Journal of Computer-Aided Molecular Design **17:** 525–536, 2003.

Automated docking of ligands to an artificial active site: augmenting crystallographic analysis with computer modeling

Robin J. Rosenfeld[a,*], David S. Goodsell[a], Rabi A. Musah[b], Garrett M. Morris[a], David B. Goodin[a] & Arthur J. Olson[a,*]

[a]*The Department of Molecular Biology and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, CA 92037, U.S.A.;* [b]*Chemistry Department, State University of New York, Albany, New York, U.S.A.*

Received 18 March 2003; accepted in revised form 5 August 2003

Key words: AutoDock, automated docking, binding free energy, cytochrome *c* peroxidase, protein engineering, protein-ligand interaction

Summary

The W191G cavity of cytochrome *c* peroxidase is useful as a model system for introducing small molecule oxidation in an artificially created cavity. A set of small, cyclic, organic cations was previously shown to bind in the buried, solvent-filled pocket created by the W191G mutation. We docked these ligands and a set of non-binders in the W191G cavity using AutoDock 3.0. For the ligands, we compared docking predictions with experimentally determined binding energies and X-ray crystal structure complexes. For the ligands, predicted binding energies differed from measured values by ± 0.8 kcal/mol. For most ligands, the docking simulation clearly predicted a single binding mode that matched the crystallographic binding mode within 1.0 Å RMSD. For 2 ligands, where the docking procedure yielded an ambiguous result, solutions matching the crystallographic result could be obtained by including an additional crystallographically observed water molecule in the protein model. For the remaining 2 ligands, docking indicated multiple binding modes, consistent with the original electron density, suggesting disordered binding of these ligands. Visual inspection of the atomic affinity grid maps used in docking calculations revealed two patches of high affinity for hydrogen bond donating groups. Multiple solutions are predicted as these two sites compete for polar hydrogens in the ligand during the docking simulation. Ligands could be distinguished, to some extent, from non-binders using a combination of two trends: predicted binding energy and level of clustering. In summary, AutoDock 3.0 appears to be useful in predicting key structural and energetic features of ligand binding in the W191G cavity.

Introduction

Cytochrome *c* peroxidase is a heme enzyme that catalyzes the oxidation of ferricytochrome c in yeast mitochondria [1, 2]. In contrast to most heme enzymes, cytochrome *c* peroxidase has the unusual property that the second oxidizing potential is stored as a cation radical on the Trp191 side chain [3–6]. The protein environment surrounding Trp191 is thought to have a role in stabilizing the cation radical. Therefore, the Trp191 site is an excellent target for engineering small molecule oxidation [7, 8]. The W191G mutation creates a structurally intact buried polar cavity that contains five well-ordered structural waters and a potassium ion in place of the Trp191 side chain [9, 10]. The W191G cavity binds small cationic ligands and can catalyze oxidation of small substrates [7]. A series of small cationic thiazoles, imidazoles, and pyridines have been shown to bind in this cavity (Figure 1) [7, 10–12]. In addition to the ligands, a set of non-binders was discovered during the experimental assessment of W191G binding specificity. Non-binders included

*To whom correspondence should be addressed. E-mail: robin@scripps.edu, olson@scripps.edu

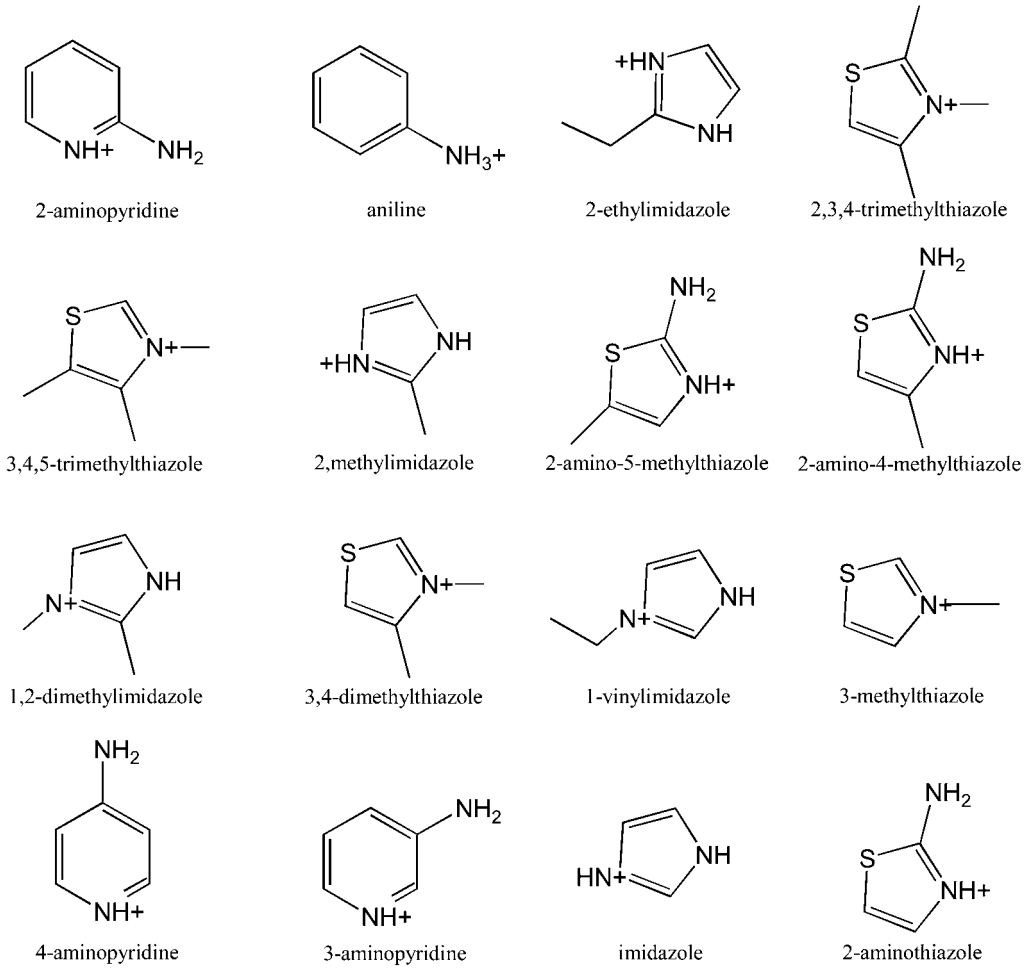

Figure 1. The set of ligands docked to the W191G cavity in this study. These compounds were shown in a previous study to bind in the engineered W191G cavity [7, 11, 43].

small, cyclic, and/or cationic compounds, and thus were similar to binding ligands (Figure 2).

