CMLS, Cell. Mol. Life Sci. 61 (2004) 783–793
1420-682X/04/080783-11
DOI 10.1007/s00018-003-3406-z

CMLS Cellular and Molecular Life Sciences

Strategies for the design of inhibitors of aldose reductase, an enzyme showing pronounced induced-fit adaptations

G. Klebe*, O. Krämer and C. Sotriffer

Philipps-Universität Marburg, Institute for Pharmaceutical Chemistry, Marbacher Weg 6, 35037 Marburg (Germany), Fax: +49 6421 28 2 89 94, e-mail: klebe@mailer.uni-marburg.de

Abstract. Aldose reductase is involved in the polyol pathway, catalyzing the reduction of glucose to sorbitol. However, due to pronounced binding site adaptations, the enzyme can operate on a broad palette of structurally diverse substrates ranging from small aliphatic and aromatic aldehydes up to steroid-type ligands. A comparative analysis of the presently accessible crystal structures of aldose reductase complexes reveals four binding-competent protein conformations. Additional relevant conformers are detected through molecular dynamics simulations. They indicate an equilibrium of several conformers which is shifted towards the binding-competent geometries upon ligand binding. Such a manifold system with several alternative binding site conformers requires some tailored concepts in virtual screening. We followed two strategies, both successfully suggesting new micromolar inhibitors. In a first attempt, we concentrated on one preferred conformer and performed a virtual screening, assuming that the binding pocket of aldose reductase adopts only this conformation. In a second approach, we followed a ligand superpositioning method. Ligands were extracted in their bound conformations from three different crystal structures, all accommodating the ligands with different active site conformations. After merging these ligands into one supermolecule, mutual alignments were computed, taking candidate ligands from a screening database. The latter strategy also retrieved several structurally new inhibitors of micromolar potency.

Key words. Drug design; virtual screening; protein-based pharmacophore; ligand-based pharmacophore; docking; molecular superpositioning; ligand scorting.

Introduction

The enzyme aldose reductase (AR) belongs to the superfamily of aldo-keto reductases. It adopts a triose phosphate isomerase (TIM)-barrel fold. Using NADPH as cofactor, it catalyzes the reduction of various aromatic and aliphatic aldehydes to the corresponding alcohols [1, 2]. As a key enzyme in the polyol pathway, its physiological role involves the reduction of glucose to sorbitol. Due to its polar properties, sorbitol does not easily cross the cell membrane, and its further conversion to fructose via sorbitol dehydrogenase is the rate-limiting step in the polyol pathway [3]. Accordingly, in situations of increased glucose flux, sorbitol can accumulate in cells of insulin-independent glucose uptake, leading to augmented osmotic pressure in the cells. In addition, enhanced oxidative stress arises as a result of the depletion of NADPH reservoirs. These biochemical phenomena have been related to a variety of diabetic complications and, accordingly, AR has received considerable attention as possible target for therapeutic intervention. Inhibiting AR and thus preventing the entry of glucose into the sorbitol pathway can reduce the damaging effects of late-onset diabetic disorders [4]. Considerable effort has been spent in the discovery and optimization of potent AR inhibitors (ARIs), resulting in a large variety of inhibitor-compound classes [5]. However, so far only a very small number of compounds has met the criteria of sufficient potency, oral activity and an acceptable side-effect profile. Accordingly, the area still requires further efforts, in particular with respect to the discovery of new leads. The present status of ARI development has recently been summarized by Suzen and Buyukbingol [6].

Besides its therapeutic potential for the treatment of diabetes, AR is particularly challenging for drug design due to its remarkably broad substrate promiscuity. The ac-

* Corresponding author.

commodation of structurally diverse substrates and, similarly, of a broad range of inhibitors is achieved through pronounced induced-fit adaptations of the binding pocket next to the catalytic centre. This intrinsic property of AR provides special challenge to structure-based inhibitor design and complicates the prediction of putative binding modes in docking and virtual screening. On the other hand, and this renders AR an ideal test case for structure-based drug design, the highest-resolution crystal structures hitherto available for an enzyme inhibitor complex have been obtained for AR [7].

Several successful studies for structure-based inhibitor design have been reported in case of AR [8–14]. Most of the presently used strategies for drug design were developed to handle rather rigid enzymes that do not allow for pronounced induced-fit adaptations of their binding sites. In the present review we want to summarize some strategies how to successfully perform virtual screening and ligand design on a mobile target showing significant flexibility.

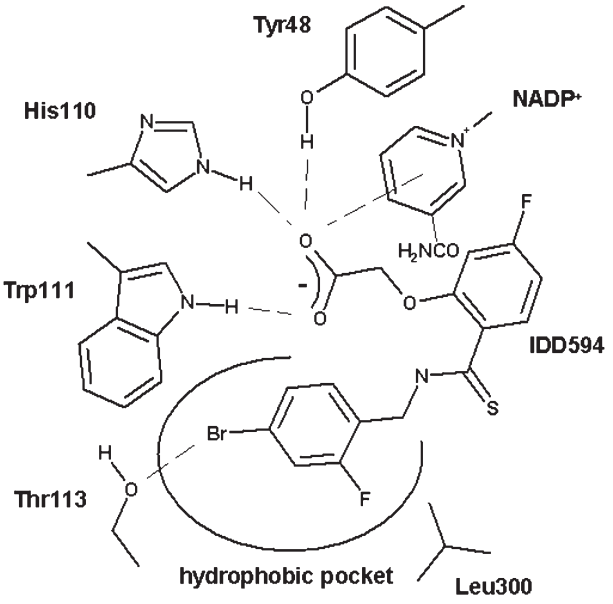

Figure 1. Sketch of the binding mode of IDD594 in the binding pocket of AR. The so-called anion binding pocket is composed by Trp111, His110, Tyr48 and the nicotinamide moiety of the bound cofactor NADPH/NADP+. A hydrophobic specificit pocket is opened next to the residues Thr113 and Leu300.

A starting point: comparative analysis of crystallographically determined AR inhibitor complexes

A sample set of 12 AR crystal structures[1] has been compared [15] using Relibase [16]. As reference the recently determined ultra-high-resolution structure (0.66 Å) with the bound inhibitor IDD594 has been selected [7, 17]. In this complex the inhibitor binds with its deprotonated carboxylate group to the anion binding pocket, supposedly adopting a binding mode similar to that of the natural aldehyde substrate (fig. 1). One of the carboxylate oxygens forms a hydrogen bond to Tyr48 OH and $N_{\varepsilon 2}$ of His110. The carboxylate carbon (or carbonyl carbon of the substrate) is placed next to C4 of the nicotinamide of the adjacent cofactor. The second carboxylate oxygen of IDD594 is involved in a hydrogen bond to $N_{\varepsilon 1}$ of Trp111. The three residues Tyr48, His110 and Trp111 are important for catalysis. Together with the nicotinamide moiety of the cofactor they compose the so-called anion-binding pocket. The nonpolar portion of IDD594 is accommodated in a second binding pocket outlined by Trp111 and Leu300 on either side. This specificity pocket is mainly of hydrophobic nature.

