Published in final edited form as:
J Med Chem. 2004 October 7; 47(21): 5076–5084.

Soft Docking and Multiple Receptor Conformations in Virtual Screening

Anna Maria Ferrari[†,§,‖], **Binqing Q. Wei**[†,§,⊥], **Luca Costantino**[‡], and **Brian K. Shoichet**[*,†]
Department of Pharmaceutical Chemistry, University of California San Francisco, Genentech Hall, 600 16th Street, San Francisico, California 94143-2240, and Dipartimento di Scienze Farmaceutiche, Università di Modena e Reggio Emilia, via G. Campi 183, 41100 Modena, Italy

Abstract

Protein conformational change is an important consideration in ligand-docking screens, but it is difficult to predict. A simple way to account for protein flexibility is to soften the criterion for steric fit between ligand and receptor. A more comprehensive but more expensive method would be to sample multiple receptor conformations explicitly. Here, these two approaches are compared. A "soft" scoring function was created by attenuating the repulsive term in the Lennard-Jones potential, allowing for a closer approach between ligand and protein. The standard, "hard" Lennard-Jones potential was used for docking to multiple receptor conformations. The Available Chemicals Directory (ACD) was screened against two cavity sites in the T4 lysozyme. These sites undergo small but significant conformational changes on ligand binding, making them good systems for soft docking. The ACD was also screened against the drug target aldose reductase, which can undergo large conformational changes on ligand binding. We evaluated the ability of the scoring functions to identify known ligands from among the over 200 000 decoy molecules in the database. The soft potential was always better at identifying known ligands than the hard scoring function when only a single receptor conformation was used. Conversely, the soft function was worse at identifying known leads than the hard function when multiple receptor conformations were used. This was true even for the cavity sites and was especially true for aldose reductase. To test the multiple-conformation method predictively, we screened the ACD for molecules that preferentially docked to the expanded conformation of aldose reductase, known to bind larger ligands. Six novel molecules that ranked among the top 0.66% of hits from the multiple-conformation calculation, but ranked relatively poorly in the soft docking calculation, were tested experimentally for enzyme inhibition. Four of these six inhibited the enzyme, the best with an IC_{50} of 8 μM. Although ligands can get better scores in soft docking, the same is also true for decoys. The improved ranking of such decoys can come at the expense of true ligands.

Introduction

Protein flexibility is a frontier problem in molecular docking, especially for virtual screening applications. In such calculations, a large database of molecules is screened for complementarity to a macromolecular target. Typically, the target is represented by a single conformation. This can fully represent only rigid structures, and many receptors change

* Corresponding author. Phone: 415-514-4126. Fax 415-502-1411. E-mail:shoichet@cgl.ucsf.edu..
[†]University of California San Francisco.
[‡]Università di Modena e Reggio Emilia.
[§]These authors contributed equally to this work.
[‖]Current address: Dipartimento di Scienze Farmaceutiche, Università di Modena.
[⊥]Current address: The Scripps Research Institute.

Supporting Information Available: Table of van der Waals parameters used for soft docking. This material is available free of charge via the Internet at http://pubs.acs.org.

conformation on ligand binding. Unfortunately, allowing for even a small amount of conformational change increases the number of receptor conformations substantially, scaling exponentially with degrees of freedom. A brute-force exploration of accessible states for a fully flexible active site at the time of docking is currently infeasible.

Investigators have, therefore, attempted to restrict the number of states sampled by a flexible receptor in docking screens. One way to do this is to restrict the region of the receptor that can change conformation. Thus, an induced fit of a restricted set of flexible side chains has been treated by torsion-angle optimization either during docking[1] or after approximate positions of ligand are found.[2] Rotamer libraries[3] have been used to sample particular side-chain states discretely.[4–6] A second way to restrict the number of states is to take discrete snapshots of different receptor conformations. Low-energy conformations can be sampled using molecular dynamics simulations; an advantage of this approach is that it naturally combines side-chain and backbone movements.[7,8] Similarly, ensembles of experimental structures have been used to calculate average potential-energy grids of the flexible site, which were then used in the docking calculation.[9] Olson et al. extended this approach to structurally ordered water molecules.[10] The FlexE algorithm, recently introduced by Lengauer and co-workers, samples discrete receptor conformations in a combinatorial fashion while incrementally building ligands into the site;[11] we have proposed a similar method.[12] These combinatorial methods can treat whatever movements are observed in the experimental structures, including both side-chain and backbone movements.

One way to avoid sampling multiple conformations is to permit some steric clashes between ligand and protein. This implicitly models receptor accommodation by loosening the criterion for steric fit, for instance, by reducing the steepness of the repulsion term in the Lennard-Jones potential function.[13] Applications of such an approach have been considered by Abagyan and colleagues[14] for protein docking and by Vieth, Brooks, and colleagues, among others, for ligand docking.[15–17] An attractive feature of “soft docking” is that it is simple to implement and is much faster than explicit sampling of multiple receptor conformations. A disadvantage is that it can address only small conformational changes.

Here, we compare soft docking and explicit sampling for docking screens of large libraries against protein structures. We chose target sites that were more or less suited to the two different approaches. As ideal soft-docking sites, we used two cavities in the T4 lysozyme.[18–20] Both are small and completely sequestered from solvent. One, the mutant L99A (created by the substitution Leu99 → Ala in the core of the protein), is almost completely hydrophobic; the second, L99A/M102Q (created by the double mutant Leu99 → Ala, Met102 → Gln), introduces a single polar residue into the cavity. Despite their complete burial in the core of the protein, many small aryl hydrocarbons bind to these cavities. The complexed structures of many of these ligands have been determined,[12,19,20] and the protein motions involved in their binding have been studied.[21] Whereas relatively small ligands, such as benzene and toluene, are accommodated with little conformational change, larger ligands, such as butyl-benzene, expand the cavities through side-chain and backbone movements on the order of 1 Å (Figure 1A). In doing so, the cavity volume expands from 209 $Å^3$ for the benzene complex to 295 $Å^3$ for the butyl-benzene complex (as calculated by the CASTP server[22]).

A third site, that of aldose reductase, seems more suited to explicit sampling of receptor conformations. On binding several larger inhibitors, such as Tolrestat, aldose reductase undergoes a large conformational change to reveal a second subsite that is not present when smaller ligands, such as Sorbinil, bind (Figure 1B). The opening of this subsite is necessary to accommodate the larger ligands, which would intersect the surface of the smaller, Sorbinil-bound conformation of the protein (yellow surface in Figure 1B). A soft-docking approach might have difficulty in allowing for such a large conformational change.

