U89. Therefore, residues with high temperature factors do not indicate the entire set of flexible residues in a protein.

2.6. Defining a flexible receptor

For this enzyme, the discrepancy of rigid receptor docking scores was reduced by applying the rules of prediction of conformational change to identify three flexible residues in the active site of PcTS. Two strategies were compared: 1) switching the conformations of the three flexible residues in one site to those seen in the other site and docking the inhibitors and, 2) varying the rotamers of the flexible residues prior to docking the inhibitors. The second strategy is preferred since it uses no prior knowledge of the conformational change and can be employed without the existence of multiple structures. Employing the first strategy, switching the conformations of only three residues from those in the PcTS/dUMP/CB3717 structure to those seen in the PcTS/dUMP/BW1843U89 structure, was a proof-of-concept test that shows that improved docking scores could result from the modification of only a small number of residues in the active site. With this switch in the PcTS/dUMP/CB3717 structure, for example, the docking score for BW1843U89 improves to −55.1 (from −44.7 in the unaltered PcTS/dUMP/CB3717 structure).

The goal of this work, however, is to achieve an accurate minimal manifold of active site conformations without the prior knowledge of multiple structures and therefore this manifold was achieved with a rotamer library for the identified side chains. Using both conformations of the active site of PcTS, the rotamers of Phe 209, Ile 86 and Phe 58 were varied according to the rotamer library composed by Lovell et al. [30], which contains approximately four rotamers per side chain only, and BW1843U89 and CB3717 were docked into the resulting remodeled active site. Rotamer variation was chosen as a method for generating new protein conformations since it has the potential to bring residues into position to interact with a ligand or to remove a potential clash between a residue and ligand as well as representing a limited, low-energy subset of new conformations. We also made the simplifying assumption that all rotamers are iso-energetic. In this case, the question we asked was whether any one of these rotamer combinations would give an acceptably good score for the known ligand. Therefore we chose rotamers that were close to the conformation of the 'opposite' structure. For example, in docking CB3717 into the BW1843U89 structure, we chose rotamers for Phe 209, Ile 86 and Phe 58 that best represented the conformations of those residues in the CB3717 structure, and vice versa for docking BW1843U89 into the CB3717 structure. The structure of the enzyme with docked inhibitor was minimized with a simple conjugate gradient minimization [31].

In any genuine case, without the knowledge of the structure of a particular ligand/enzyme combination, it would be necessary to compute the entire manifold of structures based on different rotamers for each flexible residue and dock the ligands against each structure. We used a rotamer library with 153 total states for all residues. This library was sufficient to predict a correct alternative conformation for the identified three residues and the manifold of possible structures with alternative conformations is represented by only 12 structures (4 rotamers per 3 residues). Docking a database of possible inhibitors against the entire manifold is computationally manageable. In a case where finer sampling of rotamers is desired, a very comprehensive rotamer search using a large library such as that described by De Maeyer et al. [32], including 859 rotamer states, could be explored, given the limited number of residues for which alternative conformations need to be predicted.

The energy score for BW1843U89 docked into the newly remodeled PcTS receptor from the CB3717 structure, using rotamer variation, is −51.5 (Fig. 6a and Table 2), reflecting a value much closer to that of CB3717 docked into its own active site (−52.9). The energy score for CB3717 docked into the remodeled receptor from the BW1843U89 structure is −49.7 (Fig. 6b and Table 2). At this step, minimization of the structure of the enzyme with the docked inhibitor present is important for an accurate determination of the energy score since before minimization the scores for the BW1843U89 and CB3717 compounds were −43.7 and −40.6, respectively. Both new scores show a great reduction in the discrepancy of energy scores for two potent inhibitors into the same site with alternate conformations. The drastic improvement in energy scores for docking either BW1843U89 or CB3717 into the remodeled, minimized receptors represents a large step toward the solution of the rigid receptor problem.