In the present study, a set of compounds (16 ligands, 17 non-binders) whose interaction with the W191G cavity was characterized experimentally [11] was used for docking analysis. The dissociation free energies of the ligands range from −4.4 to −6.9 kcal/mol, as measured by following changes in the heme Soret absorbance [11]. The binding orientation of each ligand within the cavity was determined using X-ray crystallography (resolution 1.9 to 2.2 Å) [11]. Among the ligand-bound structures, there are no significant deviations in protein structure [10, 11]. In fact, the only differences amongst ligand-bound structures are the orientations of the ligands and the number of waters observed in the cavity. Specifically, cavity water HOH-308 is observed in all structures with and without ligand bound, whereas cavity water HOH-401 is displaced by all but three ligands.

W191G presents a rigid protein template that is experimentally well-characterized [7, 8, 10, 11]. Despite the rigid template, visual inspection of the unliganded crystal structure is not sufficient for ligand design or predicting ligand-binding orientations. These features make the W191G cavity an excellent model system for exploring the limits of ligand binding specificity using automated docking. Automated docking is a useful tool for predicting ligand binding orientations, binding affinities [13], and the role of solvent in protein-ligand interactions [14]. Use of well-characterized model systems where the results are known can aid in comparative molecular field analysis [15, 16]. We chose the AutoDock 3.0 program to model ligand binding within the W191G cavity because of the recently

implemented empirical free energy force field, improvements in searching efficiency, and the advantage of using one set of precomputed atomic affinity maps to dock numerous ligands [17]. AutoDock 3.0 uses a Lamarckian genetic algorithm to find the most favorable ligand binding orientations. The force field, represented by atomic affinity grid maps computed prior to ligand docking, is a linear combination of van der Waals, hydrogen bonding, hydrophobic desolvation, electrostatics, and torsional free energy empirically determined to reproduce ligand-protein binding free energies [17]. AutoDock has been used to model diverse protein-ligand interactions including ligand binding to antibodies [18], diacylglycerol binding to protein kinase C [19], cofactor binding to nitric oxide synthase [20], factor-Xa-inhibitor complexes [21], β-amylase substrate binding [22], cannanibinoid receptor interactions [23], and monosaccharide binding to glucoamylase [24]. Here we demonstrate the utility of AutoDock for analyzing ligand binding in a cavity with broad specificity.

Methods and materials

Preparation of protein coordinate files and grid map calculations

We constructed two models of the W191G cavity, protein model A and protein model B, to reflect the two states of solvent occupancy observed in the ligand-bound crystal structures. Protein model A contained cavity water HOH-308, which was observed in all 16 ligand-bound crystal structures. HOH-308 is tightly hydrogen bonded to the amino group of Gly191, the carbonyl of Ala176, and the carbonyl of Gly178 in the ligand-bound and ligand-free states of W191G, and thus treating it as part of the protein structure was warranted. Protein model B contained HOH-308 and an additional cavity water, HOH-401, which was present in three of the ligand-bound crystal structures.

Polar hydrogens, partial charges, and solvation parameters were added to the W191G protein coordinate files in preparation for computing grid maps. Polar hydrogens and Kollman united-atom charges [25, 26] were added to the protein using SYBYL version 6.1 (Sybyl, Tripos Associated, Inc., St. Louis, MO). Amber potentials were used for the heme [25, 27]. Solvation parameters, based on the Stouten model [28], were added to the protein file in accordance with the AutoDock 3.0 force field [29].

Atomic affinity grid maps were computed for each atom type in the ligand set, as well as an electrostatics grid map, from the protein models using AutoGrid 3.0 and the standard AutoDock 3.0 force field. The cubic grid maps were centered on the W191G cavity and had dimensions of $9 \times 9 \times 9$ Å with 0.15 Å spacing between grid points. Protonated nitrogen atoms in W191G were modeled as non-hydrogen bond acceptors and assigned the standard 6-12 Lennard-Jones parameters, rather than 10–12 hydrogen-bond parameters. Parameters for the heme iron, which are not standard to AutoDock 3.0, were taken from Amber [25, 27] and the well depths multiplied by the AutoDock empirical free energy force field scaling factor (0.1485) [17].

Preparing ligands and non-binding compounds for docking

The coordinates of each ligand bound in the cavity were determined crystallographically as described previously and are deposited in the Protein Data Bank [9–11]. Non-binding compounds were built and energy minimized using SYBYL version 6.1. The chemical structures of the ligands and non-binding compounds are shown in Figure 1 and Figure 2, respectively. Density functional charges, computed for each compound, as described elsewhere [30] with Gaussian94 [31] and the 6-31 and TZ2P basis sets [32], were added to the coordinate file. The AutoTors application of the AutoDock program suite was used to merge nonpolar hydrogens and initialize torsions for each ligand [33].