The two methods were compared in docking screens of over 200 000 diverse molecules, including 56, 77, and 908 known ligands for the L99A, L99A/M102Q, and aldose reductase sites, respectively. The algorithms were first evaluated by the enrichment of the known ligands as high-ranking "hits". A second criterion was whether the docked geometries resembled those from the crystal structures. Both soft-docking and multicon-former protein-sampling algorithms were implemented in the same docking program, DOCK3.5.54,[20,23,24] a derivative of DOCK3.5[25–27] that docks multiconformational ligand ensembles and corrects for ligand solvation energy. For soft docking, the repulsive term in the AMBER-derived Lennard-Jones potential function was attenuated to vary with the 9th power of ligand–protein interatomic distance rather than the usual 12th power (Figure 2). For these soft-docking calculations, the ACD database was screened against apo or apolike conformations of the target proteins. For multiconformation receptor docking, the standard 12-6 Lennard-Jones potential was employed with a method that samples different explicit receptor conformations and allows for recombination among predefined flexible regions, as previously described.[12]

We expected that soft docking would do well in the cavity sites, leading to higher enrichments compared to hard docking against a single rigid, apo conformation. Compared to explicit sampling of multiple receptor conformations, we expected soft docking to lead to competitive enrichments and geometries. Whereas the former expectation was realized, we were surprised to find that even in the cavity sites the softer steric potential led to worse rankings and worse geometries than the explicit sampling method. Explicit sampling had an even greater advantage in aldose reductase.

To test the multiconformational-receptor method prospectively, we turned to experimental testing of novel molecules for aldose reductase. From the screen of the ACD database against this site, we looked for new molecules, previously unknown to bind to the enzyme, that were predicted to bind to the larger conformations of the enzyme. Six such compounds were tested for inhibition in an enzyme assay.

Results

Flexible Cavities

The 202 383 molecules of the ACD database were docked against the L99A and L99A/M102Q cavities with DOCK3.5.54,[23,24] using either soft docking or multiconformational docking. For soft docking, the $1/R^{12}$ term of the Lennard-Jones energy was substituted by a more permissive $1/R^{9}$ term. Parameters in the Lennard-Jones equation were weighted to affect minimally the magnitude and position of the energy minimum while making the repulsive portion of the curve less restrictive (Figure 2). (See Methods.) For explicit receptor sampling, we chose 4 characteristic large and small conformations of the cavity sites on the basis of X-ray crystal structures and docked them against an ensemble of 12 conformations based on these 4 structures in a single multiconformational-receptor calculation. As in previous work,[12] we found it important to adjust the docking energies for each receptor conformation by the relative internal energy of forming the cavities, which is expected to increase as the cavity volume increases.[28]

There are 56[19,29] and 77[12,20] known ligands for the L99A and L99A/M102Q cavities, respectively. The success of the two treatments was evaluated by two criteria: the enrichment of these known ligands versus other database molecules and the geometries of their docked structures. In both apolar cavity L99A and slightly polar cavity L99A/M102Q, soft docking against the apo cavity structures led to higher ligand enrichments than did hard docking, using the standard 12-6 potential, against the same single-cavity conformation (Figure 3A and 3B). Thus, soft docking successfully found 57% of the 56 L99A ligands and 64% of the 77 L99A/M102Q ligands among the top 1% of ranked molecules in the database. Hard docking found

only 51% of the L99A ligands and 49% of the L99A/M102Q ligands among the top 1% of ranked molecules. This improvement reflects the ability to recognize larger ligands, which cannot be easily accommodated by the apo cavity with the standard 12-6 potential but which the more permissive 9-6 potential penalizes much less (Table 1).

The enrichment factors improved yet again when multiple receptor conformations were explicitly sampled (Figure 3A and B). In these calculations, 72% of the L99A ligands and 68% of the L99A/M102Q ligands were found in the top 1% of the ranked database. Although the improvement over soft docking might seem small (72 versus 57% for the L99A cavity and 68 versus 64% for the L99A/M102Q cavity), its significance is supported by the types of molecules that were highly ranked and their geometries in the cavities. When multiple receptor conformations are explicitly sampled, both smaller and larger ligands for these cavities are highly ranked (Table 1), which is sensible because their binding energies are often similar. However, the rankings of smaller ligands in soft docking typically fell significantly (Table 1). For instance, benzene ranks well (98th overall) in hard docking against the apo cavity of L99A and is still ranked well in the multireceptor conformer docking (164th overall), presumably because the apo structure is represented in the explicitly sampled conformations. In contrast, the rank of benzene drops to 994th in the soft-docking calculation; similar trends were observed for other small ligands. This is explained not by a drop in the magnitude of their energies, which remained the same or even improved, but by an improvement in the energy of larger molecules, some of which are ligands but some of which are decoys. However, the very largest cavity ligands continued to have unfavorable interactions, even in the soft potential, suggesting that there is a limit to the accommodation allowed by this softening, whereas their energies improved significantly for the multiconformation method (Table 1).

In the polar L99A/M102Q cavity, which binds both polar and apolar aryl hydrocarbons, the balance between the types of ligands also differed between soft docking and multiconformational docking. In the soft-docking calculation, the relative ranking of polar ligands dropped (Table 2). For instance, phenol, which hydrogen bonds with Gln102, ranks 306th overall in the multi-conformational calculation and also ranks well by hard docking against a single conformation of the apo cavity. In the soft-docking calculation, phenol's rank drops to 875th. The overall good enrichment achieved by soft docking appears to be attended by a shift in the balance of interactions toward apolar ligands at the expense of polar ligands.

A final criterion by which to judge the performance of the soft and multiconformer methods is the geometries they produce. Although both techniques calculated fairly accurate geometries, those of the multiconformational method were closer to the crystallographic structures (Figure 4, Table 3; all rmsd values compare 3D structures calculated from 2D information, i.e., without crystallographic bias, to those observed in the actual crystal structures). The difference was largest among the larger ligands. For instance, the docked geometry of isobutylbenzene predicted by multiconformer docking differs from the crystallographic result by a 0.16-Å rmsd, whereas the geometry of the same molecule predicted by soft docking differs from the crystallographic result by a 0.50-Å rmsd (Figure 4).