As further entries holoenzyme structures and additional complexes have been considered [15]. These include the complexes with citrate, cacodylate and glucose-6-phosphate. With respect to inhibitor complexes, the crystal structures with the carboxylates tolrestat, zopolrestat, zenarestat, alrestatin, and IDD384 are available. Out of the series of spirohydantoins, the complexes with sorbinil and fidarestat could be determined. All the latter inhibitors place their basic group into the same area as the carboxylate group of IDD594. With respect to the remaining ligand portions, rather distinct binding modes are observed. Despite a good overall rmsd achieved for the mutual C_α-fit (0.6–0.7 Å) of these complexes, the inhibitor IDD594 would for example penetrate into protein residues when being transferred from its complex structure to any of the other superimposed binding pockets. Interestingly enough, most frequently such mutual overlaps or clashes are observed with Leu300, while short contacts to Trp20, His110, Trp111, Thr113 and Phe122 are also observed, but less pronounced. The structure with bound zenarestat (1iei) resembles the IDD594 complex most closely in binding site geometry. Compared with the unbound state, these latter two inhibitors as well as tolrestat and zopolrestat open up new hydrophobic pockets. These cavities are not present in the holoenzyme or with the bound hydantoin sorbinil.

The composite picture of the various crystal structures suggests that the AR binding site splits into two parts with distinct flexibility properties (fig. 2). The part which is well superimposable across all known structures and shows virtually no variance is mainly formed by the residues of the catalytic site (Trp20, Tyr48, Val47, His110, Trp79 and Trp 111) and the flanking cofactor. Trp111 borders the anion binding pocket and exposes its π-face to the hydrophobic specificit pocket.

[1] PDBcodes: 1ads, 1ah0, 1ah3, 1ah4, 1az1, 1eko, 1el3, 1iei, 2acq, 2acr, 2acs, 2acu.

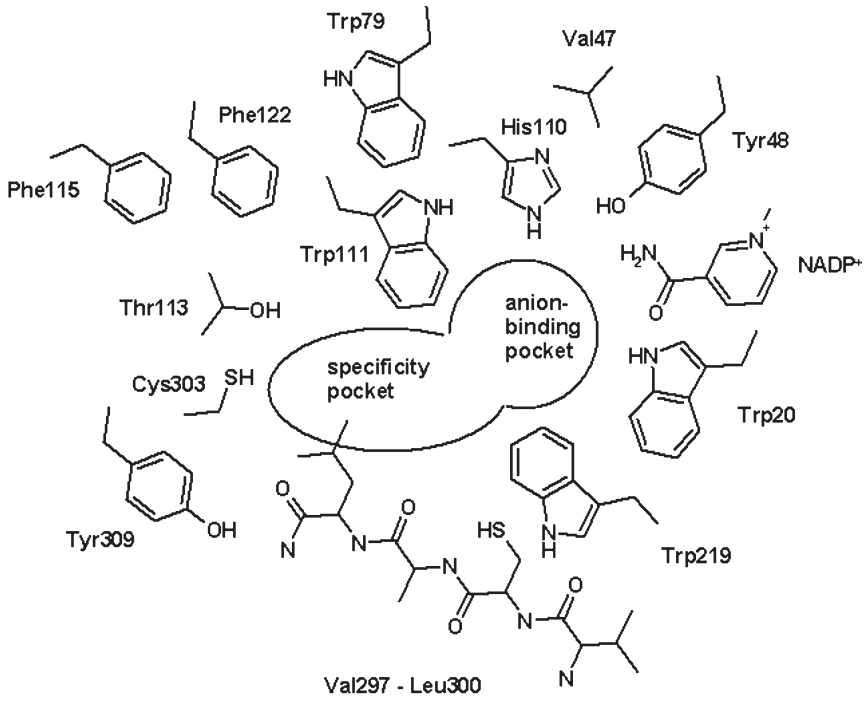

Figure 2. Sketch of the important residues involved in the formation of the AR binding pockets. The residues next to the anion binding pocket (Trp79, His79, Val47, Tyr48, Trp20 and $NADP^+$) all adopt very similar positions in the presently known crystal structures of protein-ligand complexes. They also perform minor motions in the MD simulations. The residues flanking the specificity pocket adopt different geometry depending on the actual binding site conformer experienced by AR. Also in the MD simulations these residues exhibit enhanced mobility, in particular Leu300, Phe122 and Trp219.

In contrast, the second binding site portion is flexible and shows frequent, but recurring changes. Most strongly affected is the segment Val297 to Leu300 and the adjacent region composed by Trp219, Cys303 and Tyr309. To a minor extent also the residues Thr113, Phe121 and Phe122 can be involved. However, the most variable residue among all is Leu300, being the determinant for the appearance of the occasionally formed hydrophobic binding pocket.

In summary, the analysis of the presently available crystal data on AR suggests the formation of essentially four distinct binding site conformations. They roughly resemble the bound-free situation, and the IDD594-, tolrestat-, and zenarestat-bound conformations, respectively. The ligand-free conformation found for the holoenzyme is virtually conserved upon binding of sorbinil (1ah0), alrestatin (1az1), IDD384 (1eko, 1el3), glucose-6-phosphate (2acq), cacodylate (2acr) and citrate (2acs). Binding of the other inhibitors mainly involves rearrangements of Leu300, operating as a type of gatekeeper initiating the formation of the size-adaptable hydrophobic binding pocket. In addition, the Cys298 to Leu300 backbone plays an important role by modulating the exposed hydrogen-bond donor and acceptor functionalities towards the ligand binding site. Additional adaptations involve Cys303, Tyr309, Thr113 and the aromatic portions of Phe115 and Phe122 [15].