Docking compounds using AutoDock 3.0

Ligands and non-binding compounds were first docked to the protein model containing only cavity water HOH-308 (protein model A), because this water was observed in all ligand-bound crystal structures. Two ligands, 2-aminothiazole and 3-aminopyridine, were later docked to the protein model containing HOH-308 and HOH-401 (protein model B). A Lamarckian genetic algorithm was used to search each ligand's configuration space for low energy binding orientations. Lamarckian genetic algorithms are inspired by Jean Baptiste de Lamarck's theory that phenotype changes that occur during an individual's lifetime can be passed on as genetic changes to its offspring. In AutoDock, the Lamarckian genetic algorithm is implemented by combining a global genetic

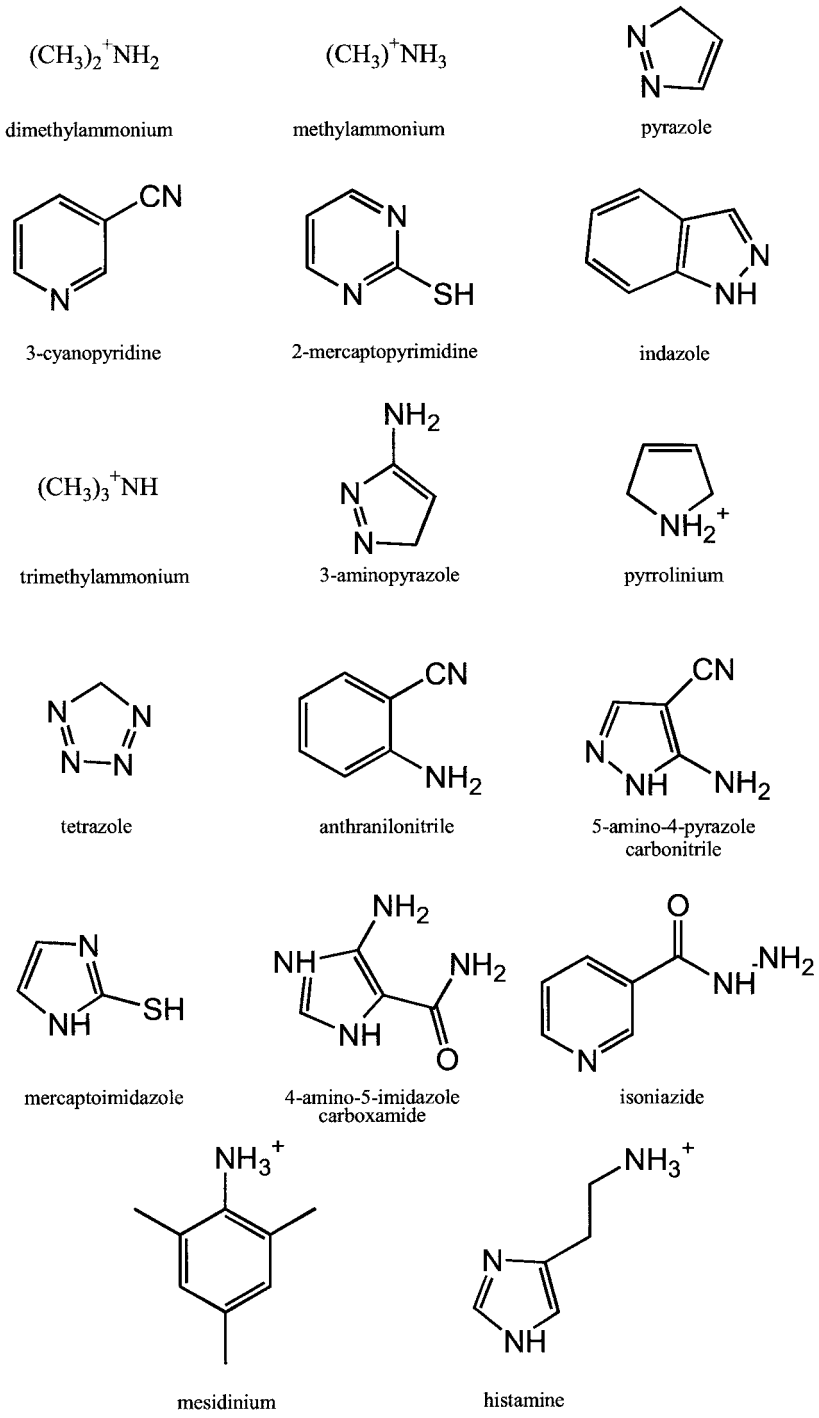

Figure 2. The set of non-binders docked to the W191G cavity in this study. These compounds were classified as 'non-binders' based on previous experimental assessment of W191G ligand-binding specificity [11].

algorithm with local minimization. For local minimization, AutoDock uses the Solis & Wets adaptive search algorithm. During local minimization, the step size increases if the last step improved the binding energy and decreases if the last step resulted in a worse binding energy. Each docking trial was initiated with a randomly generated population of 50 binding orientations and completed after 1.5 million energy evaluations had been performed. The point mutation and crossover rates were set at 2% and 80%, respectively. The probability of performing a local search on an individual was set at 6%. Due to the stochastic nature of genetic search algorithms, 100 trials of each docking were performed. The resulting docked conformations were clustered together using an RMSD tolerance of 0.5 Å. The output files contained the final predicted conformations, orientations, positions and the RMSD from the bound crystal structure, the docked energy and the estimated free energy of binding for each cluster and each individual docking. The free energy was computed as the sum of the docked energy, the torsional energy, and the intramolecular interaction energy for the ligand [29]. AutoDock was run on SGI Power Challenges with R8000 IP21 processor chips operating at 90 MHz using IRIX Release 6.2. The docked conformations were visualized in the AVS graphics program, version 5.02 [34] (Advanced Visual Systems, Inc., Waltham, MA).

Results and discussion

In order to study ligand recognition in the engineered W191G cavity, we docked 33 compounds to the cavity using AutoDock 3.0 [17]. The 16 ligands and 17 non-binders are shown in Figures 1 and 2, respectively [11]. For most of the ligands, docking unambiguously reproduced the crystallographic binding orientation within the protein cavity. AutoDock predicted the binding affinity of all ligands within the estimated error of the force field (2.2 kcal/mol). Initially, all compounds were docked computationally to the structure of the unliganded protein, with an invariant water molecule (HOH-308) included as part of the protein structure (protein model A). For each compound, 100 docking experiments were initiated with randomized populations and solutions for individual runs were clustered if their final docked positions were within a tolerance of 0.5 Å RMSD.