We considered the possibility that the improved performance of multiconformational docking reflected the use of holo conformations, whereas apo or apolike conformations were used in the soft-docking screens. To control for this, we conducted soft- and hard-docking screens of the ACD against the isobutylbenzene complex conformation of the hydrophobic cavity. In both calculations, the ranks of the larger ligands improved relative to those of smaller ligands in the apo-cavity screen; this effect was more pronounced for the soft-docking calculation. The number of large decoys among the top-ranking hits also increased. Correspondingly, enrichment factors were reduced. Thus, 72% of the known ligands were found in the top 1% of the docking-ranked database in the multiconformation-cavity screen; in the soft-docking

screen against the isobutylbenzene-bound cavity conformation only 14% of the known ligands were found in this top 1% (results not shown). This suggests that the differences we observe reflect differences between the different scoring functions and not simply the increased reliance on bound conformations in the multiconformer receptor method.

Aldose Reductase

In contrast to the cavity sites, which undergo relatively small accommodations on ligand binding, large inhibitors of aldose reductase (AR) lead to a large conformational change, resulting in the opening of a second pocket in the ligand binding site (Figure 1B). Thus, whereas inhibitors such as Sorbinil bind to a conformation closely resembling the apo enzyme, inhibitors such as Tolrestat and Zenarestat can be well accommodated only by the opening of this second subsite. This enzyme would thus seem to be better suited to a method that explicitly samples multiple receptor conformations.

A multiconformational ensemble of AR was built on the basis of the crystallographic structures of the enzyme in complex with Sorbinil, Tolrestat, Zenarestat, and Alrestatin, [30–32] which, owing to recombination, allowed for six explicit receptor conformations (Methods section). This ensemble was then used to screen a version of the ACD database that had been seeded with 908 annotated AR inhibitors from the MDL Drug Data Report database (MDDR). This database was also screened against the single, apolike, Sorbinil-bound conformation of the enzyme using both soft and hard docking.

As in the cavity sites, soft docking against the single apo conformation of the enzyme led to better enrichment of known ligands among the best-ranked docked molecules than did hard docking against the same target (Figure 5). Thus, the soft-docking calculation found 14.5% of the known ligands among the top 1% of the ranked database, whereas the hard-docking calculation against exactly the same site found only 12.6% of the known. The multiconformer-docking calculation did better still, finding 17.3% of the annotated 908 ligands in the top 1% of the ranked database. The difference between soft docking and the multiple conformation method increases as one looks further down the hit list; thus, by the top 2% of the ranked database, the soft-docking calculation has found 16.1% of the known ligands, most of which are small, Sorbinil-like compounds, whereas the multiconformational method has found 22.6% of the known ligands, here a mixture of larger and smaller inhibitors.

The difference between multiconformational receptor docking and soft docking is further born out by comparing the geometries of the docked ligands. Whereas large AR ligands such as Zenarestat and Tolrestat can be soft docked against the small, Sorbinil-bound conformation of the enzyme, their predicted geometries are incorrect (Figure 6). For these large ligands, the attenuation of the repulsive term in the Lennard-Jones equation cannot overcome the increase in active-site volume that is necessary to accommodate these inhibitors. Conversely, when these larger receptor conformations are explicitly considered by the multiconformational approach, these ligands are easily accommodated, and their best-scoring docked orientation corresponds closely to the crystallographic orientation (Figure 6).

To investigate the ability of the multiconformation receptor docking to predict inhibitors, we experimentally tested six novel molecules as inhibitors of AR. These molecules were selected on the basis of having high ranks in the multiconformational-receptor docking screens (within the top 0.66% of the docking-ranked ACD molecules), significantly lower ranks in the soft-docking calculation, docked orientations that placed them at least partially in the second subsite of the enzyme, and finally, whether the docked complexes looked sensible to us on graphical inspection. Four of these six molecules, compounds **1**–**4**, inhibited the enzyme with IC_{50} values of 8, 14, 336, and 274 μM, respectively (Table 4). By comparison, AR targeting drugs Tolrestat and Sorbinil have IC_{50} values of 1.1 and 0.038 μM, respectively, under the same conditions

as those used in these assays (Table 4). Compounds **1** and **2** were counter-screened against the chance of aggregation-based promiscuous inhibition; they were found neither to inhibit β-lactamase nor to form aggregates based on dynamic light scattering.[33,34] They thus seem to be well-behaved inhibitors of AR.

Discussion

In all three binding sites, the enrichment of known ligands by database screens was better for soft docking versus hard docking against a single conformation of the receptor and was better again for hard docking against multiple receptor conformations (Figures 3 and 5). The improved enrichment performance of soft docking was partially offset by less accurate docked geometries (Figure 4, Table 3), and by the increased weight given to apolar over polar interactions (Table 2). This may be readily understood: by reducing the penalty for close contacts, one favors larger ligands over smaller ones and van der Waals over polar interactions. For instance, in the polar cavity L99A/M102Q, the median van der Waals energy of the 100 top-ranking compounds in the soft-docking screen was 6 kcal/mol better than that in the hard-docking screen, out of a total median van der Waals energy of 22.2 for the soft-docking screen, whereas the electrostatic energies were much less affected. We note that we made no effort to rescore the initial docking poses from the soft-docking calculation with a more sophisticated calculation that allowed for the relaxation of the system.[35–38] Such a two-stage scoring method, which restores the hard potential in the second step, might well remove some of the decoy geometries and hits that we found here.

The decoy geometries and hits favored by soft docking are, by and large, not found with hard, multiconformational sampling, as long as one applies an internal energy correction for the different binding site conformations (Methods section).[12] This is because one is not changing the balance of energies in the scoring function. Moreover, in a site where a genuinely large conformational change can occur, such as AR, the greater permissiveness of the softer potential cannot make up for the large reorganization required to accommodate the larger ligands that induce it (Figure 1B). At least in these three systems, docking against multiple explicit receptor conformations gives the best results in database screening calculations. The questions become can one afford to dock against multiple receptor conformations, how often will one know enough receptor conformations to be able to do so, and what sorts of results might one expect in such cases?