3. Significance

It is absolutely necessary to include receptor side-chain flexibility during the search for novel inhibitors in docking and scoring experiments in order to properly model ac-

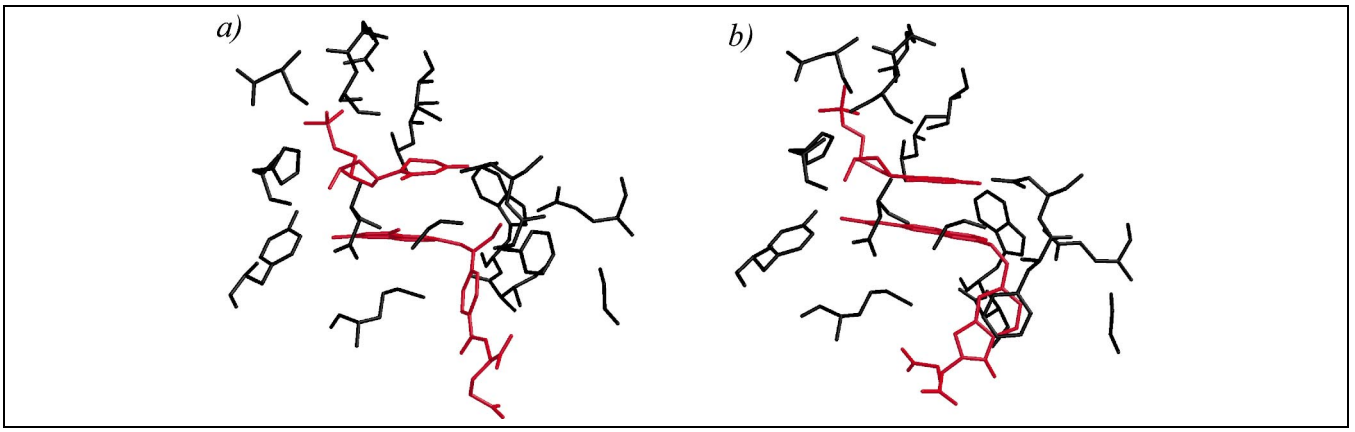

Fig. 6. Docked conformations of antifolates in the remodeled receptors. (a) CB3717 (red) docked into the altered PcTS/dUMP/BW1843U89 structure with new rotamer conformations for the three identified flexible residues in the active site. (b) BW1843U89 (red) docked into the altered PcTS/dUMP/CB3717 structure. dUMP is also depicted in red, all protein residues are shown in black.

tive sites with conformational flexibility. Two known inhibitors, BW1843U89 and CB3717, of TS, an enzyme that undergoes significant ligand-induced conformational changes, receive poor scores when computationally docked into the protein crystal structure from the complex with the other ligand, revealing the breakdown of the rigid receptor search model. With an emphasis on the role of hydrophobic residues in conformational changes, a standard seen with TS and other enzymes, it is possible to identify the principal side chains responsible for ligand-induced structural accommodation using the predictive algorithm SOFTSPOTS presented here. Varying the rotamer positions of only the side chains (generated in PLASTIC) allows the necessary active site plasticity that minimizes the discrepancy of energy scores of potent inhibitors, as compared with their ranking in experimental determinations. This algorithm likewise makes the conformational sampling tractable computationally. With the procedures described here it is not necessary to have additional structural information beyond a single inhibitor bound enzyme structure in order to identify flexible residues in an enzyme active site, apply a rotamer database to create a minimal manifold of new conformations, and dock against that manifold to dramatically improve discovery and scoring of new candidate inhibitors.

4. Materials and methods

4.1. Expression and assays

PcTS was expressed and purified as described previously [33]. Enzyme activity was assayed spectrophotometrically according to the methods of Wahba and Friedkin [34]. Inhibition assays measured the initial velocities of reactions with 1.9 μg enzyme, 80 μM dUMP, 100 μM mTHF and varying concentrations of inhibitor. Since the ratio of the total enzyme to K_i value was much greater than 1000, the data were plotted as v_o/v_i vs. $I_t/(1-v_1/v_o)$. The slope of this line yields the K_i value [35].

4.2. Crystallization

100 μM enzyme, 2 mM BW1843U89, 2.2 mM dUMP and 2.2 mM DTT were mixed and allowed to incubate for 1 h. Hanging drop crystallizations used a well solution of 14.5% PEG 8000, 0.25 M $(NH_4)_2SO_4$, 5 mM DTT, 100 mM Tris pH 7.4. Crystals grew within 5 days. The crystals belong to the space group $P2_1$ and have unit cell dimensions: $a=55$ Å, $b=66$ Å, $c=184$ Å, $\beta=90.2$.