Successful prediction of single ligand binding modes

Convergence of dockings into a common conformational cluster was a major indicator that the cluster matched the crystallographic ligand-binding orientation (Table 1). In 12 of 16 cases, the results were clear and unambiguous (e.g. 2,3,4-trimethylthiazole, Figure 3), often docking to one single converged cluster. For all 12 of these ligands, at least 90% of the docked conformations were within a single dominant cluster. These highly populated clusters corresponded closely to the crystallographically determined positions, having RMS differences from the crystal structure ranging from 0.2 Å to 0.8 Å.

Effect of crystallographically observed waters in docking

Four of the known ligands (2-aminothiazole, 3-aminopyridine, 4-aminopyridine and imidazole) gave more ambiguous results, with docked solutions clustering into several (3 to 5) different conformations with similar energies (Tables 2 and 3). We hypothesized that these multiple clusters have two possible explanations: (i) different conformations may result from variations in cavity solvent in the ligand bound structures that were not accounted for in the AutoDock simulation; or (ii) the different conformations may reflect actual disorder present in the complex. For 2-aminothiazole and 3-aminopyridine, case (i) clearly applies, as the ligand-bound crystal structures of these complexes showed density for an additional cavity water molecule (HOH-401). Docking results for these two ligands became consistent with the crystallographic structures when HOH-401 was included as part of the protein structure (protein model B) (Table 2, Figure 4). The complex with aniline also showed strong electron density for HOH-401, but in this case, AutoDock predicted the proper conformation in all 100 dockings because the size of the aniline excluded it from docking in a position such that the amino group could satisfy the hydrogen bonds made by HOH-401.

Water plays an important role in many protein-ligand interactions [35–39]. In some cases, water mediated ligand recognition can be modeled by predicting conserved waters in a ligand binding site [40], predicting potential water positions [36], hydrating the ligand [35], or using a weighted average of multiple target structures [41]. Computational tools, such as Consolv [40], predict conserved waters in a binding site based upon hydrogen-bond analysis,

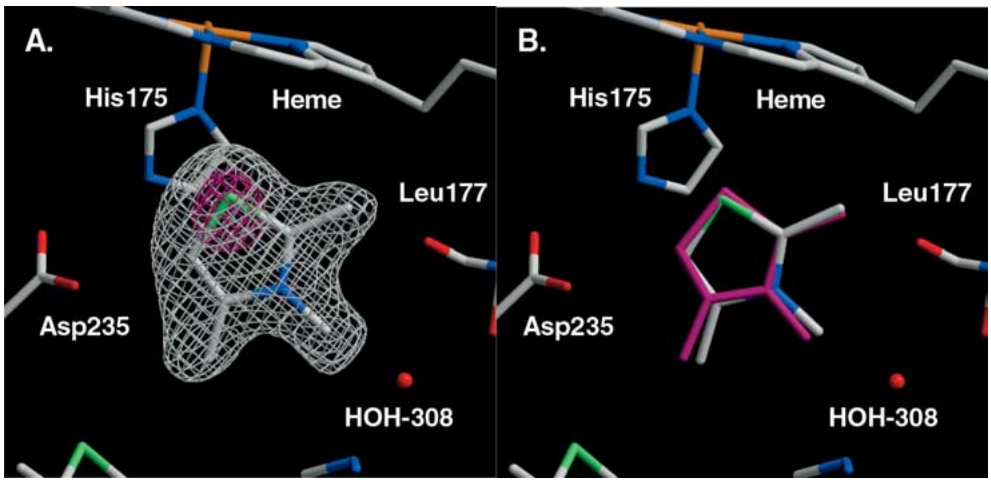

Figure 3. Example of single binding mode successfully predicted by AutoDock. (A) The crystallographic binding orientation of 2,3,4-trimethylthiazole fit into F_o-F_c omit electron density (white contours 3σ, magenta contours 10σ). (B) Comparison of the predicted (magenta) and crystallographic (white carbon atoms) orientations of 2,3,4-trimethylthizole in the W191G cavity. A single cluster 0.5 Å RMSD away from the assigned crystal orientation was predicted by 100 docking trials.

Table 1. Clustering analysis of ligands with single binding modes. For each ligand, the most populated cluster predicted by AutoDock is within 0.8 Å RMSD from the crystallographically determined orientation. The maximum difference between predicted and experimental binding energies is 1.4 kcal/mol. The average absolute difference between predicted and measured binding affinities is ± 0.8 kcal/mol.

Ligand	Cluster	Cluster population	Binding free energy (kcal/mol)			RMSD (Å)	PDB code [11, 43]
			Predicted	Measured [11]	Predicted – measured		
Aniline	1	100	−5.8	−6.2	0.4	0.5	1AEE
2-Aminopyridine	1	100	−5.8	−5.7	−0.1	0.3	1AEO
2,3,4-Trimethylthiazole	1	100	−5.0	−4.4	−1.4	0.5	1AC4
3,4,5-Trimethylthiazole	1	100	−5.0	−5.3	0.3	0.2	1AC8
2-Methylimidazole	1	100	−4.9	−5.7	0.8	0.7	1AEU
2-Ethylimidazole	1	100	−5.4	−4.2	−1.2	0.3	1AEQ
3-Methylthiazole	1	99	−4.5	−5.4	0.9	0.8	1AEB
	2	1	−4.5			2.5	
1,2-Dimethylimidazole	1	99	−5.3	−6.0	0.7	0.4	1CMP
	2	1	−5.3			1.5	
2-Amino-5-methylthiazole	1	99	−5.6	−6.9	1.3	0.3	1AEN
	2	1	−5.6			1.4	
2-Amino-4-methylthiazole	1	4	−5.3	−4.9	−0.4	3.1	1AEH
	2	96	−5.2			0.6	
3,4-Dimethylthiazole	1	90	−5.0	−5.7	0.7	0.2	1AED
	2	10	−5.0			2.4	
1-Vinylimidazole	1	90	−4.5	−5.3	0.8	0.6	1AEJ
	2	2	−4.5			0.9	
	3	8	−4.5			2.5	

crystallographic B-factors, and solvent-accessible surface area of water molecules in a protein crystal structure. Similar factors were used to analyze conserved waters in major histocompatibility complex structures [37]. In the W191G cavity all five crystallographic waters have low B-factors and are in a tight hydrogen-bonding network, and thus are all predicted to be conserved. Therefore, our ability to