Whereas brute-force docking against multiple receptor conformations in the general case is infeasible, there are several algorithmic alternatives that are relatively fast.[5,7,9–11] Here, we used a method that explicitly represents multiple potential energy grids for different receptor conformations and recombines grids from different mobile regions to extend the number of explicit receptor conformations further.[12] Admittedly, this method is limited by the number of explicit conformations it can hold in memory, but it appears to scale well with the number of receptor conformations.[12] For this and related algorithms, the time sacrifice to dock against a limited number of receptor conformations is not burdensome.

A second question is how often one will have the luxury of docking against multiple receptor conformations. Here, we restricted ourselves to the conformations that had been explicitly observed by X-ray crystallography and the combination of different movable regions from among these structures.[12] Both the cavity sites and AR have been intensely studied, and multiple structures are available for each site. Rarely will so many experimental conformations be available, and one would have to rely on calculated receptor conformations. The reliability of such conformations remains a matter of ongoing research; our experience is that injudicious use of calculated conformations can actually worsen docking performance.[12]

These caveats should not obscure the observation that multiconformation-receptor docking led to qualitatively different and better results versus soft docking against a single receptor conformation. Whereas we do not pretend to have experimentally tested the new algorithm on a large scale, the experimental results against AR are encouraging—most of the compounds tested turned out to be inhibitors, with two having IC_{50} values in the micromolar range. Not only would these novel inhibitors have been overlooked had we used a soft-docking approach but the hits that were returned by soft docking were also qualitatively different—typically less flexible and smaller than the hits from docking multiple receptor conformations. Overall, these studies suggest that when one has little trustworthy information as to alternate receptor conformations, soft docking can better enrich likely ligands than more traditional, hard potential functions in single-receptor conformation docking. This improved enrichment does come at a cost to geometric accuracy and to the balance between polar and apolar terms. Conversely, reliable information about multiple receptor conformations can be directly exploited to improve docking hit lists in database screens.

Methods

Soft-Docking Calculations

To soften the repulsive term in the van der Waals interaction energy, we replaced the original Lennard-Jones 12-6 potential (eq 1) with a 9-6 potential (eq 2) by modifying the program CHEMGRID,[26] which precalculates the van der Waals potential grid for the receptor. To maintain the same energy minima and the separating distances at which the minima occurred, it was necessary to modify the AMBER-based repulsive and attractive parameters (eq 3 and Figure 2). These changes were made in the CHEMGRID parameter file vdw.parms.amb.mindock (Supporting Information).

$$E(\varepsilon) = \frac{A}{r^{12}} - \frac{B}{r^{6}} \tag{1}$$

$$E(\varepsilon) = \frac{a}{r^{9}} - \frac{b}{r^{6}} \tag{2}$$

where

$$a = \sqrt{2AB} \quad b = \frac{3}{2}B \tag{3}$$

Receptors and Ligands

For hydrophobic cavity L99A, soft (9-6)- and hard-docking calculations were undertaken against the benzene-bound conformation of the cavity (181L), which is essentially the same as the apo conformation. Multiple receptor–conformation docking was conducted using an ensemble of 12 receptor conformations, derived from the structures of L99A in complex with the ligands *p*-xylene, indene, and isobutylbenzene (187L, 185L, 184L), in addition to the benzene-bound conformation, as described.[12] For these calculations, a version of the docking program DOCK3.5.54 that can treat receptor conformational ensembles was used.[12] For the polar cavity, L99A/M102Q, the structure of the apo cavity (1LGU) was used for the soft and hard single-conformation docking calculations. To calculate an ensemble for L99A/M102Q, we computationally substituted Met102 in the ensemble of conformations of L99A with the Gln102 from 1LGU, as described.[12] A total of 12 conformations were represented in the ensemble for L99A/M102Q. For AR, the crystallographic structures of pig AR in complex with inhibitors Sorbinil, Tolrestat, or Zenarestat and the Alrestatin-bound structure of the human enzyme were superposed by buried core residues that were not involved with inhibitor binding (1AH0, 1AH3, 1IEI, 1AZ1, respectively). Thus, the parts of the enzyme that moved

significantly upon inhibitor binding were identified; the receptor conformational ensemble was composed of three parts: (1) residues Cys298 through Cys303, Tyr309, and Phe122; one conformation from each of the three structures 1AH0, 1AH3, 1IEI (three conformations in total); (2) residues Arg217 through Lys221; one conformation from each of the two structures 1AH0 and 1AZ1 (two in total); (3) the rest of the protein, using the coordinates from the structure 1AH0. A total of six conformations are in the ensemble. We checked the conformations of those flexible parts to ensure that their recombination would not create an obvious steric clash.

The 56 known ligands of the apolar L99A cavity and 77 ligands for the polar L99A/M102Q cavity were compiled from previous work,[12,19,20,29] and the 908 inhibitors of AR were collected from the MDDR database (v2000, MDL Inc., San Leandro CA). We note that we made no effort to control for differences among the ligands and the decoy molecules in the ACD database. Especially for the cavity sites, where only about 60 000 ACD molecules will fit any of the site sterically, this has the effect of improving our absolute enrichment factors over what would be found had we only docked molecules that, for instance, were in the same molecular weight range as the ligands. Had we controlled for this, our cavity enrichment factors would have been 3- to 4-fold lower across the board. This would have no effect on the relative performance of the different docking methods, which are all compared using the same ligands and the same overall database.

Docking Preparation and Parameters

Docking “spheres” for the cavity sites were generated as described,[12,20] and a total of 53 spheres were used. A total of 90 spheres were generated for AR. In all docking calculations, initial ligand poses were filtered for steric complementarity using an excluded volume grid calculated by DISTMAP,[26] and the electrostatic component of the interaction energy was calculated on a grid computed with DelPhi.[39] The van der Waals potential grids were calculated by CHEMGRID[26] using either standard parameters or the modified parameters for the soft-docking calculation.

In all docking screens, a distance tolerance (dislim) of 0.3 Å was applied for matching the ligand onto the spheres, and 20 steps of rigid-body minimization were conducted after the molecules were docked. For the cavities, bin sizes were set at 0.2 Å and overlaps at 0.1 Å for both ligands and receptors; for AR, these were set at 0.3 and 0.2 Å, respectively.