4.3. Data collection

Data were collected at 100 K on a Mar image plate at Stanford Synchrotron Radiation Laboratory, beamline 9-1. The data were indexed, merged and scaled with the program DENZO [36] and converted to structure factors with the program, Truncate, from the CCP4 package [37].

4.4. Structure solution

The structure was solved using the AMoRe molecular replacement method [38] with a model of PcTS [15] from which all ligands were removed. Two molecules of the dimeric enzyme were located in the asymmetric unit with a starting R_{factor} of 36%. After setting aside 10% of the reflections for a free R calculation, the model was refined with 50 cycles of conjugate gradient refinement [31]. Two-fold non-crystallographic symmetry constraints were applied to the two dimers but the monomers within each dimer were structurally independent and remained so during the entire course of the refinement. Electron density for the ligands was clear from the initial maps and therefore the ligands were placed in the active site [39] and included in the refinement. The model was subjected to simulated annealing refinement, manually rebuilt from simulated annealed omit maps and refined with conjugate gradient minimizations. Water molecules with 3σ difference density (F_o-F_c) were included in the final calculations. The final model with refined values of position and temperature factors converged with the residuals $R_{free}=0.247$ and $R_{cryst}=0.216$.

4.5. Docking and scoring

Computational chemistry experiments were carried out with the program DOCK 4.0 [40] and its associated programs. The active sites of the two structures of PcTS bound to dUMP and either CB3717 or BW1843U89 formed the target receptors. All water molecules and the inhibitor present in the ternary bound structure were removed but the substrate, dUMP, was retained since dUMP forms an essential binding surface for inhibitors. Proton positions and charges were calculated for these coordinates using SYBYL 6.5 (Tripos Associates, St. Louis, MO) and an AMBER united atom representation [41]. Spheres filling the active site cavity were calculated with the program, SPHGEN. The active sites were probed with 110 070 molecules from the ACD 97.0 (MDL Inc., San Leandro, CA) representing those with more 'drug-like' characteristics (no known toxicity, molecular weight less than 600, no charges greater than 2 or less than −2). Four hundred orientations were sampled for each molecule and those that passed a steric filter, using a bump grid [23], were scored using a combination of van der Waals and electrostatic potentials as calculated by CHEMGRID [23,42].

Flexible docking experiments were carried out using the anchor-first search algorithm described in Leach et al. [43]. This algorithm divides each potential ligand at rotatable bonds and initially docks and minimizes an anchor derived from one of these segments. Remaining segments are re-attached to the anchor during the conformational search.

Acknowledgements

This work was supported by a grant to R.M.S. (CA63081). A.C.A. is supported by an NIH Postdoctoral Fellowship (GM20067). We thank D. Santi for the overexpression plasmid for *P. carinii* TS and M. Yu for assistance in purification. We also thank Dr. Irwin Kuntz for assistance with the computational chemistry experiments and Dr. Kenneth Brameld for helpful comments on the manuscript. BW1843U89 was a gift from Robert Ferone, Burroughs Wellcome, Inc.