Probing flexibility by MD simulations

To further probe potential binding site flexibility and adaptability of AR towards bound ligands, molecular dynamics simulations have been performed [15]. The goal of this study was to confirm the relevance of the four indicated protein conformers available to accommodate ARIs and to investigate whether additional stable conformers could potentially be experienced by AR giving rise to other yet unknown ligand binding modes. The simulations were set up based on the IDD594 complex. Of particular interest were those simulations where the ligand was removed from the binding cavity. Detailed evaluation of the obtained trajectories by clustering the various frames observed during the different simulations revealed a fairly consistent picture of the dynamic characteristics of AR. As already indicated by the sample of static crystal structures, flexibility is clearly exhibited by a limited number of binding site residues: most of the mobility is transmitted by about a third of the 15 binding site amino acids. Among these, Leu300 and the backbone stretch from residue number 297 to 300 play the key role. The side-chain motions of Leu300 amplified by the backbone movements lead to significant alterations and are responsible for the major part of conformational adaptations of the binding pocket. Among the other residues, mainly Phe122 and Trp219 show the most pro-

nounced mobility, both being in spatial proximity to Leu300. The molecular dynamics simulations support the dual character of the AR binding site being split into a dynamically rather rigid catalytic centre and a flexible area obviously responsible for accommodating the side chains of different substrates. In this context, it appears appropriate to think of the observed binding site adaptations not simply as a result of a ligand-induced fit, but of a preferential selection of protein conformations populated in an equilibrium [18, 19]. They are stabilized through the binding and local interactions with parts of a bound ligand. As a result, the conformational equilibrium is shifted towards the binding-competent conformations and the relevant conformational state becomes predominant and experimentally observable. This consideration also suggests that it appears a reasonable strategy in structure-based drug design to address induced-fit adaptations by the analysis of a set of pregenerated protein conformers.

In the present example of AR, would it be sufficient to consider only the four genuine conformers indicated by the comparative analysis of known crystal structures? The molecular dynamics simulations are in agreement with these experimentally confirmed conformers; however, there is no reason to believe that no further conformational variations are possible which could be stabilized upon binding of appropriate substrate or inhibitor molecules [15].

A first attempt: virtual screening focussed on the most prominent binding site conformation

Given the pronounced conformational flexibility of AR, the question is which of the latent conformational states should be used as reference to start a virtual screening study [20–22]. We decided to select the binding pocket adopted in the ultra-high-resolution structure of AR with IDD594 [23]. Besides the extraordinary resolution disclosing many structural details, IDD594, similarly to zopolrestat, binds with high affinity to this conformer. Obviously, in this conformational state the enzyme provides a binding pocket suited to accommodate inhibitors with nanomolar affinity.

A virtual screening strategy established in our laboratory has up to now successfully predicted submicromolar to nanomolar inhibitors for four systems [24, 25]. It follows a series of consecutive hierarchical filters. For the present search with the IDD594 pocket of AR as reference, we started with 260,000 compounds possessing a functional group suited to anchor to the anion binding pocket (s. fig. 1). Furthermore, the molecular weight of possible candidates was limited to 350 D and an upper limit of eight rotatable bonds was set. Along with these size- and complexity-limiting factors, the criteria defined by Lipinski et al. [26] were requested. These filters reduced the original data sample to 12,545 entries.

In a subsequent step, a three-dimensional (3D) search using Tripos' UNITY search engine was accomplished [27]. This search tries to match each candidate molecule onto a predefined pharmacophore model. The accuracy and relevance of this pharmacophore hypothesis is crucial for this important filter step. Accordingly, we try to extract the pharmacophore hypothesis directly from the reference protein structure. The protein binding site environment thus imposes the requirements to be matched by a putative ligand. To discover the most favorable areas likely to accommodate a particular type of ligand functional group, we map the binding pocket in terms of 'hot spots' of binding [28].

A number of alternative methods are available to map such areas, e. g. the force field-based method GRID [29], originally developed by Peter Goodford, and the approaches SuperStar [30] and DrugScore [31]. The latter ones are both based on crystal data. SuperStar calculates preferred interaction sites by mapping scatter plots or derived propensity distributions onto active site-exposed residues. The maps are obtained as composite crystal-field environments observed in the crystal packing of small organic molecules. DrugScore is a knowledge-based approach that uses statistically derived potentials using the so-called inverse Boltzmann principle. In a large sample of protein-ligand complexes the protein-to-ligand interface is analyzed in terms of distance-dependent occurrence frequencies by which a particular ligand atom type is found in contact with a protein atom type. For all three methods, a regularly spaced grid is embedded into the binding site. Subsequently, for the different ligand atom types interaction energies or contact propensities are calculated by systematically placing these probe atoms to the various grid intersections. As functional form to describe these interactions, the force field potential of GRID, the propensity distributions of SuperStar or the knowledge-based contact preferences of DrugScore are applied. In a final step, the obtained grid values are contoured according to a predefined level above the detected global minimum.

In our study of AR, SuperStar and DrugScore were used to detect favorable sites for an H-bond acceptor group and a hydrophobic aliphatic or aromatic group [23]. In figure 3, the crystallographically observed binding mode of IDD594 is shown together with the predicted acceptor sites. Acceptor hot spots are indicated in the anion-binding pocket and in the specificity pocket next to Thr113, where IDD594 places its bromine atom.

In a subsequent step the indicated hot spots were translated into a pharmacophore hypothesis appropriate for a database search using UNITY. In total, a set of six centres defining the pharmacophore were assigned, each centre addressing a particular property (donor, acceptor,

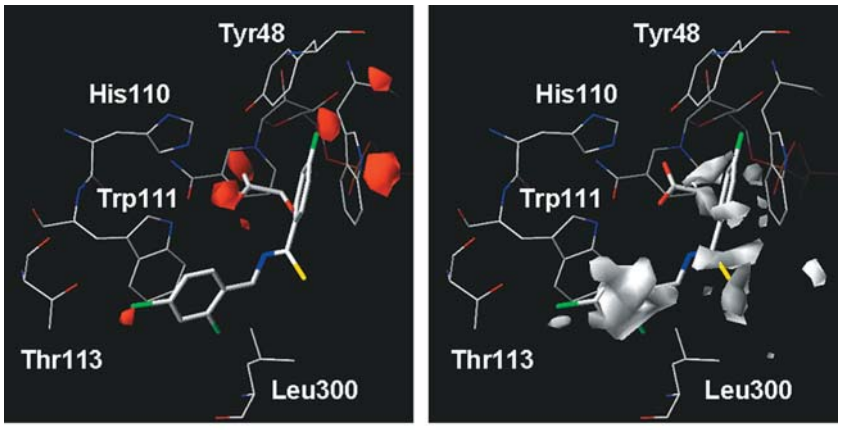

Figure 3. 'Hot spot' analysis of binding of the active site of AR. To facilitate orientation, the binding geometry of IDD594 is also shown as determined by crystal structure analysis. Using SuperStar, preferred areas to bind either a carbonyl oxygen as hydrogen-bond acceptor group (left) or a hydrophobic group (methyl probe, right) were mapped.