Each screen of the ACD database (v2000, MDL Inc., San Leandro CA) was performed using DOCK3.5.54 with either the standard rigid receptor algorithm or a recent multireceptor conformation algorithm.[12] Both versions allow for ligand flexibility using pregenerated conformational ensembles [23,29,40] and account for ligand desolvation.[20,24] For the multireceptor conformation docking, the energetic cost of forming a particular receptor conformation, relative to the ground-state apo structure, was subtracted to docking pose energies particular to that conformation. For the cavity sites, this calculation was based on the differential volumes of the various conformations of the sites using a relation proposed by Matthews and co-workers.[18] As the ligands grow larger, so too do the cavity volumes; the correction is a linear penalty for the differential cavity volume engendered on ligand binding, as described.[12] Whereas this simple treatment is well accepted for the cavity sites, which are completely buried and largely hydrophobic, the energetics of most enzymes are more complicated. As we found previously for thymidylate synthase,[12] with AR we needed to consider differential surface area exposure and a term for how the electrostatic self-energy of the enzyme changed. The ground-state structure for AR was taken to be the Sorbinil-bound conformation, and the relative energies of the other five conformations were calculated on the basis of the differential nonpolar accessible surface areas and the differential electrostatic self-energies, calculated using a Poisson–Boltzmann treatment, as described.[12]

Synthesis

Tolrestat was synthesized according to a procedure reported in the literature.[41] Sorbinil was kindly provided by Pfizer.

Enzymology

AR (E.C. 1.1.1.21) was partially purified from pig lenses as reported.[42] Briefly, pig lenses were obtained locally from freshly slaughtered animals. Capsules were incised, and frozen lenses were suspended in sodium phosphate buffer pH 7.0 containing 5 mM DTT (1 g tissue/3.5 mL buffer) and stirred in an ice-cold bath for 1 h. The suspension was then centrifuged at 22 000*g* at 4°C for 20 min, and the supernatant was subjected to ion-exchange chromatography on DE52. Enzyme activity was measured by monitoring the change in absorbance at 340 nm, which accompanies the oxidation of NADPH cofactor, catalyzed by ALR2. The enzyme inhibition assays were performed at 37°C as previously described, using 4.7 mM D, L-glyceraldehyde as the substrate in 0.25 M sodium phosphate buffer pH 6.8 containing 0.38 M ammonium sulfate and 0.11 mM NADPH. The sensitivity of ALR2 to the inhibition by the compounds under study was tested in the above assay conditions by including the inhibitors dissolved in DMSO at the desired concentration in the reaction mixture. The DMSO in the assay mixture was kept at a constant concentration of 1%. A reference blank containing all of the above reagents, except the substrate, was used to correct for the nonenzymatic oxidation of NADPH. IC_{50} values (the concentration of the inhibitor required to produce 50% inhibition of the enzyme-catalyzed reaction) were determined from least-squares analyses of the linear portion of the log dose–inhibition curves. Each curve was generated using at least three concentrations of inhibitor causing an inhibition between 20 and 80% with duplicates at each concentration. The 95% confidence limits were calculated from T values for $n - 2$, where n is the total number of determinations.[43] The K_m for the substrate D,L-glyceraldehyde is 10 μM.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by GM59957 from the NIH (to B.K.S.). We thank John Irwin, Veena Thomas, Giulio Rastelli, and Ruth Brenk for reading this manuscript.

References

1. Totrov M, Abagyan R. Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins 1997;Suppl:215–220. [PubMed: 9485515]
2. Schnecke V, Swanson CA, Getzoff ED, Tainer JA, Kuhn LA. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. Proteins 1998;33:74–87. [PubMed: 9741846]
3. Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol 1987;193:775–791. [PubMed: 2441069]
4. Leach AR. Ligand docking to proteins with discrete side-chain flexibility. J Mol Biol 1994;235:345–356. [PubMed: 8289255]
5. Kallblad P, Dean PM. Efficient conformational sampling of local side-chain flexibility. J Mol Biol 2003;326:1651–1665. [PubMed: 12595271]
6. Frimurer TM, Peters GH, Iversen LF, Andersen HS, Moller NP, et al. Ligand-induced conformational changes: improved predictions of ligand binding conformations and affinities. Biophys J 2003;84:2273–2281. [PubMed: 12668436]
7. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, et al. Developing a dynamic pharmacophore model for HIV-1 integrase. J Med Chem 2000;43:2100–2114. [PubMed: 10841789]

8. Broughton HB. A method for including protein flexibility in protein–ligand docking: improving tools for database mining and virtual screening. J Mol Graphics Modell 2000;18:247–257.302–244
9. Knegtel RM, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. J Mol Biol 1997;266:424–440. [PubMed: 9047373]
10. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins 2002;46:34–40. [PubMed: 11746701]
11. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308:377–395. [PubMed: 11327774]
12. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK. Testing a flexible-receptor docking algorithm in a model binding site. J Mol Biol 2004 33;7:1161–1182. [PubMed: 15046985]
13. Carlson HA, Masukawa KM, McCammon JA. Method for including the dynamic fluctuations of a protein in computer-aided drug design. J Phys Chem A 1999;103:10213–10219.
14. Fernandez-Recio J, Totrov M, Abagyan R. Soft protein–protein docking in internal coordinates. Protein Sci 2002;11:280–291. [PubMed: 11790838]
15. Wu G, Robertson DH, Brooks CL 3rd, Vieth M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMm-based MD docking algorithm. J Comput Chem 2003;24:1549–1562. [PubMed: 12925999]
16. Vieth M, Hirst JD, Kolinski A, Brooks CL. Assessing energy functions for flexible docking. J Comput Chem 1998;19:1612–1622.
17. Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. J Med Chem 2004;47:45–55. [PubMed: 14695819]
18. Eriksson AE, Baase WA, Wozniak JA, Matthews BW. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. Nature 1992;355:371–373. [PubMed: 1731252]
19. Morton A, Baase WA, Matthews BW. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. Biochemistry 1995;34:8564–8575. [PubMed: 7612598]
20. Wei BQ, Baase WA, Weaver LH, Matthews BW, Shoichet BK. A model binding site for testing scoring functions in molecular docking. J Mol Biol 2002;322:339–355. [PubMed: 12217695]
21. Mulder FA, Hon B, Muhandiram DR, Dahlquist FW, Kay LE. Flexibility and ligand exchange in a buried cavity mutant of T4 lysozyme studied by multinuclear NMR. Biochemistry 2000;39:12614–12622. [PubMed: 11027141]
22. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci 1998;7:1884–1897. [PubMed: 9761470]
23. Lorber DM, Shoichet BK. Flexible ligand docking using conformational ensembles. Protein Sci 1998;7:938–950. [PubMed: 9568900]
24. Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. Proteins 1999;34:4–16. [PubMed: 10336382]
25. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982;161:269–288. [PubMed: 7154081]
26. Meng EC, Shoichet BK, Kuntz ID. Automated Docking with Grid-Based Energy Evaluation. J Comput Chem 1992;13:505–524.
27. Gschwend DA, Kuntz ID. Orientational sampling and rigid-body minimization in molecular docking revisited: on-the-fly optimization and degeneracy removal. J Comput Aided Mol Des 1996;10:123–132. [PubMed: 8741016]
28. Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, et al. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science 1992;255:178–183. [PubMed: 1553543]
29. Su AI, Lorber DM, Weston GS, Baase WA, Matthews BW, et al. Docking molecules by families to increase the diversity of hits in database screens: Computational strategy and experimental evaluation. Proteins 2001;42:279–293. [PubMed: 11119652]