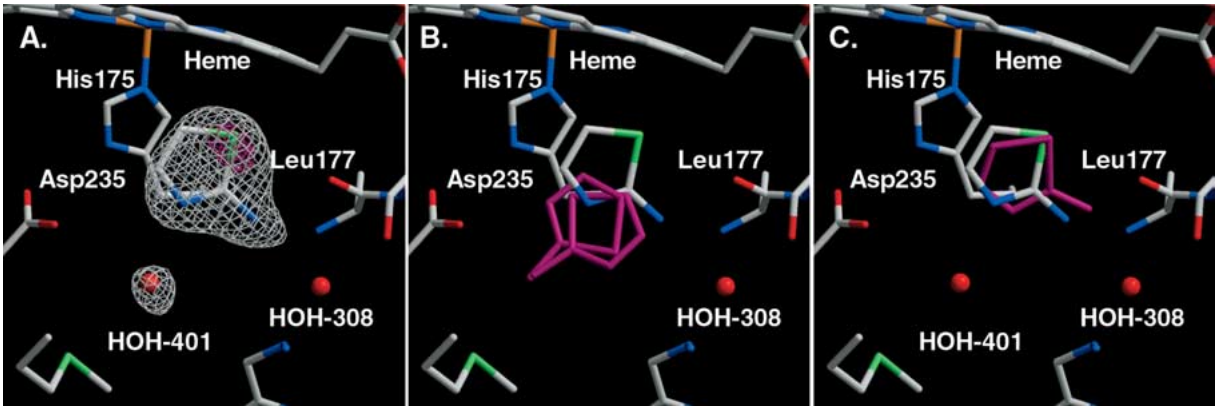

Figure 4. Structural waters influence docking results. (A) Crystal structure of 2-aminothiazole fit into the F_o-F_c omit electron density in the W191G cavity (white contours 3σ, red contours 6σ). (B) Docking 2-aminothiazole to W191G protein model A, without explicitly including HOH-401 results in multiple clusters in which the amino group docks to the position occupied by HOH-401. The top two clusters (magenta) are superimposed onto the crystallographically observed orientation (white carbon atoms). (C) 2-Aminothiazole docked to protein model B, a model that explicitly includes structural water HOH-401. The docked prediction (magenta) matches the crystallographically observed orientation of 2-aminothiazole. One cluster was predicted that contained 100 dockings with an RMSD of 0.8 Å.

include conserved water molecules based on superposition of ligand-bound structures was useful. Ideally, methods that account for changes in solvent molecules will be implemented in future versions of AutoDock. For instance, it should be possible to adapt AutoDock 4.0, which is currently being developed to model protein side-chain flexibility [42], to model solvent flexibility (http://www.scripps.edu/pub/olson-web/doc/autodock).

Multiple binding modes predicted by multiple clusters of solutions

The remaining two multi-conformational docking results could not be explained by changes in solvent occupation, and we postulated that these multiple clusters reflected actual disordered ligand binding within the cavity. Three clusters are found for 4-aminopyridine (Table 3, Figure 5), which align the two hydrogen-bonding groups of the ligand towards two different sites within the cavity. In the top 2 clusters, the amino group is oriented to form a hydrogen bond with Asp235, and the ring nitrogen is directed towards the carbonyl of Leu177. In cluster three (as sorted by increasing energy), the ligand is oriented so that the amino group hydrogen bonds with the Leu177 carbonyl and the ring nitrogen hydrogen bonds with Asp235. In this case, re-examination of the experimental electron density maps revealed that disorder is evident in this structure. A similar result is observed with the imidazole-bound complex. Imidazole and 4-aminopyridine are roughly symmetrical molecules, with two similar hydrogen-bonding groups arranged on opposite sides. Multiple docked orientations are found as two sites in the cavity vie for the pseudo-symmetrical hydrogen bonding groups. In this case, the multiple orientations could not be verified in the experimental structure due to the symmetry of the ligand.

These docking results, when analyzed together with the electron density maps, suggest that AutoDock simulations may be capable of predicting multiple binding orientations. Similar results have been reported from studies using AutoDock to analyze inhibitors binding to nitric oxide synthase [42]. Therefore, AutoDock may have an important unanticipated utility: the ability to identify and validate multiple binding orientations. AutoDock was not trained to recognize multiple energetically similar binding modes *per se*, but has this capability as a consequence of modeling the binding landscape using its empirical free energy force field. In the nitric oxide synthase study, dual orientations predicted by AutoDock were used to support crystallographic assignment of a dual binding mode for a nitroindazole-class inhibitor [42]. Understanding the dual binding mode was important in developing a general model for molecular recognition of this class of compounds in the nitric oxide synthase active site. In both the nitric oxide synthase study and in our study of W191G ligands, the prediction of multiple binding orientations by AutoDock augmented the crystallographic analysis and provided new understanding of the ligand binding sites.

Table 2. Results of docking ligands that hydrogen bond with HOH-401 to protein models with and without HOH-401 explicitly included.

Ligand	Protein model A[a]				Protein model B[b]				Measured binding energy [7, 11] (kcal/mol)	Predicted – measured binding energy	PDB [7, 11]
	Cluster	Cluster population	Predicted binding energy (kcal/mol)	RMSD (Å)	Cluster	Cluster population	Predicted binding energy (kcal/mol)	RMSD (Å)			
3-Aminopyridine	1	34	−5.4	1.7	1	100	−5.3	0.6	−5.7	0.4	1AEF
	2	26	−5.4	1.3							
	3	2	−5.3	1.7							
	4	5	−5.2	1.5							
	5	33	−5.0	0.6							
2-Aminothiazole	1	64	−5.1	2.3	1	100	−6.7	0.8	−0.6	−0.7	1AEV
	2	28	−4.8	2.3							
	3	3	−4.5	1.7							
	4	4	−4.4	2.0							
	5	1	−4.4	2.4							

[a]Protein model A explicitly includes cavity water HOH-308 only.
[b]Protein model B explicitly includes both cavity waters HOH-308 and HOH-401.