References

[1] H. Carlson, J. McCammon, Accommodating protein flexibility in computational drug design, Mol. Pharmacol. 57 (2000) 213–218.
[2] H. Carlson, K. Masukawa, J. McCammon, Method for including the dynamic fluctuations of a protein in computer-aided drug design, J. Phys. Chem. A 103 (1999) 10213–10219.
[3] Y.-P. Pang, A. Kozikowski, Prediction of the binding sites of huperizine A in acetylcholinesterase by docking studies, J. Comput. Aided Mol. Des. 8 (1994) 669–681.
[4] R. Knegtel, I. Kuntz, C. Oshiro, Molecular docking to ensembles of protein structures, J. Mol. Biol. 266 (1997) 424–440.
[5] D. Gschwend, A. Good, I. Kuntz, Molcular docking towards drug discovery, J. Mol. Rec. 9 (1996) 175–186.
[6] V. Schnecke, C. Swanson, E. Getzoff, J. Tainer, L. Kuhn, Screening a peptidyl database for potential ligands to proteins with side-chain flexibility, Proteins 33 (1998) 74–87.
[7] L. Schaffer, G. Verkhivker, Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side chain optimization, Proteins 33 (1998) 295–310.
[8] A. Leach, Ligand docking to proteins with discrete side-chain flexibility, J. Mol. Biol. 235 (1994) 345–356.
[9] N. Nakajima, J. Higo, A. Kidera, H. Nakamura, Flexible docking of a ligand peptide to a receptor protein by multicanonical molecular dynamics simulation, Chem. Phys. Lett. 278 (1997) 297–301.
[10] A. Jackman, G. Taylor, W. Gibson, R. Kimbell, M. Brown, A. Calvert, I. Judson, L. Hughes, ICI D1694, a quinazoline antifolate thymidylate synthase inhibitor that is a potent inhibitor of L1210 tumor cell growth in vitro and in vivo: a new agent for clinical study, Cancer Res. 51 (1991) 5579–5586.
[11] T. Jones, A. Calvert, A. Jackman, S. Brown, M. Jones, K. Harrap, A potent antitumor quinazoline inhibitor of thymidylate synthetase: synthesis, biological properties, and therapeutic results in mice, Eur. J. Cancer 17 (1981) 11–19.
[12] A. Gangjee, F. Mavandadi, R. Kisliuk, J. McGuire, S. Queener, 2-Amino-4-oxo-5-substituted-pyrrole[2,3-d]pyrimidines as non-classical antifolate inhibitors of thymidylate synthase, J. Med. Chem. 39 (1996) 4563–4568.
[13] M. Varney, G. Marzoni, C. Palmer, J. Deal, S. Webber, K. Welsh, R. Bacquet, C. Bartlett, C. Morse, C. Booth, S. Herrmann, E. Howland, R. Ward, J. White, Crystal structure-based drug design and synthesis of benz[cd]indole-containing inhibitors of thymidylate synthase, J. Med. Chem. 35 (1992) 663–676.
[14] I. Duch, S. Banks, I. Dev, S. Dickerson, R. Ferone, L. Heath, J. Humphreys, V. Knick, W. Pendergast, S. Singer, G. Smith, K. Waters, H. Wilson, Biochemical and cellular pharmacology of 1843U89, a novel benzoquinazoline inhibitor of thymidylate synthase, Cancer Res. 53 (1993) 810–818.
[15] A. Anderson, K. Perry, D. Freymann, R. Stroud, The crystal structure of thymidylate synthase from *Pneumocystis carinii* reveals a fungal insert important for drug design, J. Mol. Biol. 297 (2000) 645–657.
[16] M. Bretner, T. Kulikowski, J. Dzik, M. Balinska, W. Rode, D. Shugar, 2-Thio derivatives of dUrd and 5-fluoro-dUrd and their 5′-monophosphates: synthesis, interaction with tumor thymidylate synthase, and in vitro anti-tumor activity, J. Med. Chem. 36 (1993) 3611–3617.
[17] R. Jackson, A. Jackman, H. Calvert, Biochemical effects of a quinazoline inhibitor of thymidylate synthase, CB3717, on human lymphoblastoid cells, Biochem. Pharmacol. 32 (1983) 3783–3790.
[18] W. Pendergast, S. Dickerson, J. Johnson, I. Dev, R. Ferone, D. Duch, G. Smith, Benzoquinazoline inhibitors of thymidylate synthase: enzyme inhibitory activity and cytotoxicity of some sulfonamidobenzolylglutamate and related derivatives, J. Med. Chem. 36 (1993) 3464–3471.
[19] B. Shoichet, R. Stroud, D. Santi, I. Kuntz, K. Perry, Structure-based discovery of inhibitors of thymidylate synthase, Science 259 (1993) 1445–1450.
[20] T. Stout, D. Tondi, M. Rinaldi, D. Barlocco, P. Pecorari, D. Santi, I. Kuntz, R. Stroud, B. Shoichet, P. Costi, Structure-based design of inhibitors specific for bacterial thymidylate synthase, Biochemistry 38 (1999) 1607–1617.
[21] W. Montfort, A. Weichsel, Thymidylate synthase: structure, inhibition, and strained conformations during catalysis, Pharmacol. Ther. 76 (1997) 29–43.
[22] A. Anderson, R. O'Neil, W. DeLano, R. Stroud, A structural mechanism for half-the-sites reactivity in an enzyme, thymidylate synthase, involves a relay of changes between subunits, Biochemistry 38 (1999) 13829–13836.
[23] E. Meng, B. Shoichet, I. Kuntz, Automated docking with grid-based energy evaluation, J. Comp. Chem. 13 (1992) 505–524.
[24] W. Montfort, K. Perry, E. Fauman, J. Finer-Moore, G. Maley, L.