hydrophobic) and allowing for a limited spatial tolerance for the actual functional group match in the database search. On the basis of this query a set of 1261 compounds was retrieved. Among these hits, nearly 97% showed a terminal carboxylate group as anchor to be placed into the anion binding pocket. In a final step, the remaining 1261 hits were subjected to a docking with FlexX, taking the coordinates of the AR conformer accommodating IDD594 as reference receptor. The various binding modes generated by FlexX were initially ranked using DrugScore. As a knowledge-based scoring function, DrugScore samples over all contacts formed between ligand and protein atoms in the various binding poses. We observed some degree of surface dependence of DrugScore, ranking larger ligands automatically higher due to their increasing surface portions. This fact can easily exaggerate the affinity estimate of docking solutions of candidate molecules with higher molecular weight. To obtain a less biased ranking, we scaled our hits by the molecular weight, as recently suggested by Pan et al. [32]. In addition, we also considered the number of rotatable bonds in the scoring. Upon binding, ligands are immobilized at the binding site due to a loss of internal molecular degrees of freedom. This loss in mobility corresponds to a loss in binding entropy. In a recent study we could demonstrate experimentally that this loss in molecular motion is an important determinant in descriminating the turnover rate of stereoisomers in chiral resolution [33]. Since rigid molecules will experience a smaller conformational space and will accordingly suffer from smaller entropic losses upon immobilization in a binding pocket, the affinity will scale to some extent with the number of rotatable bonds. A knowledge-based scoring function such as DrugScore does not comprise a particular term considering the conformational flexibility of candidate ligands. To take the latter two aspects into account, we applied a ranking that considers the squared DrugScore score divided by molecular weight and the number of rotatable bonds.

Taken together, the achieved empirical ranking and the quality of the actual correspondence with the originally defined pharmacophore hypothesis reduced our search to the 206 most promising hits. Among these, we selected 120 candidates for further consideration. Based on the UNITY 2D fingerprint descriptors these hits were mutually clustered to facilitate their final visual inspection. In this step, the suggested binding conformations, the mutual surface complementarities between ligand and protein, and the possible occurrence of unfilled voids along the protein-ligand interface were analyzed. As a result, 9 compounds were selected for acquisition and subsequent testing in an enzyme inhibition assay (fig. 4). Out of these 9 compounds, 6 exhibited micromolar affinity, whereas 3 turned out to be unexpected false-positive hits. With the rather low number of tested compounds, it is difficult to assign a statistically significant hit rate. However, taking into account our other virtual screening studies applied to further enzymes that produced even better success rates, the hierachical filtering approach to virtual screening may be considered as reliable in retrieving active molecules. In the present study only carboxylic acids were suggested. This clearly matches with the experience that many of the previously described ARIs exhibit this functional group to anchor to the anion binding pocket. One can try to reason why 3 of the 9 compounds apparently correspond to unfavorable decoy structures even though docking had suggested them to fit into the AR binding site. Supposedly, insufficiently considered solvation/desolvation contributions or conformational restrictions of these suggested hits are responsible for the false-positive assignment.

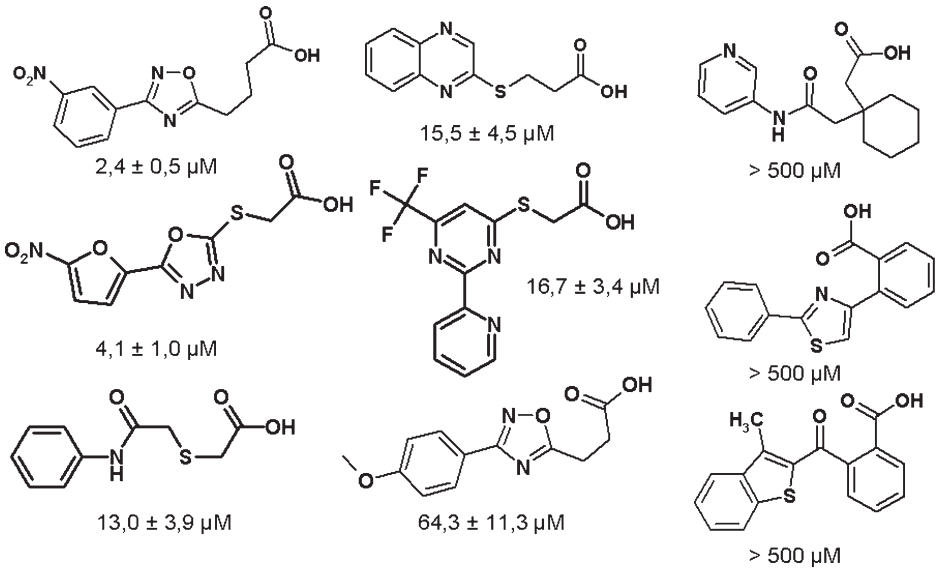

Figure 4. Chemical formulae of the nine virtual screening hits selected for experimental testing together with the measured IC_{50} values.

Strategies to screen simultaneously against a set of alternative binding site conformers

The analyses of known ARI crystal structures and the molecular dynamics (MD) simulations, described above, have suggested that AR adopts several binding-competent conformers, giving access to alternative binding pockets. Our first virtual screening was focussed on one single binding site conformation. However, the question is whether a more sophisticated approach could consider the pronounced adaptability of the AR binding site. One possible strategy could be to follow a very similar protocol as described in the previous section, but to consider all relevant binding pockets at the same time. This certainly will provide the chance to screen for a broader range of hits. However, in several cases, retrieval of the same candidates suggested that binding to the alternative binding site conformers can be expected. Consequently, such a search strategy is likely to produce a rather fuzzy picture of multiply retrieved hits. Mutual ranking of the different solutions obtained from several separate virtual screening runs constitutes a hardly tractable problem. This is particularly due to the unknown energy contribution imposed by the conformational adaptation of the protein. Accordingly, in order to avoid overwhelming and unsolvable scoring problems, a strategy that allows simultaneous consideration of various binding site conformations in a docking search seems to be more promising. Here, the ligand can explore at the same time the best possible orientations with respect to the multiple binding site conformers.

At present, only a small number of concepts have been described in the literature to approach the docking problem by simultaneously addressing multiple sites. One example has been presented with AutoDock. It considers several superimposed grids at the same time, weighting in various ways the different grid points according to their accessibility and interaction energies [34]. Starting from a different concept, the program FlexE enables the simultaneous docking of a ligand into an ensemble of alternative protein conformers [35]. During the incremental build-up procedure followed in this program to assemble the entire ligand from a pre-placed base fragment, the various orientations of the binding site residues are taken into account and mutually combined. As a result, the most likely binding mode of a candidate ligand is generated in one of the possible, mutually exclusive binding site conformations. Most crucial in this approach is the reliable ranking of the different binding modes generated during the docking procedure. Furthermore, the approach has to rank the various protein conformations screened during the docking procedure. This requires finding a reliable energy criterion during the build-up procedure to select among the alternative protein conformations. This dual-scoring problem does not make the problem simpler; even so, the strategy appears very convincing.

We have focussed on another alternative which takes advantage of the successful superpositioning of ligand molecules according to their spatial similarity in terms of physicochemical properties [O. Krämer, Dissertation, University of Marburg, 2003] [36, 37]. In previous studies, we have applied the program SEAL [38] to produce reasonable ligand alignments as a prerequisite for subsequent comparative molecular field analysis (CoMFA) [39] and comparative molecular similarity indices (CoMSIA) [40]. Since reasonable geometries are produced in agreement with near-native binding modes, we decided to follow this approach to perform a virtual screening considering multiple binding modes.