Table 3. AutoDock predicts multiple clusters for ligands that have multiple binding modes. These two ligands have symmetrical hydrogen bond donating groups that dock interchangeably towards hydrogen bond acceptors at opposite ends of the W191G cavity. Electron density supports multiple binding modes for these ligands. The RMSD is calculated using the crystallographic orientation that was initially fit based on electron density and analysis of hydrogen bonding interactions.

Ligand	Cluster	Cluster population	Binding free energy (kcal/mol)			RMSD (Å)	PDB code [11]
			Predicted	Measured [11]	Predicted – measured		
Imidazole	1	55	−4.0	−5.7	1.7	1.0	1AES
	2	35	−4.0			2.3	
	3	3	−3.9			0.7	
	4	7	−3.9			1.8	
4-Aminopyridine	1	21	−5.4	−5.9	0.5	1.2	1AEG
	2	77	−5.2			0.3	
	3	2	−5.2			1.0	

Calculated binding energies agree with experimentally measured values

Accurate prediction of binding affinities is an important goal of automated docking and is particularly relevant to rational drug design. For this set of 16 ligands, the average difference between the predicted binding energy of the most populated cluster and measured binding affinity is ±0.8 kcal/mol (Tables 1–3). This is less than the reported residual standard error for the AutoDock 3.0 force field (2.2 kcal/mol), indicating good overall agreement of the results with the previously calibrated force field. The range in binding affinity of these ligands is only 2.5 kcal/mol, thus the binding energies are expected to be roughly equivalent within the AutoDock force field. In fact, the average binding energy computed by AutoDock (5.1 ± 0.5 kcal/mol) was remarkably similar to the average measured binding energy (5.6 ± 0.7 kcal/mol) for these ligands.

A rank ordering based on predicted binding energy is not realistic for this set of compounds because of the narrow range of binding affinities and the uncertainty in predicted values. Despite the low correlation, the strongest binding ligand in this series, according to in vitro experiments, 2-amino-5-methylthiazole, was predicted to be one of the strongest binders by docking. Only 2-aminopyridine and aniline were predicted to have a higher absolute affinity and were predicted

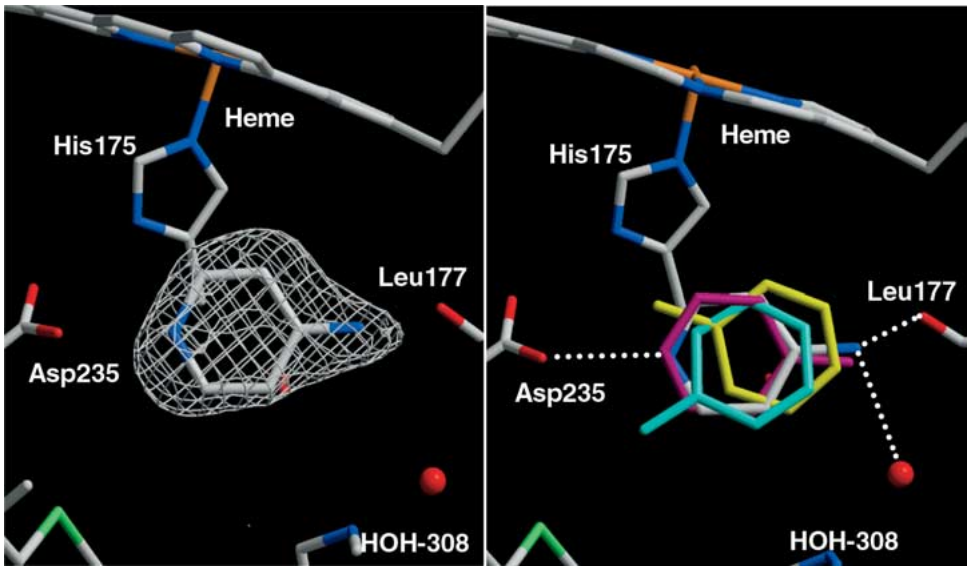

Figure 5. Multiple conformations predicted by AutoDock. (Left) A single crystallographic binding orientation was assigned for 4-aminopyridine, however, the elongated shape of the electron density suggests additional binding modes. (Right) Three clusters of binding orientations were predicted for 4-aminopyridine (cyan, magenta, yellow, in order of least to highest predicted energy), which may account for the observed electron density. The three predicted orientations for 4-aminopyridine satisfy hydrogen bonds to Asp235 and Leu177.

to bind only 0.2 kcal/mol more tightly. While AutoDock cannot be used to rank this particular series of chemically similar ligands by binding affinity, the results show promise that such a rank ordering might be possible given a larger spread in binding affinities. These results therefore support the utility of AutoDock being useful in distinguishing between ligand dissociation constants that differ by about two orders of magnitude (roughly equivalent to 1–2 kcal/mol of binding energy).

Binding and non-binding compounds can be distinguished by predicted energy and clustering

Predicting which compounds from a set of potential ligands will bind in an active site is one of the preeminent challenges in computer aided drug discovery. In addition to ligand binding data for W191G, we had an available database of 17 compounds for which specific binding to the W191G cavity was not detected experimentally (Figure 2). The non-binders are compounds that were designed and screened for binding during the initial efforts to assess the specificity of W191G, and therefore resemble the binders. The non-binders included compounds that are small, planar, positively charged and/or aromatic, and thus are not precluded from binding simply based on size and shape. To see if AutoDock could distinguish between binders and non-binders, we docked both sets to W191G protein model A and compared the results.