30. Harrison DH, Bohren KM, Petsko GA, Ringe D, Gabbay KH, et al. The alrestatin double-decker: binding of two inhibitor molecules to human aldose reductase reveals a new specificity determinant. Biochemistry 1997;36:16134–16140. [PubMed: 9405046]

31. Urzhumtsev A, Tete-Favier F, Mitschler A, Barbanton J, Barth P, et al. A “specificity” pocket inferred from the crystal structures of the complexes of aldose reductase with the pharmaceutically important inhibitors tolrestat and sorbinil. Structure 1997;5:601–612. [PubMed: 9195881]

32. Kinoshita T, Miyake H, Fujii T, Takakura S, Goto T. The structure of human recombinant aldose reductase complexed with the potent inhibitor zenarestat. Acta Crystallogr, Sect D 2002;58:622–626. [PubMed: 11914486]

33. McGovern SL, Caselli E, Grigorieff N, Shoichet BK. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. J Med Chem 2002;45:1712–1722. [PubMed: 11931626]

34. McGovern SL, Helfand BT, Feng B, Shoichet BK, Caselli E, et al. A specific mechanism of nonspecific inhibition. J Med Chem 2003;46:4265–4272. [PubMed: 13678405]

35. Taylor RD, Jewsbury PJ, Essex JW. FDS: Flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. J Comput Chem 2003;24:1637–1656. [PubMed: 12926007]

36. Fernandez-Recio J, Totrov M, Abagyan R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. Proteins: Struct Funct Genet 2003;52:113–117. [PubMed: 12784376]

37. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004;47:1739–1749. [PubMed: 15027865]

38. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Arthurs S, et al. Deciphering common failures in molecular docking of ligand-protein complexes. J Comput Aided Mol Des 2000;14:731–751. [PubMed: 11131967]

39. Nicholls A, Honig B. A Rapid Finite Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson–Boltzmann Equation. J Comput Chem 1991;12:435–445.

40. Lorber DM, Udo MK, Shoichet BK. Protein–protein docking with multiple residue conformations and residue substitutions. Protein Sci 2002;11:1393–1408. [PubMed: 12021438]

41. Sestanj K, Bellini F, Fung S, Abraham N, Treasurywala A, et al. *N*-[5-(trifluoromethyl)-6-methoxy-1-naphthalenyl]thioxomethyl]-*N*-methylglycine (Tolrestat), a potent, orally active aldose reductase inhibitor. J Med Chem 1984;27:255–256. [PubMed: 6422042]

42. Costantino L, Rastelli G, Vescovini K, Cignarella G, Vianello P, et al. Synthesis, activity, and molecular modeling of a new series of tricyclic pyridazinones as selective aldose reductase inhibitors. J Med Chem 1996;39:4396–4405. [PubMed: 8893834]

43. Tallarida, R. J.; Murray, R. B. Manual of Pharmacologic Calculations with Computer Programs, 2nd ed.; Springer-Verlag: New York, 1987.

44. Ferrin TE, Huang CC, Jarvis LE, Langridge R. The MIDAS display system. J Mol Graph 1988;6:13–27.

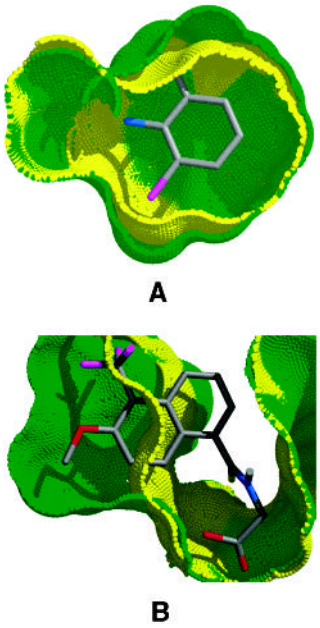


Figure 1.

Binding sites and conformational changes that the protein structures undergo. (A) Cutaway of the molecular surfaces of the polar cavity L99A/M102Q in its apo conformation (yellow), and the conformation that it adopts in complex with 2-fluoro-6-methylaniline (green). The crystallographic configuration of 2-fluoro-6-methylaniline is shown (carbon in gray, nitrogen in blue, fluorine in magenta). (B) Molecular surfaces of aldose reductase in its smaller, Sorbinil-bound conformations (yellow) and its larger, Tolrestat-bound conformation (green). The crystallographic configuration of Tolrestat is shown. Images were made using NEON in Midas-Plus,[44] as were Figures 4, 6, and 7.

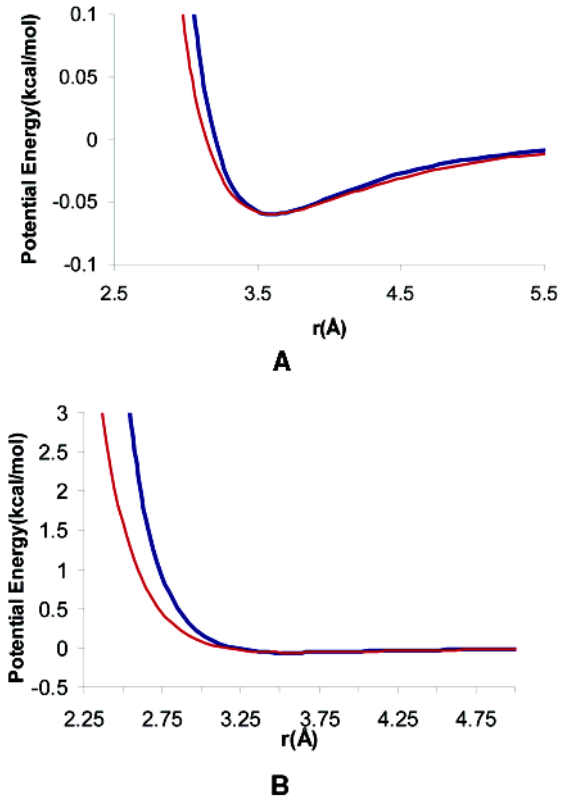


Figure 2.