Molecular superpositioning with SEAL

SEAL (steric and electrostatic alignment), originally developed by Kearsley and Smith [38] and further improved by us [36, 37], compares a ligand and a reference molecule in terms of associated property-based Gaussian functions. This reference could either be one individual molecule or, as attempted in our present AR case, a combination of several merged ligands ('supermolecule') corresponding to the alternative binding modes observed in the various AR conformers. Thus, as a major advantage, this strategy reflects the multiple orientations of ligands in exclusively opened binding pockets simultaneously in terms of the combined supermolecule as reference. The associated Gaussian functions can be localized either at the atomic positions in a molecule, or to reduce computational demands, they can be attributed to entire fragments. They represent the shape of the molecules in space (figs 5, 6). In addition, they are loaded with a vector describing five different physicochemical properties (steric, electrostatic, hydrophobic, hydrogen-bond donor and acceptor features). The mutual similarity is measured by spatially superimposing the molecules and computing the scalar product of the associated property vectors, weighted by the integral of the overlapping Gaussian functions (fig. 7). Our extended and modified version of SEAL first performs a superposition of rigid multiple conformers of a probe ligand molecule onto the reference molecule. The multiple conformers are generated by MIMUMBA [41], a knowledge-based approach, to produce well-distributed conformers in that part of conformational space which is relevant for ligands in protein-bound situations. In a final step, the best-ranked rigidly superimposed ligand molecules are conformationally relaxed by optimizing the similarity condition with the reference while keeping their conformations as close as possible to the local minima obtained from MIMUMBA. Details of the method have been described elsewhere [36, 37].

Similar to the approach described above based on a protein-based pharmacophore we developed a ligand-based pharmacophore [O. Krämer, Dissertation, University of Marburg 2003]. A generally applicable procedure to derive a pharmacophore based on several superimposed ligands (e.g. the supermolecule reference) involves the following steps. Each of the superimposed molecules is expressed in terms of a set of atom-based Gaussian functions associated with a vector describing the steric, electrostatic, hydrophobic and hydrogen-bonding properties of each molecule (SEAL approach, fig. 7). All considered molecules are embedded into a regularly spaced grid. Subsequently, at each grid point the contributions of the Gaussians of all molecules are summed, forming

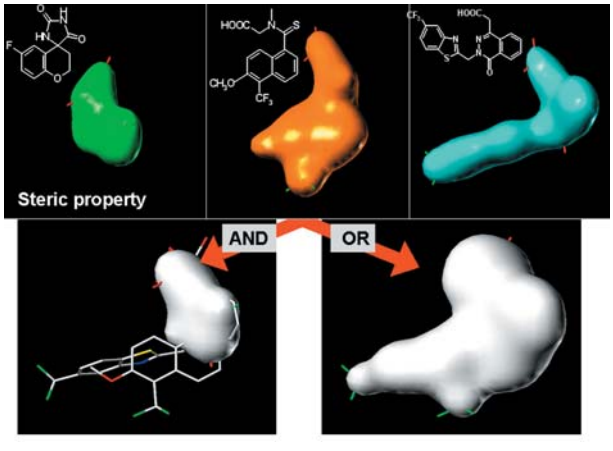

Figure 5. The steric properties of the three inhibitors sorbinil (upper left), tolrestat (upper middle) and zopolrestat (upper right) were expressed by associated Gaussian functions contoured at a predefined level. Similarly, the electrostatic, hydrophobic or H-bond donor and acceptor properties can be expressed (fig. 6). Through the formation of the arithmetic (lower left) or geometric mean the averaged properties of a combined supermolecule reference can be formed. This reference is subsequently used either to define the ligand-based pharmacophore shown in figure 8 or to mutually align with putative candidate molecules from a database search.

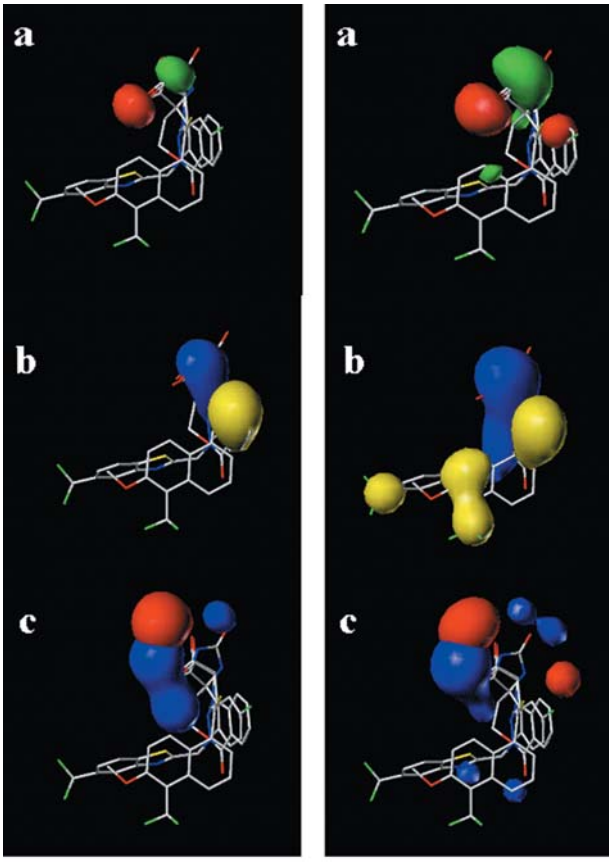

Figure 6. Field-based pharmacophore assigned to the supermolecule reference formed by sorbinil, tolrestat and zopolrestat. On the left, the AND, on the right the OR pharmacophore (fig. 5) is shown. (a) Electrostatic properties (green, positive; red, negative charge); (b) hydrophobic properties (yellow, hydrophobic; blue, hydrophobic); (c) H-bond donor (red) and acceptor (blue) facilities.