We found that two outputs from docking help to discriminate W191G binders from non-binders: (i) small number of clusters; and (ii) low predicted binding energy (Table 4). Of these two trends, clustering was the stronger indicator of ligand binding. Most of the W191G binders dock to only 1–2 converged clusters, whereas most non-binders docked into 5 or more clusters. Even binders with multiple orientations and those that bind protein model B dock to protein model A in fewer clusters (3–5) than most non-binders. Second, in contrast to most of the non-binders, most binders have a predicted binding energy less than -4.5 kcal/mol. Neither of the criteria alone is sufficient, but the combination of clustering and predicted energy appears to discriminate binders from non-binders. Note that classification of an individual compound as a binder or non-binder may be hindered by the relative uncertainty in predicted binding energy. However, overall there are significant statistical differences in docking results for the two sets of compounds ($p < 0.001$). Our results suggest that the AutoDock force field is on the cusp of discriminating weak binders from similar compounds that do not bind in the cavity. The ability to identify binders may be improved by empirically fitting the force field for a specific case of interest. Also, a test set containing binders with

Table 4. Results of docking binders and non-binders to W191G (protein model A). The overall docking statistics for binders and non-binders differed significantly, both in terms of clustering ($p < 0.001$) and predicted binding energy ($p < 0.001$).

Compound	Number of clusters (100 dockings)	Avg. predicted energy (kcal/mol)	Clusters + docked energy
BINDERS			
2-Aminopyridine	1	−5.8	−4.8
Aniline*	1	−5.8	−4.8
2-Ethylimidazole	1	−5.4	−4.4
2,3,4-Trimethylthiazole	1	−5.0	−4.0
3,4,5-Trimethylthiazole	1	−5.0	−4.0
2-Methylimidazole	1	−4.9	−3.9
2-Amino-5-methylthiazole	2	−5.6	−3.6
2-Amino-4-methylthiazole	2	−5.4	−3.4
1,2-Dimethylimidazole	2	−5.3	−3.3
3,4-Dimethylthiazole	2	−5.0	−3.0
1-Vinylimidazole	2	−4.5	−2.5
3-Methylthiazole	2	−4.5	−2.5
4-Aminopyridine†	3	−5.2	−2.2
3-Aminopyridine*	5	−5.3	−0.3
Imidazole†	4	−4.0	0
2-Aminothiazole*	5	−5.0	0
Average	2 ± 1	-5.1 ± 0.5	-3 ± 2
Median	2	−5.1	−3.3
NON-BINDERS			
Dimethylammonium	3	−3.0	0
Methylammonium	3	−2.6	0.4
Pyrazole	4	−3.2	0.8
3-Cyanopyridine	5	−4.2	0.8
2-Mercaptopyrimidine	6	−3.9	2.1
Indazole	7	−4.6	2.4
Trimethylammonium	6	−3.1	2.9
3-Aminopyrazole	7	−4.0	3.0
Pyralinium	8	−4.9	3.1
Tetrazole	6	−2.3	3.7
Anthranitrile	9	−4.1	4.9
5-Amino-4-pyrazole-carbonitrile	10	−4.0	6.0
Mercaptoimidazole	11	−3.5	7.5
4-Amino-5-imidazole-carboxamide	13	−4.1	8.9
Isoniazide	12	−1.3	10.7
Mesidinium	12	−0.7	11.3
Histamine	28	−5.7	22.3
Average[a]	9 ± 6	-3 ± 1	$+5 \pm 6$
Average[b]	8 ± 3	-3 ± 1	$+4 \pm 4$
Median[a]	7	−3.7	+3.6

[a]Calculated using all non-binders.
[b]Calculated without histamine (outlier).
*Ligand does not displace HOH-401.
†Ligand has multiple binding orientations.

a large range of affinities would be advantageous for making statistical comparisons. This docking study provides an important benchmark illustrating the 'limits of detection' for modeling ligand recognition using the current version of AutoDock.

Conclusions

The W191G cavity is an engineered binding site that binds and can oxidize small cyclic compounds. W191G represents a nascent active site: removal of the Trp191 side chain created a binding cavity, but substrate-binding specificity has not yet evolved. In spite of the low specificity of the W191G cavity, illustrated by the small range of binding affinities for many different ligands, it is remarkable that AutoDock is able to predict correct ligand orientations for a majority of the cases tested. AutoDock 3.0 reproduced binding orientations within 1.0 Å RMSD and binding energies within ± 0.8 kcal/mol. AutoDock captures the subtle features of binding in the cavity and distinguishes between ligands with single and multiple orientations. A high degree of clustering appears to be associated with correct predictions of single binding modes by AutoDock. Multiple modes were predicted either when the binding site waters were not explicitly included or when there was experimental evidence of multiple binding modes. In the cases where binding site waters varied, AutoDock correctly predicted the binding orientation when the binding site waters were explicitly included in the docking model. The prediction of multiple binding modes by AutoDock is an unanticipated utility that may enhance crystallographic analysis of disordered binding. Finally, we found trends in the level of clustering and predicted binding energy help to distinguish ligands from non-binders. This approach should prove useful for future design and evaluation of ligand binding in engineered protein active sites.

Acknowledgements

Support for this work was provided by fellowships from DOE-CSGF and the Burroughs-Wellcome fund (LJIS) to R.J.R., grants GM-41049 to D.B.G., GM-17844 to R.A.M., and P01GM48870 and P41RR08605 to A.J.O. We thank Gerard Jensen, Samantha Zeitlin, Anna-Marie Hayes, Vickie Roberts and Joseph Bonaventura for discussion and critical reading of the manuscript.