Soft 9-6 (red) versus hard 12-6 (blue) van der Waals potential energies between two sp^3-hybridized carbon atoms. Parameters were adjusted so that the magnitude and location of the minima for the two functions were the same. (A) In the region near the energy minima. (B) In a broader range of distance.

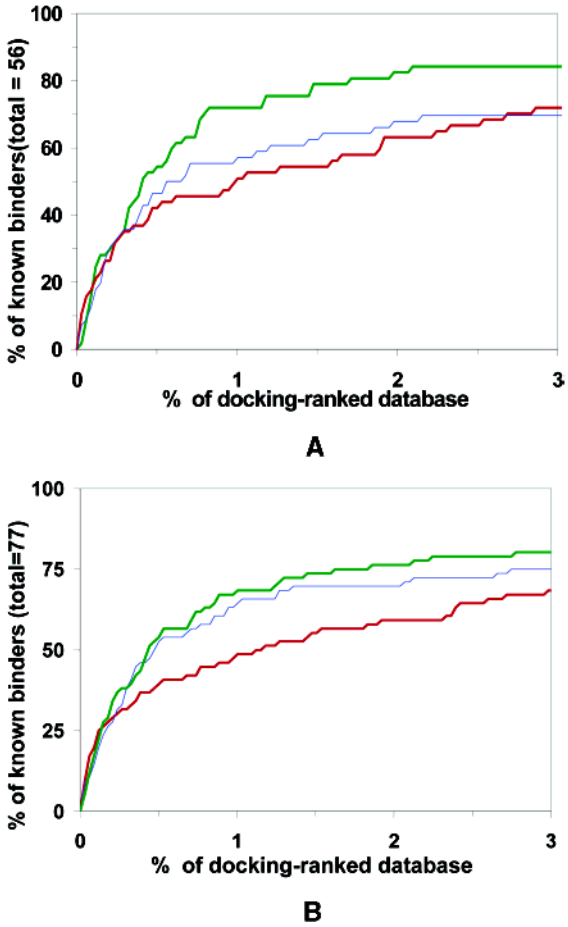


Figure 3.
Enrichment of known ligands for (A) the L99A cavity and (B) the L99A/M102Q cavity from docking screens of the ACD database. Three different scoring functions were used to rank the docked molecules: a 12-6 hard Lennard–Jones potential against the apo-cavity conformation (red), a soft 9-6 Lennard–Jones potential against the same structure (blue), and a multiconformation calculation using a hard potential (green).

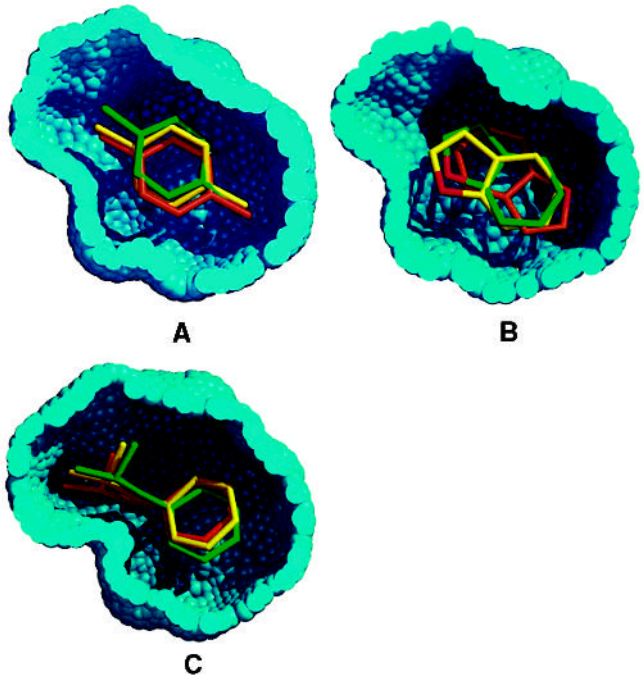


Figure 4.

Comparing docked and crystallographic geometries of three L99A cavity ligands: (A) *p*-xylene, (B) benzofuran, and (C) isobutylbenzene. Carbon atoms are colored in green in the crystallographic structures, in yellow in the configurations predicted by multiconformational docking, and in orange in the configurations predicted by 9-6 soft docking. Oxygen atoms are colored in red.

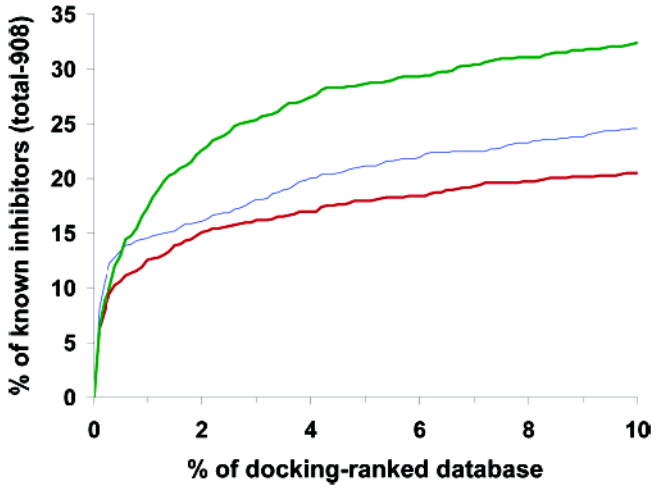


Figure 5.
Enrichments of known ligands for aldose reductase from docking screens of the ACD database. As in the cavity sites, three different scoring functions were used to rank the database: a hard-docking calculation against the smaller, Sorbinil-bound conformation of the enzyme (red), a soft-docking calculation against the same structure (blue), and docking using a multiconformational approach (green).