$$f_{mol1}(x) = \sum_{i=1}^{m} \begin{pmatrix} w_S \cdot steric \\ w_E \cdot elec \\ w_L \cdot lipo \\ w_D \cdot H_{don} \\ w_A \cdot H_{acc} \end{pmatrix}_i^{mol1} e^{-2\alpha \left| x - x_i^{mol1} \right|^2}$$

$$f_{mol2}(x) = \sum_{j=1}^{n} \begin{pmatrix} w_S \cdot steric \\ w_E \cdot elec \\ w_L \cdot lipo \\ w_D \cdot H_{don} \\ w_A \cdot H_{acc} \end{pmatrix}_j^{mol2} e^{-2\alpha \left| x - x_i^{mol2} \right|^2}$$

$$A_F = \int \left\langle f_{mol1}(x) \middle| f_{mol2}(x) \right\rangle dx = \sum_{i=1}^{m} \sum_{j=1}^{n} \left\langle \begin{pmatrix} w_S \cdot steric \\ w_E \cdot elec \\ w_L \cdot lipo \\ w_D \cdot H_{don} \\ w_A \cdot H_{acc} \end{pmatrix}_i^{mol1} \middle| \begin{pmatrix} w_S \cdot steric \\ w_E \cdot elec \\ w_L \cdot lipo \\ w_D \cdot H_{don} \\ w_A \cdot H_{acc} \end{pmatrix}_j^{mol2} \right\rangle \cdot e^{-\alpha \left| x_i^{mol1} - x_j^{mol2} \right|^2} \cdot C$$

$$C = \int e^{-4\alpha \left| x - \frac{1}{2}\left(x_i^{mol1} + x_j^{mol2} \right) \right|^2} dx = const.$$

geometric mean $$V(x) = \sqrt[N]{\prod_{i=1}^{N} v_i(x)}$$

If the physicochemical properties have opposite signs the geometric mean is set to zero.

arithmetic mean $$V(x) = \frac{\sum_{i=1}^{N} v_i(x)}{N}$$

Figure 7. The ligand and reference molecule (mol1, mol2) are mathematically expressed by associated Gaussian functions. The functions are either localized at the atomic position in a molecule, or to reduce computational effort, they are attributed to entire fragments. They express the shape of the molecules in space. Furthermore, they are loaded with a vector describing five different physicochemical properties (steric, electrostatic, hydrophobic, hydrogen-bond donor and acceptor features). The mutual similarity (A_F) is measured by spatially superimposing the molecules and computing the scalar product of the associated property vectors, weighted by the integral of the overlapping Gaussian functions. To compute the mean, all considered molecules are embedded into a regularly spaced grid. Subsequently, at each grid point the contributions of the Gaussians of all molecules are summed, either forming the arithmetic or geometric mean. Finally, the resulting values at the different grid points are translated back into a set of associated Gaussian functions (fig. 5)

either the arithmetic or the geometric mean. Finally, the resulting values at the different grid points are translated back into a set of associated Gaussian functions (figs 5, 6). Accordingly, the derived field-based pharmacophore reflects the properties of superimposed ligands in terms of associated Gaussian functions. The two alternatives for calculating the mean (fig. 7) represent either the properties that all molecules allot similarly in the same spatial area (AND-pharmacophore, geometric mean), or the properties that are represented in a particular region in space by at least one molecule of the superimposed set (OR-pharmacophore, arithmetic mean). Figs 5 and 6 illustrate this averaging for the three potent hAR inhibitors sorbinil, tolrestat, and zopolrestat. The ligands were used in the binding site conformation as found in the corresponding protein structures. They were mutually superimposed based on a rmsd fit considering the C_α-atoms of the protein backbone. To assess the scope and reliability of this superpositioning procedure, the three ligands used to define the supermolecule reference, as well as IDD594, fidarestat, zenarestat and zopolrestat were selected to test the predictive power with respect to the binding modes. Each test molecule was pairwise aligned with the OR field-based pharmacophore (supermolecule reference). The obtained RMS deviations between the best-scored SEAL solution and the crystallographically known binding mode are found in a range between 0.5 and 1.6 Å. These small deviations demonstrate the validity of our approach as a strategy to perform a VS for the multiple conformers in AR. In addition, we also tried to superimpose IDD384 with the reference ligand-based pharmacophore. In the complex with AR, IDD384 adopts a different binding mode not explored by any of the three reference molecules considered in the ligand-based pharmacophore. Not surprisingly, the approach is not able to reproduce this binding mode in agreement with experiment. This definitely exceeds the scope of this concept, which 'only' maps similarities of ligands in space.

Keeping this inherent limitation in mind, we used in the following the OR field-based pharmacophore to generate structural alignments with candidate molecules in a virtual screening run. As reference sample we selected the Maybridge database composed of 60,000 entries [42]. Prior to the actual superpositioning step using SEAL, we preselected molecules comprising a suitable anchoring group to bind to the anion binding pocket. Accordingly, we demanded that all candidate molecules possess either a carboxylate, acid amide or hydantoin moiety. This connectivity criterion limited the database entries to 3800 candidate molecules. Furthermore, we applied a filter

considering the criteria introduced by Lipinski et al. [26], and we concentrated on compounds with a molecular weight <350 D and possessing less than five rotatable bonds. This reduced our list of putative candidates to about 2700 entries.

Subsequent to these 2D connectivity searches, performed with UNITY, we conducted a 3D search with the same program based on the predefined field-based pharmacophore hypothesis (based on zopolrestat, tolrestat and sorbinil), which was translated into a flexible 3D search query in UNITY (fig. 8). The centres 1 to 4 to be matched were assigned according to the AND field-based pharmacophore. The additional centres 5 and 6 correspond to areas only indicated by the OR field-based pharmacophore. The latter centers reflect the occupancy of the alternative binding pockets. To regard this optional placement, the 'partial match' command in UNITY was used. To avoid clashes with the protein, a series of excluded volumes was defined in the UNITY search query. These areas were assigned in a way to roughly approximate the accessible surface of the protein in the rigid anion binding pocket (fig. 2). After applying these additional filter criteria, 150 entries remained for further analysis.

With the 150 entries thus retrieved, a mutual flexible superposition onto the OR-based pharmacophore was performed with the combined MIMUMBA/SEAL approach. The entries were ranked according to their SEAL score, and the 75 best-ranked solutions were visually inspected. After careful analysis of the best-predicted compounds with respect to the geometry of the three alternative AR binding site conformers adopted in the complexes with sorbinil, tolrestat and zopolrestat, we acquired four molecules and subjected them to enzyme inhibition testing.