References

1. Finzel, B.C., Poulos, T.L. and Kraut, J., J. Biol. Chem., 259 (1984) 13027.
2. Yonetani, T., Enzymes, Academic Press, San Diego, CA, 1976.
3. Erman, J.E., Vitello, L.B., Miller, M.A., Shaw, A., Brown, K.A. and Kraut, J., Biochemistry, 32 (1993) 9798.
4. Houseman, A.L., Doan, P.E., Goodin, D.B. and Hoffman, B.M., Biochemistry, 32 (1993) 4430.
5. Sivaraja, M., Goodin, D.B., Smith, M. and Hoffman, B.M., Science, 245 (1989) 738.
6. Huyett, J.E., Doan, P.E., Gurgiel, R., Houseman, A.L.P., Sivaraja, M., Goodin, D.B. and Hoffman, B.M., J. Am. Chem. Soc., 117 (1995) 9033.
7. Musah, R.A. and Goodin, D.B., Biochemistry, 36 (1997) 11665.
8. Fitzgerald, M.M., Trester, M.L., Jensen, G.M., McRee, D.E. and Goodin, D.B., Protein Sci., 4 (1995) 1844.
9. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., Nucleic Acids Res., 28 (2000) 235.
10. Fitzgerald, M.M., Churchill, M.J., McRee, D.E. and Goodin, D.B., Biochemistry, 33 (1994) 3807.
11. Musah, R.A., Jensen, G.M., Bunte, S.W., Rosenfeld, R.J. and Goodin, D.B., J. Mol. Biol., 315 (2002) 845.
12. Fitzgerald, M.M., Musah, R.A., McRee, D.E. and Goodin, D.B., Nat. Struct. Biol., 3 (1996) 626.
13. Olson, A.J. and Goodsell, D.S., SAR & QSAR in Environmental Research, 8 (1998) 273.
14. Minke, W.E., Diller, D.J., Hol, W.G. and Verlinde, C.L., J. Med. Chem., 42 (1999) 1778.
15. Bernard, P., Kireev, D.B., Chretien, J.R., Fortier, P.L. and Coppet, L., J. Comput.-Aided Mol. Des., 13 (1999) 355.
16. Gamper, A.M., Winger, R.H., Liedl, K.R., Sotriffer, C.A., Varga, J.M., Kroemer, R.T. and Rode, B.M., J. Med. Chem., 39 (1996) 3882.
17. Morris, G.M., Goodsell, D.S., Halliday, S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., J. Comput. Chem., 19 (1998) 1639.
18. Sotriffer, C.A., Flader, W., Winger, R.H., Rode, B.M., Liedl, K.R. and Varga, J.M., Methods, 20 (2000) 280.
19. Nacro, K., Bienfait, B., Lee, J., Han, K.C., Kang, J.H., Benzaria, S., Lewin, N.E., Bhattacharyya, D.K., Blumberg, P.M. and Marquez, V.E., J. Med. Chem., 43 (2000) 921.
20. Hyndman, M.E., Verma, S., Rosenfeld, R.J., Anderson, T.J. and Parsons, H.G., Am. J. Physiol. Heart Circulat. Physiol., 282 (2002) H2167.
21. Rao, M.S. and Olson, A.J., Proteins, 34 (1999) 173.
22. Laederach, A., Dowd, M.K., Coutinho, P.M. and Reilly, P.J., Proteins, 37 (1999) 166.
23. Mahmoudian, M., J. Mol. Graph. Model., 15 (1997) 149.
24. Coutinho, P.M., Dowd, M.K. and Reilly, P.J., Proteins, 27 (1997) 235.
25. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Caterina, G., Alagona, G., Profeta Jr, S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
26. Bessler, B.H., Merz Jr., K.M. and Kollman, P.A., J. Comput. Chem., 11 (1990) 431.

27. Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A., J. Comput. Chem., 7 (1986) 230.
28. Stouten, P.F.W., Frömmel, C., Nakamura, H. and Sander, C., Mol. Simul., 10 (1993) 97.
29. Morris, G.M., Goodsell, D.S., Huey, R. and Olson, A.J., J. Comput.-Aided Mol. Des., 10 (1996) 293.
30. Jensen, G.M., Goodin, D.B. and Bunte, S.W., J. Phys. Chem., 199 (1996) 954.
31. Frisch, M.W., GAUSSIAN92/DFT; GAUSSIAN94 Revision B.1, Gaussian, Inc., Pittsburgh, PA, 1995.
32. Hehre, W.J., Radom, L., Schleyer, P.R. and Pople, J.A., Ab Initio Molecular Orbital Theory, Wiley, New York, 1986.
33. Goodsell, D.S., Morris, G.M. and Olson, A.J., J. Mol. Recognit., 9 (1996) 1.
34. Upson, C., Faulhaber, T., Kamins, D., Laidlaw, D., Schlegel, D., Vroom, J., Gurwitz, R. and van Dam, A., IEEE Comp. Graphics App., 9 (1989) 30.
35. Shoichet, B.K., Leach, A.R. and Kuntz, I.D., Proteins, 34 (1999) 4.
36. Rarey, M., Kramer, B. and Lengauer, T., Proteins Struct. Funct. Genet., 34 (1999) 17.
37. Ogata, K. and Wodak, S.J., Protein Eng., 15 (2002) 697.
38. Ladbury, J.E., Chem. Biol., 3 (1996) 973.
39. Tilton, R.F., Jr., Singh, U.C., Weiner, S.J., Connolly, M.L., Kuntz, I.D., Jr., Kollman, P.A., Max, N. and Case, D.A., J. Mol. Biol., 192 (1986) 443.
40. Raymer, M.L., Sanschagrin, P.C., Punch, W.F., Venkataraman, S., Goodman, E.D. and Kuhn, L.A., J. Mol. Biol., 265 (1997) 445.
41. Osterberg, F., Morris, G.M., Sanner, M.F., Olson, A.J. and Goodsell, D.S., Proteins, 46 (2002) 34.
42. Rosenfeld, R.J., Garcin, E.D., Panda, K., Andersson, G., Aberg, A., Wallace, A.V., Morris, G.M., Olson, A.J., Stuehr, D.J., Tainer, J.A. and Getzoff, E.D., Biochemistry, 41 (2002) 13915.
43. Musah, R.A., Jensen, G.M., Rosenfeld, R.J., McRee, D.M., Bunte, S.W. and Goodin, D.B., J. Am. Chem. Soc., 119 (1997) 9083.