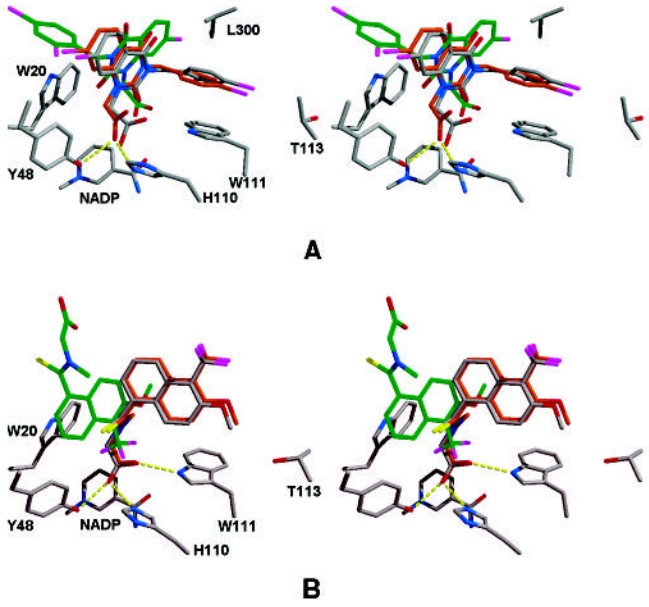


Figure 6.

Comparing the binding poses predicted by hard docking against multiple receptor conformations (carbons in orange) or by soft docking against the Sorbinil-bound conformation (carbons in green) to the corresponding crystallographic structures (carbons in gray) for two aldose reductase inhibitors: (A) Zenarestat and (B) Tolrestat. For Zenarestat, the rmsd values of the two predictions are 0.4 and 5.3 Å, respectively; for Tolrestat, the rmsd values are 0.3 and 7.2 Å, respectively. Pictures are in stereo. Dashed lines illustrate hydrogen bonds. Color scheme is as in Figure 1.

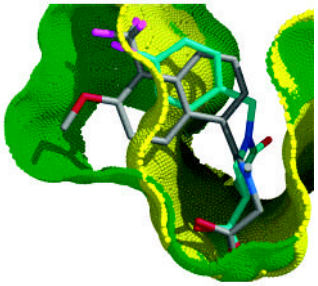


Figure 7.

Docking-predicted pose for compound **1** (carbon in cyan), a new 8 μM inhibitor, in comparison to the observed geometry of Tolrestat (carbon in gray) in the aldose reductase site. The molecular surfaces of the Sorbinil-bound (yellow) and Tolrestat-bound conformations (green) of the enzyme are shown.

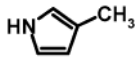
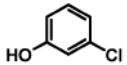
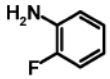
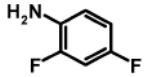
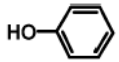
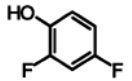
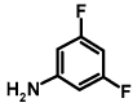
Table 1

Ranking of Small and Large Ligands from Docking 202 383 ACD Molecules against the L99A Cavity

ligand	docking rank		
	rigid cavity 12-6 potential	rigid cavity 9-6 potential	multiconformation 12-6 potential
		small ligands	
benzene	98	994	164
fluorobenzene	71	687	105
cyclohexene	543	2029	1311
thiophenol	103	447	173
toluene	28	157	33
		large ligands	
isobutylbenzene	10695	7724	1027
n-amylbenzene	21453	15229	2457
sec-butylbenzene	11885	8251	664
n-butylbenzene	12030	6488	2026
n-hexylbenzene	34698	34572	16528
propylbenzene	4307	991	717

Table 2

Difference in Rankings of Seven Representative Ligands of the Polar L99A/M102Q Cavity from Docking Screens of the ACD[a]

Compound	Structure	Ranking by soft docking	Ranking by multi-conformer docking	Δrank[b]
3-methylpyrrole		1377	350	−1027
3-chlorophenol		333	926	593
2-fluoroaniline		164	52	−112
2,4-difluoroaniline		155	122	−130
phenol		875	306	−569
2,4-difluorophenol		104	21	−83
3,5-difluoroaniline		44	26	−18

[a] Also see Figure 4.

[b] A negative value means a better rank was received using the multiconformational approach.

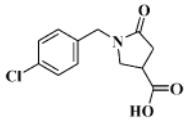
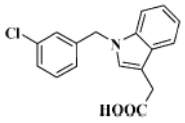
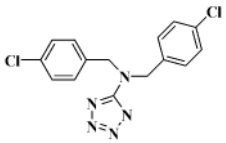
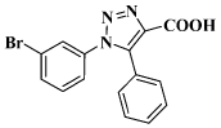
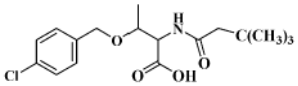
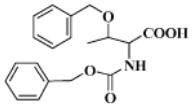
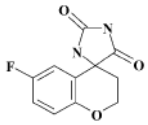
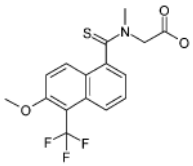
Table 3

rms Deviations from the Crystallographic Structures for Apolar (L99A) and Polar (L99A/M102Q) Cavity Ligands

compound	multiconformer: rms to X-ray (Å)	soft docking: rms to X-ray (Å)
	L99A cavity ligands	
benzene	0.08	0.09
p-xylene	0.15	0.49
indene	0.22	0.28
benzofuran	0.21	1.02
n-butylbenzene	0.39	0.36
isobutylbenzene	0.16	0.50
indole	0.83	0.84
	L99A/M102Q cavity ligands	
2-fluoroaniline	0.24	0.25
3,5-difluoroaniline	0.20	0.32
phenol	0.32	0.63
3-chlorophenol	0.89	0.94
3-methylpyrrole	0.07	0.12
2-allylphenol	0.37	0.41
3-fluoro-2-methylaniline	0.10	0.35

Table 4

IC_{50} Values of Novel Compounds Tested as Aldose Reductase Inhibitors Based on Multiconformer Docking[a]

Compound	Structure	Multi-conformer rank	Soft docking rank	Hard docking rank	IC_{50} (μM)
1		62	958	517	8
2		88	617	2949	14
3		263	68164	69,468	336
4		280	20614	30,978	274
5		790	158465	153,049	NI[b]
6		1256	157312	152,311	NI[c]
Sorbinil		NA	NA	NA	1.1
Tolrestat		NA	NA	NA	0.038

[a] Sorbinil and Tolrestat are included for comparison.

[b] IC_{50} > 400 μM.

[c] IC_{50} > 300 μM.