All selected compounds possess a carboxylate group to anchor the anion binding pocket. Figure 9 gives the obtained IC_{50} (concentration of the inhibitor that produced half-maximal effect) values of the selected hits, which are in the micromolar inhibition range. Most remarkable is the fact that this virtual screening search retrieves a set of new molecular skeletons, distinct from the hits obtained by the above-described docking approach using the protein-based pharmacophore extracted from the IDD594 binding pocket. Furthermore, this search suggests interesting compound classes, e.g. the benzothiazepines, not yet described as possible lead structure for AR inhibition. The latter ligand-based pharmacophore search also has limitations. Apparently, the similarity score used to rank and discriminate the generated solutions is only of limited scope. Accordingly, the final visual inspection of the solutions generated becomes even more important com-

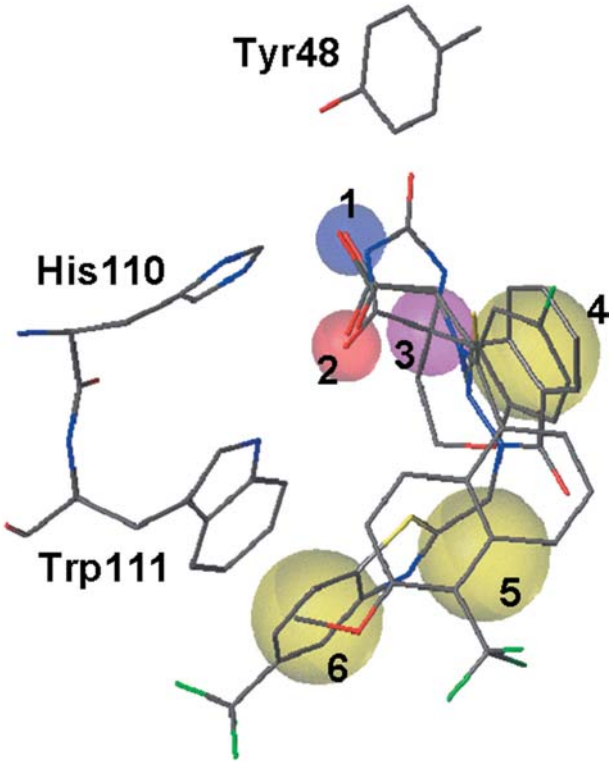

Figure 8. Ligand-based pharmacophore hypothesis extracted from the properties of the merged supermolecule reference. In figure 5 the averaging in terms of the OR and AND pharmacophores with repect to the steric properties is shown. Similarly produced maps indicate the electrostatic, hydrophobic and H-bonding properties. In area 1 a donor facility is required; 2 requests an acceptor. Area 3 requires a C, N, P, O or S atom, whereas 4–6 are assigned to hydrophobic groups. Areas 5 and 6 can be matched optionally.

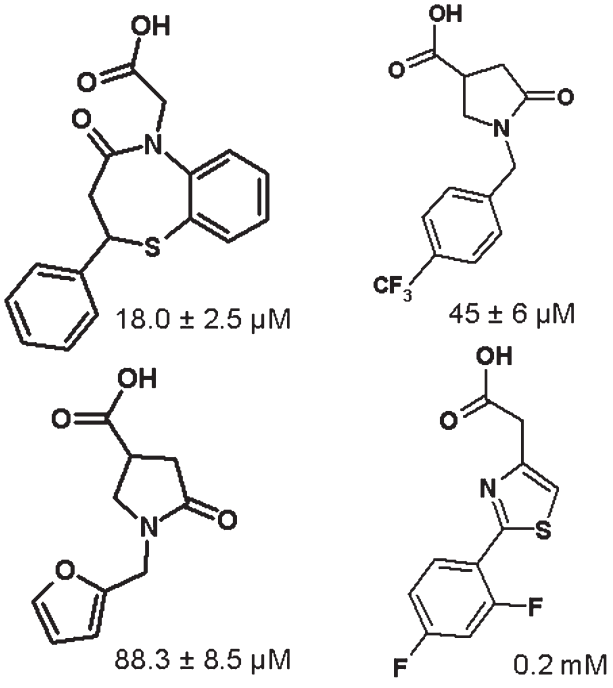

Figure 9. Chemical formulae of the four virtual screening hits obtained from the second ligand-based search which were selected for experimental testing. The measured IC_{50} values are given.

pared with the docking approach with the protein-based pharmacophore. In this step, the bound conformations produced and the complementarity with the different binding site conformers have to be analyzed carefully. This step requires some experience and – most important – can only be applied to a limited set of putative hits. Thus, this second screening was performed only on a rather limited database holding about 60,000 initial entries. Nevertheless, the results obtained demonstrate the relevance and feasibility of this concept.

Conclusions

AR is an enzyme of pronounced conformational adaptability that accepts various substrate and inhibitor skeletons. The detailed analysis of the presently available crystallographic information together with the frames obtained from MD trajectories suggests a composite picture of various protein conformers competent to accommodate a broad range of ligand skeletons with different binding site geometries. This fact provides additional challenge to virtual screening for inhibitors of AR and makes comparative scoring of the ligand binding modes produced extremely difficult. Besides the scoring of the actual protein-to-ligand interaction contacts, the energetic differences of various protein conformers would have to be regarded. We have followed two different concepts to retrieve active molecules from virtual screening. Both approaches were successful, leading to the discovery of micromolar inhibitors. The first approach arbitrarily considers the enzyme to be rigid and used one conformer likely to accommodate strong-binding AR inhibitors. A protein-based pharmacophore hypothesis is used to select candidate molecules in a hierachical filtering process and to dock them to the protein. In a second alternative approach simultaneous consideration of the multiple protein conformers is attempted. However, the various binding pockets are not considered as reference, but a set of different ligand molecules is merged to describe the multiplicity of putative ligand binding modes. As a search criterion to be matched, the spatial similarity of ligands merged into a supermolecule reference is used. The advantage of this approach is that it circumvents several of the above-mentioned scoring aspects and uses only the similarity with the supermolecule reference as criterion. Upon flexible superpositioning, the candidate molecule to be tested decides which of the considered binding modes is most appropriate for its placement. This advantage of simplifying the problem is to some degree compromised by the reduced discriminating power of the similarity measure used for scoring the subsequently suggested binding modes. Due to this limitation elaborate visual inspection of the considered hits has to be performed. Both approaches are tractable concepts for virtual screening using targets with pronounced conformational adaptability. The successful retrieval of several new molecular skeletons suggests this conclusion. However, the described concepts are only two possibilities. The field still requires the development of other strategies and ideas to come up with a general and reliable concept for the application of virtual screening to flexible protein targets.

Acknowledgement. The authors are grateful to Dr Alberto Podjarny (IGBMC, Strasbourg, France) for providing the coordinates of the high-resolution AR complex prior to publication. Very fruitful discussions with the Strasbourg group are appreciated. The present project has been funded by the Deutsche Forschungsgemeinschaft (grant KL 1204/3).