Hardy, F. Maley, R. Stroud, Structure, multiple site binding, and segmental accommodation in thymidylate synthase on binding dump and an anti-folate, Biochemistry 29 (1990) 6964–6977.

[25] S. Zographos, N. Oikonomakos, K. Tsitsanou, D. Leonidas, E. Chrysina, V. Skamnaki, H. Bischoff, S. Goldmann, K. Watson, L. Johnson, The structure of glycogen phosphorylase b with an alkyldihydropyridine-dicarboxylic acid compound, a novel and potent inhibitor, Structure 5 (1997) 1413–1425.

[26] D. Wilson, I. Tarle, J. Petrash, F. Quiocho, Refined 1.8 Å structure of human aldose reductase complexed with the potent inhibitor zopolrestat, Proc. Natl. Acad. Sci. USA 90 (1993) 9847–9851.

[27] A. Urzhumtsev, F. Tete-Favier, A. Mitschler, J. Barbanton, P. Barth, L. Urzhumtseva, J. Biellman, A. Podjarny, D. Moras, A 'specificity' pocket inferred from the crystal structures of the complexes of aldose reductase with the pharmaceutically important inhibitors tolrestat and sorbinol, Structure 5 (1997) 601–612.

[28] J. Ren, R. Esnouf, E. Garman, D. Somers, C. Ross, I. Kirby, J. Keeling, G. Dardy, Y. Jones, D. Stuart, D. Stammers, High resolution structures of HIV-1 RT from four RT-inhibitor complexes, Nat. Struct. Biol. 2 (1995) 293–302.

[29] A. Davis, S. Teague, Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis, Angew. Chem. 38 (1999) 736–749.

[30] S. Lovell, J. Word, J. Richardson, D. Richardson, The penultimate rotamer library, Proteins 40 (2000) 389–408.

[31] A. Brunger, Crystallography and NMR System: a new software suite for macromolecular structure determination, Acta Crystallogr. D 54 (1998) 905–921.

[32] M. De Maeyer, J. Desmet, I. Lasteres, All in one: a highly detailed rotamer library improves both accuracy and speed in the modeling of side chains by dead-end elimination, Fold. Des. 2 (1997) 53–66.

[33] D. Santi, U. Edman, S. Minkin, P. Greene, Purification and characterization of recombinant *Pneumocystis carinii* thymidylate synthase, Protein Express. Purif. 2 (1991) 350–354.

[34] A. Wahba, M. Friedkin, Direct spectrophotometric evidence for the oxidation of tetrahydrofolate during the enzymatic synthesis of thymidylate, J. Biol. Chem. 236 (1961) PC11–PC12.

[35] P. Henderson, A linear equation that describes the steady-state kinetics of enzymes and subcellular particles interacting with tightly bound inhibitors, Biochem. J. 127 (1972) 321–333.

[36] Z. Otwinowski, DENZO 1-56-62, in: SERC, Daresbury Laboratory, Warrington, 1993.

[37] Collaborative Computational Project Number 4, The CCP4 suite of programs for protein crystallography, Acta Crystallogr. D 50 (1994) 760–763.

[38] J. Navaza, AMoRe: an automated package for molecular replacement, Acta Crystallogr. A 50 (1994) 157–163.

[39] T. Jones, J.-Y. Zou, S. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models, Acta Crystallogr. A 47 (1991) 110–119.

[40] T. Ewing, I. Kuntz, Critical evaluation of search algorithms used in automated molecular docking, J. Comp. Chem. 18 (1997) 1175–1189.

[41] S. Weiner, P. Kollman, D. Case, U. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, A new force field for molecular mechanics simulation of nucleic acids and proteins, J. Am. Chem. Soc. 106 (1984) 765–784.

[42] B. Shoichet, D. Bodian, I. Kuntz, Molecular docking using shape descriptors, J. Comput. Chem. 13 (1992) 380–397.

[43] A. Leach, I. Kuntz, Conformational analysis of flexible ligands in macromolecular receptor sites, J. Comput. Chem. 13 (1992) 730–748.