1 Oates P. J. (2002) Polyol pathway and diabetic peripheral neuropathy. Int. Rev. Neurobiol. **50:** 325–392
2 Yabe-Nishimura C. (1998) Aldose reductase in glucose toxicity: a potential target for the prevention of diabetic complications. Pharmacol. Rev. **50:** 21–33
3 Petrash J. M., Tarle I., Wilson D. K. and Quiocho F. A. (1994) Aldose reductase catalysis and crystallography. Insights from recent advances in enzyme structure and function. Diabetes **43:** 955–959
4 Brownlee M. (2001) Biochemistry and molecular cell biology of diabetic complications. Nature **414:** 813–820
5 Miyamoto S. (2002) Molecular modeling and structure-based drug discovery studies of aldose reductase inhibitors. ChemBio. Informatics Journal **2:** 74–85
6 Suzen S. and Buyukbingol E. (2003) Recent studies of aldose reductase enzyme inhibition for diabetic complications. Curr. Med. Chem. **10:** 1329–1352
7 Howard E., Sanishvili R., Cachau R. E., Mitschler A., Chevrier B., Barth P. et al. Human aldose reductase – inhibitor complex at 0.66 A: experimentally observed protonation states and atomic interactions have implications for the inhibition mechanism. Proteins, in press
8 Costantino L., Rastelli G., Vescovini K., Cignarella G., Vianello P., Del Corso A. et al. (1996) Synthesis, activity and molecular modeling of a new series of tricyclic pyridazinones as selective aldose reductase inhibitors. J. Med. Chem. **39:** 4396–4405
9 Rastelli G. and Costantino L. (1998) Molecular dynamics simulations of the structure of aldose reductase complexed with the inhibitor tolrestat. Bioorg. Med. Chem. Lett. **8:** 641–646
10 Lee Y. S., Chen Z. and Kador P. F. (1998) Molecular modeling studies of the binding modes of aldose reductase inhibitors at the active site of human aldose reductase. Bioorg. Med. Chem. **6:** 1811–1819
11 Rastelli G., Vianello P., Barlocco D., Costantino L., del Corso A. and Mura U. (1997) Structure-based design of an inhibitor modeled at the substrate active site of aldose reductase. Bioorg. Med. Chem. Lett. **7:** 1897–1902
12 Wilson D. K., Quiocho F. A. and Petrash J. M. Structural Studies on Aldose Reductase Inhibition, (1997) In: Structure-Based Drug Design, pp. 229–246, Marcel Dekker, New York, Veerapandian P. (ed.)
13 Iwata Y., Arisawa M., Hamada R., Kita Y., Mizutani M. Y., Tomioka N. et al. (2001) Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis. J. Med. Chem. **44:** 1718–1728
14 Rastelli G., Ferrari A. M., Costantino L. and Gamberini M. C. (2002) Discovery of new inhibitors of aldose reductase from molecular docking and database screening. Bioorg. Med. Chem. **10:** 1437–1450

15 Sotriffer C., Krämer O., Klebe G. Probing flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations. Proteins, in press
16 Hendlich M., Bergner A., Gunther J. and Klebe G. (2003) Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. J. Mol. Biol. **326:** 607–620
17 Van Zandt M. C., Sibley E. O., Combs K. J., McCann E. E., Flam B., Lavoie D. J. et al. (1999) Design and synthesis of novel inhibitors of aldose reductase for the treatment of diabetic complications, 218th ACS National Meeting, New Orleans
18 Freire E. (1998) Statistical thermodynamic linkage between conformational and binding equilibria. Adv. Protein Chem. **51:** 255–279
19 Ma B., Kumar S., Tsai C. J. and Nussinov R. (1999) Folding funnels and binding mechanisms. Protein Eng. **12:** 713–720
20 Lyne P. D. (2002) Structure-based virtual screening: an overview. Drug Discov. Today **7:** 1047–1055
21 Schneider G. and Böhm H. J. (2002) Virtual screening and fast automated docking methods. Drug Discov. Today **7:** 64–70
22 Walters W. P., Stahl M. T. and Murcko M. A. (1998) Virtual screening: an overview. Drug Discov. Today **3:** 160–178
23 Krämer O., Hazemann I., Podjarny A. D. and Klebe G. Ultrahigh resolution drug design III: virtual screening for inhibitors of human aldose reductase. Proteins, in press
24 Brenk R., Naerum L., Gradler U., Gerber H. D., Garcia G. A., Reuter K. et al. (2003) Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. J. Med. Chem. **46:** 1133–1143
25 Grüneberg S., Stubbs M. T. and Klebe G. (2002) Succesful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. J. Med. Chem. **45:** 3588–3602
26 Lipinski C. A., Lombardo F., Dominy B. W. and Feeney P. J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Delivery Rev. **23:** 3–25
27 Unity: Unity Chemical Information Software (ed 4.1.1.). St. Louis, MO, Tripos Inc., Software
28 Gohlke H., Hendlich M. and Klebe G. (2000) Predicting binding modes, binding affinities and 'hot spots' for protein-ligand complexes using a knowledge-based scoring function. Persp. Drug Discov. Design **20:** 115–144
29 Goodford P. J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J. Am. Chem. Soc. **28:** 849
30 Verdonk M. L., Cole J. C. and Taylor R. (1999) SuperStar: a knowledge-based approach for identifying interaction sites in proteins. J. Mol. Biol. **289:** 1093–1108
31 Gohlke H., Hendlich M. and Klebe G. (2000) Knowledge-based scoring function to predict protein-ligand interactions. J. Mol. Biol. **295:** 337–356
32 Pan Y., Huang N., Cho S. and MacKerell A. D. J. (2002) Consideration of molecular weight during compound selection in virtual target-based database screening. J. Chem. Inf. Comput. Sci. **43:** 267–272
33 Bocola M., Stubbs M. T., Sotriffer C., Hauer B., Friedrich T., Dittrich K. et al. (2003) Structural and energetic determinants for enantiopreferences in kinetic resolution of lipases. Protein Eng. **16:** 319–322
34 Österberg F., Morris G. M., Sanner M. F., Olson A. J., Goodsell D. S. (2002) Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins **46:** 34–40
35 Claussen H., Buning C., Rarey M. and Lengauer T. (2001) FlexE: efficient molecular docking considering protein structure variations. J. Mol. Biol. **308:** 377–395
36 Klebe G., Mietzner T. and Weber F. (1994) Different approaches toward an automatic structural alignment of drug molecules: applications to sterol mimics, thrombin and thermolysin inhibitors. J. Comput. Aided Mol. Des. **8:** 751–778
37 Klebe G., Mietzner T. and Weber F. (1999) Methodological developments and strategies for a fast flexible superposition of drug-size molecules. J. Comput. Aided Mol. Des. **13:** 35–49
38 Kearsley S. K. and Smith G. M. (1990) An alternative method for the alignment of molecular structures: maximizing electrostatic and steric overlap. Tetrahedron Comp. Methodol. **3:** 615–633
39 Cramer III R. D., Patterson D. E. and Bunce J. D. (1988) Comparative molecular field analysis (CoMFA). I. Effect of shape on binding of bteroids to carrier proteins. J. Am. Chem. Soc. **110:** 5959
40 Klebe G., Abraham U. and Mietzner T. (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J. Med. Chem. **37:** 4130
41 Klebe G. and Mietzner T. (1994) A fast and efficient method to generate biologically relevant conformations. J. Comput. Aided Mol. Des. **8:** 583–606
42 Maybridge: Database. UK, Maybridge Chemical Co. Ltd., 1999